THEORY OF COMPUTING, Volume 18 (4), 2022, pp. 1–46
www.theoryofcomputing.org

SPECIAL ISSUE: APPROX-RANDOM 2019

Improved Pseudorandom Generators from Peudorandom Multi-switching Lemmas

Rocco A. Servedio* Li-Yang Tan†

Received June 30, 2019; Revised December 30, 2020; Published March 2, 2022

Abstract. We give the best known pseudorandom generators for two touchstone classes in unconditional derandomization: small-depth circuits and sparse $\mathbb{F}_2$ polynomials. Our main results are an ε-PRG for the class of size-M depth-d AC^0 circuits with seed length $\log(M)^{d+O(1)} \cdot \log(1/\varepsilon)$, and an ε-PRG for the class of S-sparse $\mathbb{F}_2$ polynomials with seed length $2^{O(\sqrt{\log S})} \cdot \log(1/\varepsilon)$. These results bring the state of the art for unconditional derandomization of these classes into sharp alignment with the state of the art for computational hardness for all parameter settings: substantially improving on the seed lengths of either PRG would require a breakthrough on longstanding and notorious circuit lower bound problems.

The key enabling ingredient in our approach is a new *pseudorandom multi-switching lemma*. We derandomize recently developed *multi*-switching lemmas, which are powerful generalizations of Håstad's switching lemma that deal with *families* of depth-two circuits. Our pseudorandom multi-switching lemma—a randomness-efficient algorithm for sampling restrictions that simultaneously simplify all circuits in a family—achieves the parameters obtained by the (full randomness) multi-switching lemmas of Impagliazzo, Matthews, and

A preliminary version of this paper appeared in the Proceedings of the 23rd International Conference on Randomization and Computation (RANDOM 2019) [68]

*Supported by CCF-1420349 and CCF-1563155

†Supported by CCF-1563122

ACM Classification: F.1.2

AMS Classification: 68Q17

Key words and phrases: pseudorandom generators, switching lemmas, circuit complexity

 DOI: 10.4086/toc.2022.v018a004

Paturi (SODA'12) and Håstad (SICOMP 2014). This optimality of our derandomization translates into the optimality (given current circuit lower bounds) of our PRGs for AC^0 and sparse $\mathbb{F}_2$ polynomials.

1 Introduction

Switching lemmas. Switching lemmas, first established in breakthrough work in the 1980s [4, 32, 79, 37], are fundamental results stating that depth-two circuits (ORs of ANDs or vice versa) simplify dramatically when they are "hit with a random restriction." They are a powerful technique in circuit complexity, and are responsible for a remarkable suite of hardness results concerning small-depth Boolean circuits (AC^0). Switching lemmas are at the heart of several near-optimal bounds on AC^0 circuits, such as essentially optimal correlation bounds against the PARITY function [43, 38] and the worst-case and average-case depth hierarchy theorems of [37, 42, 39]. Indeed, comparably strong results are lacking (and are major open problems) for seemingly small extensions of AC^0, such as AC^0 augmented with parity or mod-p gates, for which switching lemmas do not apply; this gap highlights the importance of switching lemmas as a proof technique.

Switching lemmas are versatile as well as powerful: many results in circuit complexity rely on sophisticated variants and generalizations of the "standard" switching lemmas. Recent examples include the aforementioned correlation bounds and average-case depth hierarchy theorems, as well as powerful lower bounds on the circuit complexity of the CLIQUE problem [14, 64], lower bounds on the small-depth circuit complexity of ST-CONNECTIVITY [28], and lower bounds against AC^0 formulas [65]. Beyond the immediate arena of circuit lower bounds, switching lemmas are also important tools in diverse areas including propositional proof complexity [59, 47, 60, 40], computational learning theory [48], the design of circuit satisfiability algorithms [16, 43], and coding theory [26, 12].

This paper is about the role of switching lemmas in the study of *unconditional pseudorandomness*. Switching lemmas have a long history in this area; indeed, arguably the first work in unconditional derandomization, the seminal paper of Ajtai and Wigderson [6], was based on a *pseudorandom* switching lemma, which they used to give the first non-trivial pseudorandom generator for AC^0. (Interestingly, after many subsequent developments described in detail in Section 2, we come full circle in this paper and use the [6] framework to give a new pseudorandom generator for AC^0 that is essentially best possible without improving longstanding circuit lower bounds.) One key contribution that we make in this paper is to bring together two important generalizations of standard switching lemmas, one quite old and one very new:

(i) *pseudorandom* switching lemmas (originating in [6]), which employ pseudorandom rather than "fully random" restrictions, and

(ii) recently developed *multi-switching lemmas* [43, 38] which simultaneously simplify all of the depth-two circuits in a family of such circuits, rather than a single depth-two circuit as is the case for standard switching lemmas.

Let us discuss each of these generalizations in turn.

Pseudorandom switching lemmas. The (truly) random restrictions that are used in standard switching lemmas make an independent random choice for each input variable $x_1, \ldots, x_n$ of whether to map it to 0, to 1, or to leave it unassigned (map it to $*$); standard switching lemmas show that a depth-two circuit simplifies dramatically with very high probability when it is hit with such a random restriction. Such "truly random" restrictions are inherently incompatible with unconditional derandomization, which naturally motivates the notion of a *pseudorandom* switching lemma. Such a result defines a much smaller probability space of "pseudorandom" restrictions, and proves that a restriction drawn randomly from this space also has the effect of simplifying a depth-two circuit with high probability. While pseudorandom switching lemmas have been the subject of much research since they were first introduced by Ajtai and Wigderson [6, 5, 27, 3, 35, 43, 34, 73, 33], and have been applied in a range of different ways in unconditional derandomization, they are not yet fully understood.

The designer of a pseudorandom switching lemma faces an inherent tension between achieving strong parameters—intuitively, having a depth-two circuit simplify as much as possible while keeping a large fraction of variables alive—and using as little randomness as possible. Prior to the work of Trevisan and Xue [73], known pseudorandom switching lemmas fell short of achieving the parameters of Håstad's influential "full randomness" switching lemma [37]. In particular, a parameter of central importance in essentially all applications of switching lemmas is the probability that a given coordinate x_i remains alive under a random (or pseudorandom) restriction; this is often referred to as the "$*$-probability" and denoted by p. A crucial quantitative advantage of Håstad's switching lemma over previous results is that it can be applied even when p is as large as $\Omega(1/\log(n))$ for $\mathrm{poly}(n)$-size depth-two circuits—in contrast, the earlier articles [4, 32, 79] required $p = n^{-\Omega(1)}$—and yields a very strong conclusion, namely that with high probability the restricted circuit collapses to a shallow decision tree.[1] (For example, while the recent pseudorandom switching lemma of [34] is able to achieve a relatively large p, the conclusion of that switching lemma is that the restricted depth-two circuit can w.h.p. be sandwiched by depth-two circuits with small bottom fan-in, which is weaker than the aforementioned decision-tree conclusion.)

Trevisan and Xue [73] give a *pseudorandom* switching lemma that is highly randomness efficient and yet achieves the parameters of Håstad's fully random switching lemma (i. e., [73] achieves the same simplification, collapsing to a shallow decision tree, that follows from [37], with the same $*$-probability p as [37]). The key conceptual ingredient enabling this is a beautiful idea of "fooling the proof" of Håstad's switching lemma, exploiting its "computational simplicity." Trevisan and Xue leverage their pseudorandom switching lemma to construct a new pseudorandom generator for AC^0, obtaining the first improvement of Nisan's celebrated PRG [57] in over two decades. We elaborate on Trevisan and Xue's ideas and how they obtain their PRG later in Section 2.1.

Multi-switching lemmas. The switching lemma shows that any width-k CNF formula collapses to a shallow decision tree with high probability under a random restriction. Via a simple union bound it is of course possible to extend this result to say that a family of width-k CNF formulas will all collapse to a shallow decision tree with high probability under a random restriction; but this naive approach leads to a quantitative loss in parameters if the argument is iterated, as it typically is, $d-1$ times to analyze a depth-d circuit. (The exact nature of this quantitative loss is important but somewhat subtle; see Section 3

[1] The first published version of the switching lemma with a decision tree conclusion is due to Cai [22]; several authors subsequently noted that Håstad's argument also yields such a conclusion.

for a detailed explanation.)

Via an ingenious extension of the ideas underlying the original switching lemma, Håstad [38] developed "multi-switching lemmas" that essentially bypass this quantitative loss in parameters that results from iterating a naive union bound (see also the work of Impagliazzo, Matthews, and Paturi [43] for closely related results). Roughly speaking, [38] shows that a *family* of width-k CNF formulas will with high probability have a shallow *common partial decision tree*. Without explaining this structure in detail here (again see Section 3 for a detailed explanation), this makes it possible to iterate the argument and tackle depth-d circuits without incurring a quantitative loss in parameters. The savings thus achieved is the key new ingredient that allowed [43, 38] to achieve essentially optimal correlation bounds for AC^0 against the PARITY function, capping off a long line of work [4, 79, 37, 22, 10, 16]. These ideas have also been leveraged to achieve new algorithmic results such as better-than-brute-force satisfiability algorithms and distribution-free PAC learning algorithms for AC^0 [16, 43, 66].

A pseudorandom multi-switching lemma. A core technical contribution of this paper is to bring together these two lines of work, on pseudorandom switching lemmas and on multi-switching lemmas. Since the precise statement of our pseudorandom multi-switching lemma, Theorem 4.2, is somewhat involved we defer it to Section 4 and here merely make some remarks about it. In the spirit of Trevisan and Xue's derandomization of the original switching lemma, to obtain Theorem 4.2 we "fool the proof" of Håstad's multi-switching lemma [38], exploiting its "computational simplicity." This enables us to achieve optimal parameters in the same sense as [73], namely, that it establishes the same dramatic simplification—now of the family $\mathscr{F}$ of depth-two circuits—as [38], while only requiring the same $*$-probability p as [38]. Our pseudorandom switching lemma is highly efficient in its use of randomness; this randomness efficiency is crucial in the constructions of our pseudorandom generators for AC^0 circuits and sparse $\mathbb{F}_2$ polynomials using Theorem 4.2, which we describe in the next section.[2]

2 PRGs for AC^0 and sparse $\mathbb{F}_2$ polynomials

We employ our pseudorandom multi-switching lemma to give the best known pseudorandom generators for two canonical classes in unconditional derandomization: AC^0 circuits and sparse $\mathbb{F}_2$ polynomials. As we describe in this section, our results bring the state of the art for unconditional derandomization of these classes into sharp alignment with the state of the art for computational hardness. In this sense, our results are in the same spirit as those of Imagliazzo, Meka, and Zuckerman [44], which gave optimal (assuming current circuit lower bounds) pseudorandom generators for various classes of Boolean formulas and branching programs; however, our techniques are very different from those of [44].

[2] While our focus in this article is on unconditional derandomization, we briefly mention that recent work of Ball et al. [12] establishes a new connection between pseudorandom switching lemmas and *non-malleable codes* in coding theory [30]. Using this connection, [12] is able to leverage the randomness efficiency of Trevisan and Xue's pseudorandom switching lemma [73] in its design of new non-malleable codes for small-depth circuits. We leave the possibility of applying our techniques to obtain further-improved non-malleable codes as an interesting avenue for future work.

2.1 PRGs for AC^0 circuits

The class of small-depth Boolean circuits (AC^0) is a class of central interest in unconditional derandomization, and has been the subject of intensive research in this area over the past 30 years [6, 49, 57, 58, 54, 53, 45, 72, 74, 13, 62, 21, 46, 29, 2, 1, 69, 51, 31, 35, 34, 73, 33, 70, 36, 24]. This highly successful line of work on derandomizing AC^0 has generated a wealth of ideas and techniques that have become mainstays in the field of pseudorandomness. A prominent example is Nisan's celebrated PRG for AC^0 circuits [57], which introduced ideas that enriched the surprising connections between pseudorandomness and computational hardness [17, 78, 58]. The *hardness-versus-randomness paradigm* asserts, qualitatively, that strong explicit PRGs exist if and only if strong explicit circuit lower bounds exist. In the context of unconditional derandomization (the subject of this article), this strongly motivates the goal of constructing, for every circuit class $\mathscr{C}$, unconditional PRGs for $\mathscr{C}$ that are best possible given the current best lower bounds for $\mathscr{C}$. In other words, this is the goal of achieving an *optimal hardness-to-randomness conversion* for $\mathscr{C}$, converting "all the hardness" in our lower bounds for $\mathscr{C}$ into pseudorandomness for $\mathscr{C}$.

For $\mathscr{C}$ being the class of n-variable size-M depth-d AC^0 circuits, this amounts to constructing PRGs with seed length $\log^{d-1}(Mn)\log(1/\varepsilon)$; such seed length is essentially the best possible without improving longstanding AC^0 lower bounds that date back to the 1980s [37]. More precisely, it is well known (see, e. g., the subsection "Barriers to further progress" in [73, Section 1]) that achieving seed length $\log^{d-1.01}(Mn)\log(1/\varepsilon)$ would yield $\exp(\omega(n^{1/(d-1)}))$ size lower bounds against depth-d AC^0 circuits; this is a barrier that has stood for over 30 years even in the $d=3$ case. We give the first PRG that achieves a seed length of $\log^{d+O(1)}(Mn)\log(1/\varepsilon)$:

Theorem 2.1 (PRG for AC^0 circuits)**.** *For every $d \geq 2$, $M \in \mathbb{N}$ and $\varepsilon > 0$, there is an ε-PRG for the class of n-variable size-M depth-d circuits with seed length* $\log^{d+O(1)}(Mn)\log(1/\varepsilon)$.

We note that (to the best of our knowledge) a seed length of $\log^{d+O(1)}(Mn)+\log(1/\varepsilon)$ would not imply new circuit lower bounds, and neither would a seed length of $\log^{d-1}(Mn)\log(1/\varepsilon)$. We leave the design of pseudorandom generators achieving these seed lengths as interesting subjects for future work.

2.1.1 Background and prior PRGs for AC^0 circuits

As noted above, there has been a significant body of work on PRGs for AC^0 circuits, spanning over 30 years. In this section we give a brief overview of the history and prior state of the art for this touchstone problem in unconditional derandomization.

Ajtai–Wigderson and Nisan. Ajtai and Wigderson, in their seminal paper [6] pioneering the study of unconditional derandomization, constructed the first non-trivial PRG for AC^0 circuits with an $n^{o(1)}$ seed length; we will discuss their techniques in detail later. The seed length of the [6] generator was improved significantly in the celebrated work of Nisan [57], using what is now known as the Nisan–Wigderson framework [58], which provides a generic template for converting correlation bounds against a circuit class to PRGs for a closely related class (in the case of AC^0 these two classes essentially coincide). Via this approach Nisan showed how correlation bounds for AC^0 against the PARITY function [37] yield a PRG with seed length $\log^{2d+O(1)}(Mn/\varepsilon)$.

We remark that the generality of the Nisan–Wigderson framework comes at a quantitative price: it is straightforward to verify that a seed length of $(\log^d(Mn) + \log(1/\varepsilon))^2$ is the best that can be achieved via this framework given current AC^0 circuit lower bounds (see, e. g., [73, 36]). This is roughly quadratically worse than the best that can be achieved assuming *only* current AC^0 circuit lower bounds.

Bounded independence fools AC^0. Nisan's seed length for fooling AC^0 circuits stood unchallenged for more than two decades. However, in this interim period there was significant progress on showing that distributions with bounded independence fool AC^0, a well-known conjecture posed by Linial and Nisan [49]. Braverman's breakthrough, which built on an earlier result of Bazzi [13] and its subsequent simplification by Razborov [62], showed that polylog(n)-wise independence fools AC^0. Using standard constructions of k-wise independent distributions, this gave a PRG with seed length $\log^{O(d^2)}(Mn/\varepsilon)$; this was subsequently sharpened to $\log^{3d+O(1)}(Mn/\varepsilon)$ by Tal [70]. Recently, Harsha and Srinivasan [36] further improved the seed length of Braverman's generator to $\log^{3d+O(1)}(Mn)\log(1/\varepsilon)$, which is notable for its optimal dependence on the error parameter ε.

The work of Trevisan and Xue. Recent work of Trevisan and Xue [73] makes a significant advance towards achieving seed length $\log^{d-1}(Mn)\log(1/\varepsilon)$: their work circumvents the "quadratic loss" associated with the Nisan–Wigderson framework with a PRG of seed length $\log^{d+O(1)}(Mn/\varepsilon)$. This is the first PRG to achieve a $\log^{d+O(1)}(Mn)$ dependence, an exponent that is within an *additive* absolute constant of the sought-for $\log^{d-1}(Mn)$, and is also the first strict improvement on Nisan's seed length in more than two decades. (Note, however, that, like in Nisan's PRG, the dependence on ε is not optimal: $\log^{d+O(1)}(1/\varepsilon)$ instead of $\log(1/\varepsilon)$.)

Rather than going through the Nisan–Wigderson framework—which, as noted above, carries with it an associated quantitative loss in parameters—Trevisan and Xue construct their PRG by *derandomizing the proof* of AC^0 lower bounds, "opening up the black-box" of AC^0 lower bounds, so to speak. At a high level, Trevisan and Xue [73] adopt the strategy employed in the early work of Ajtai and Wigderson [6]. We describe this strategy in detail in Section 5, but roughly speaking, Ajtai and Wigderson introduced a powerful and generic framework for constructing PRGs from pseudorandom switching lemmas. In [6], they instantiated this framework with a derandomization of Ajtai's switching lemma [4]—which underlies his proof of the first superpolynomial lower bounds against AC^0—to obtain the first non-trivial PRG for AC^0. Trevisan and Xue obtain their PRG by revisiting this early framework of [6], instantiating it with their derandomization of Håstad's switching lemma [37]. (And as we will soon discuss, in this paper we obtain our PRG by instantiating the [6] framework with our derandomization of the [38] multi-switching lemmas.)

PRGs via polarizing random walks. Finally, in recent exciting work, Chattopadhyay, Hatami, Hosseini, and Lovett [24] have introduced an elegant new framework for obtaining pseudorandom generators which has consequences for fooling AC^0. Their framework is based on a notion of "fractional" pseudorandom generators, which are used as steps in a random walk which ultimately yields a (standard) pseudorandom generator. Chattopadhyay et al. [24] show that if a class $\mathscr{C}$ of circuits is closed under restrictions and has sufficiently strong Fourier concentration on low-degree coefficients, then almost k-wise independence suffices to yield a fractional PRG, which their random walk approach can then convert into a standard PRG against $\mathscr{C}$. Using Tal's sharp bounds [70] on the Fourier concentration of AC^0, they obtain a seed length of $O(\log(n/\varepsilon)(\log(\log(n)/\varepsilon))\log^{2d-2}M)$ for size-M depth-d circuits.

2.1.2 Our PRG and approach

To summarize, prior to our work there were three incomparable best known PRGs for AC^0, achieving three different tradeoffs in the overall dependence on M, d and $1/\varepsilon$. These were the PRG of Trevisan and Xue [73], which has seed length $\log^{d+O(1)}(Mn/\varepsilon)$; Harsha and Srinivasan's improvement of Braverman's generator [36], which has seed length $\log^{3d+O(1)}(Mn)\log(1/\varepsilon)$; and the [24] PRG, which has seed length $O(\log(n/\varepsilon)(\log(\log(n)/\varepsilon))\log^{2d-2}M)$, i. e., essentially $\log^{2d-1}(Mn)\log^2(1/\varepsilon)$.

Theorem 2.1 unifies and improves these three incomparable seed lengths. Our PRG achieves an essentially optimal hardness-to-randomness conversion for AC^0: our seed length of $\log^{d+O(1)}(Mn)\log(1/\varepsilon)$ comes very close to $\log^{d-1}(Mn)\log(1/\varepsilon)$, which is best possible without improving longstanding AC^0 circuit lower bounds that date back to the 1980s.

Table 1 provides a comparison of the seed length of our PRG (and the techniques that underlie our construction) and those of previous work.

Reference	Seed length	Techniques
[6] AW	$n^{o(1)}$ for $M = \mathrm{poly}(n)$	derandomize [4] switching lemma
[57] Nisan	$\log^{2d+O(1)}(Mn/\varepsilon)$	[58] framework, [37] correlation bounds
[21] Braverman	$\log^{O(d^2)}(Mn/\varepsilon)$	bounded independence
[73] TX	$\log^{d+O(1)}(Mn/\varepsilon)$	[6] framework, derandomize [37] switching lemma
[70] Tal	$\log^{3d+O(1)}(Mn/\varepsilon)$	bounded independence
[36] HS	$\log^{3d+O(1)}(Mn)\log(1/\varepsilon)$	bounded independence
[24] CHHL	(essentially) $\log^{2d-1}(Mn)\log^2(1/\varepsilon)$	almost bounded independence, fractional PRGs, polarizing random walks
This paper	$\log^{d+O(1)}(Mn)\log(1/\varepsilon)$	[6] framework, derandomize [38] multi-switching lemma, bounded independence

Table 1: PRGs for ε-fooling n-variable size-M depth-d AC^0 circuits.

Our approach. Our approach draws on and unifies ideas from articles [6, 73, 36] discussed above, which we use in conjunction with our derandomization of Håstad's multi-switching lemma [38] to obtain our PRG.

At a high level, we adopt the overall conceptual strategy of Ajtai and Wigderson [6] and Trevisan and Xue [73], and obtain our PRG by derandomizing the proof of AC^0 lower bounds. The key technical ingredient in our PRG construction is our pseudorandom multi-switching lemma, a derandomization of the multi-switching lemmas which underlie the optimal correlation bounds for AC^0 against PARITY by

Impagliazzo et al. and Håstad [43, 38]. Our pseudorandom multi-switching lemma improves both the pseudorandom switching lemma of [73] (a derandomization of Håstad's switching lemma [37] which underlies his exponential lower bounds against AC^0) and the pseudorandom switching lemma of [6] (a derandomization of Ajtai's switching lemma [4] which underlies his superpolynomial lower bounds against AC^0).

Our derandomization of the [38] multi-switching lemma is largely influenced by Trevisan and Xue's derandomization of Håstad's original switching lemma [37]. We describe our approach in detail in Section 4, but highlight here the simple but ingenious new idea underlying the argument in [73]. Roughly speaking, they derandomize Håstad's switching lemma by "fooling its proof": showing that Håstad's proof of his switching lemma "cannot δ-distinguish" between truly random restrictions and pseudorandom restrictions drawn from polylog(n)-wise independent distributions. Since Håstad's switching lemma holds for truly random restrictions, it thus follows that it also holds for pseudorandom restrictions drawn from polylog(n)-wise independent distributions (up to a δ additive loss in the failure probability).

To accomplish this, Trevisan and Xue exploit the fact that Håstad's proof of the switching lemma is "computationally simple": for a fixed k-CNF F, there is a small depth-3 circuit that takes as input an encoding of a restriction ρ, and returns 1 iff ρ is a bad restriction for the desired conclusion of Håstad's switching lemma, contributing to its failure probability. (More precisely, the failure event is that the "canonical decision tree" for $F \upharpoonright \rho$ has large depth.) In a similar spirit, our derandomization of the [38] multi-switching lemma also exploits the "computational simplicity" of its proof. In our case, for a fixed family $\mathscr{F}$ of k-CNF formulas we construct a small depth-4 circuit for recognizing bad restrictions. (The one additional layer of depth reflects the fact that multi-switching lemmas are, roughly speaking, "one quantifier more complex" than switching lemmas.) To obtain optimal parameters in our PRG constructions, we use the $d = 4$ case of Harsha and Srinivasan's strengthening of Braverman's generator [36] to fool this depth-4 circuit, and hence show that Håstad's proof [38] of the multi-switching lemmas "cannot distinguish" between truly random and pseudorandom restrictions. The fact that [36] achieves an optimal $\log(1/\varepsilon)$ dependence on the seed length plays a crucial role in enabling the optimal $\log(1/\varepsilon)$ seed-length dependence of our PRG.

2.2 PRGs for sparse $\mathbb{F}_2$ polynomials

Our second main result deals with the class of sparse $\mathbb{F}_2$ polynomials. Like AC^0 circuits, sparse $\mathbb{F}_2$ polynomials and low-degree $\mathbb{F}_2$ polynomials have been extensively studied in unconditional derandomization [56, 8, 54, 18, 74, 50, 76, 19, 51, 52, 25].

Via the hardness-versus-randomness paradigm, the problem of derandomizing $\mathbb{F}_2$ polynomials is intimately related to that of proving correlation bounds for $\mathbb{F}_2$ polynomials. A prominent open problem in the latter context—arguably the current flagship challenge in this area—is that of obtaining superpolynomially small correlation bounds against $\mathbb{F}_2$ polynomials of degree $\log(n)$. Degree $\log(n)$ represents the fundamental limit of our current suite of powerful techniques for proving $\mathbb{F}_2$ correlation bounds [11, 20, 23, 77], and breaking this "degree-$\log(n)$ barrier" would constitute a significant technical breakthrough[3]. See Open Question 1 of Viola's excellent survey [75] for a detailed discussion of this

[3]Breaking this "degree-$\log(n)$ barrier" is also well known (via a simple and beautiful observation by Håstad and Goldmann [41]) to be a prerequisite for breaking the notorious "$\log(n)$-party barrier" in multi-party communication complexity [11],

important open problem and its relationship to other central challenges in complexity theory.

As a second application of our pseudorandom multi-switching lemma, we give an ε-PRG for S-sparse $\mathbb{F}_2$ polynomials with seed length $2^{O(\sqrt{\log S})}\log(1/\varepsilon)$, which is best possible without breaking the aforementioned "degree-$\log(n)$ barrier" for $\mathbb{F}_2$ correlation bounds.

Theorem 2.2 (PRG for sparse $\mathbb{F}_2$ polynomials)**.** *For every $S = 2^{\omega(\log\log(n))^2}$ and $\varepsilon > 0$ there is a PRG with seed length $2^{O(\sqrt{\log S})}\log(1/\varepsilon)$ that ε-fools the class of n-variable S-sparse $\mathbb{F}_2$ polynomials.*

Background and prior PRGs for $\mathbb{F}_2$ polynomials. The first unconditional PRGs for $\mathbb{F}_2$ polynomials were given in early influential work of Luby, Veličković, and Wigderson [54], who constructed a PRG that ε-fools size-S $\mathsf{SYM} \circ \mathsf{AND}$ circuits—including S-sparse $\mathbb{F}_2$ polynomials as an important special case—with seed length $2^{O(\sqrt{\log(S/\varepsilon)})}$. To obtain their PRG, Luby et al. employed the Nisan–Wigderson framework [58] together with multi-party number-on-the-forehead (NOF) communication complexity lower bounds from the seminal paper by Babai, Nisan, and Szegedy [11]. Viola [74] subsequently extended this $2^{O(\sqrt{\log(S/\varepsilon)})}$ seed length to the broader class of $\mathsf{SYM} \circ \mathsf{AC}^0$ circuits with a more modular proof. In recent work [67], the authors have improved the seed-length dependence on ε of [54, 74] to $2^{O(\sqrt{\log(S)})} + \mathrm{polylog}(1/\varepsilon)$. We discuss the relation between our techniques and those of [67] in more detail below.

In a related line of work, PRGs for *low-degree* $\mathbb{F}_2$ polynomials have also been intensively studied. Starting with the fundamental results of Naor and Naor [56] on ε-biased distributions (which resolved the degree-1 case), this research continued through an exciting line of work on the degree $k \geq 2$ case [18, 19] and culminated in the breakthroughs by Lovett [50] and Viola [76] which are described in more detail below. We note that prior to our work, the underlying techniques used for the sparse case (multi-party communication complexity) were completely different from the techniques used for the low-degree case (Fourier analysis).

Our PRG and approach. Theorem 2.2 gives an exponential and optimal improvement of the PRG of [54] in terms of its dependence on the error parameter ε. Our PRG achieves an optimal hardness-to-randomness conversion for $\mathbb{F}_2$ polynomials: since every $\mathbb{F}_2$ polynomial of degree $\log(n)$ has at most $n^{\log(n)}$ monomials, it can be shown (using the simple Proposition 3.1 of [76]) that a PRG with seed length $2^{o(\sqrt{\log S})}\log(1/\varepsilon)$ would break the degree-$\log(n)$ barrier.

Our techniques for Theorem 2.2 are substantially different from the techniques of [67, 74]. As summarized in Table 2, the basic approach of [67], like [74] and [54], is via the Nisan-Wigderson paradigm using multi-party communication complexity bounds; the main point of departure between [67] and [74] is that [67] leverages Håstad's multi-switching lemma from [38] in place of his earlier [37] switching lemma which was used in [74]. (We note, that, similarly to the situation for AC^0 circuits, it is straightforward to verify that our optimal $\log(1/\varepsilon)$ dependence is not achievable via the Nisan–Wigderson framework without dramatic breakthroughs in correlation bounds for $\mathbb{F}_2$ polynomials, going well beyond breaking the degree-$\log(n)$ barrier.) In contrast, we do not use the Nisan–Wigderson framework or multi-party communication complexity lower bounds; instead, as for AC^0, our approach is based on the [6] framework and our *derandomization* of the [38] multi-switching lemma. Indeed, our approach to

a longstanding open problem that has resisted attack for over three decades.

obtaining Theorem 2.2 bridges the two previously disparate lines of work on pseudorandomness for sparse and low-degree polynomials: roughly speaking, it can be viewed as a reduction from PRGs for S-sparse polynomials to PRGs for polynomials of degree $\sqrt{\log S}$. This allows us to leverage the result of Viola [76] (building on the work of Lovett [50]), which gives PRGs for n-variable degree-k $\mathbb{F}_2$ polynomials with seed length

$$O(k\log(n)+k2^k\log(1/\varepsilon)).$$

(We note that the lack of PRGs for $\mathsf{SYM}\circ\mathsf{AND}_d$ circuits of comparable strength to [76, 50] — in particular, the $\log(1/\varepsilon)$ dependence — for general SYM gates beyond the parity function are a barrier for extending our PRG for $\mathbb{F}_2$ polynomials to general depth-two $\mathsf{SYM}\circ\mathsf{AND}$ circuits.) More precisely, at the heart of our reduction is a new pseudorandom switching lemma for sparse $\mathbb{F}_2$ polynomials, showing that such a polynomial is very likely to collapse to a *small-depth decision tree with low-degree* $\mathbb{F}_2$ *polynomials at its leaves* under a suitable pseudorandom restriction. This is essentially a special case of our pseudorandom multi-switching lemma. With this reduction in hand, we then exploit the strength and generality of Viola's result—roughly speaking, that the sum of k independent copies of a sufficiently strong ε-biased distribution fools degree-k polynomials—to show that his PRG extends to fool not only low-degree polynomials, but also small-depth decision trees with low-degree polynomials at their leaves.

Table 2 provides a comparison of the seed length of our PRG (and the techniques that underlie our construction) and those of previous work.

Reference/ Class	Seed length	Techniques
[54] LVW S-sparse	$2^{O(\sqrt{\log(S/\varepsilon)})}$	[58] framework, [11] multi-party NOF communication complexity
[67] ST S-sparse	$2^{O(\sqrt{\log S})}+(\log(1/\varepsilon))^{4.01}$	[58] framework, [11] multi-party NOF communication complexity, [38] multi-switching lemma
[50] Lovett degree k	$O(2^k\log(n)+4^k\log(1/\varepsilon))$	Fourier analysis
[76] Viola degree k	$O(k\log(n)+k2^k\log(1/\varepsilon))$	Fourier analysis
This paper S sparse	$2^{O(\sqrt{\log S})}\log(1/\varepsilon)$	[6] framework, derandomize [38] multi-switching lemma, Fourier analysis, bounded independence

Table 2: PRGs for ε-fooling $\mathbb{F}_2$ polynomials.

2.3 Organization

Section 2.4 recalls some basic preliminaries from unconditional pseudorandomness. We describe and contrast the original Håstad switching lemma [37] versus the [38] multi-switching lemma in Section 3.

Section 3.1 establishes some infrastructure towards derandomizing the [38] switching lemma, and the actual derandomization is carried out in Section 4, culminating in the proof of Theorem 4.2. Section 5 describes a general framework for constructing pseudorandom generators that is implicit in the work of Ajtai and Wigderson [6]; a crucial ingredient in this framework for constructing a pseudorandom generator for a class $\mathscr{C}$ is a "pseudorandom simplification lemma" for $\mathscr{C}$. In Section 6 we apply our derandomized multi-switching lemma from Section 4 to obtain the required pseudorandom simplification lemmas for AC^0 circuits and for sparse $\mathbb{F}_2$ polynomials. Finally, Section 7 puts the pieces together and establishes the PRGs for AC^0 and for sparse $\mathbb{F}_2$ polynomials that are our main PRG results.

2.4 Preliminaries

All logarithms are in base 2. For $r < n$, we say that a distribution $\mathcal{D}$ over $\{0,1\}^n$ can be *sampled efficiently with r random bits* if (i) $\mathcal{D}$ is the uniform distribution over a multiset $\{z^{(1)},\dots,z^{(s)}\}$ of strings from $\{0,1\}^n$ where $s \le 2^r$ and (ii) there is a deterministic algorithm $\mathrm{Gen}_{\mathcal{D}}$ which, given as input a uniform random element of $[s]$, runs in time $\mathrm{poly}(n,s)$ and returns a string drawn from $\mathcal{D}$.

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

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

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

Two kinds of distributions which are extremely useful in derandomization are *δ-biased* and *k-wise independent* distributions. We say that a distribution $\mathcal{D}$ over $\{0,1\}^n$ is *δ-biased* if it δ-fools the class of all 2^n parity functions $\{\textsc{Parity}_S\}_{S \subseteq [n]}$, where $\textsc{Parity}_S : \{0,1\}^n \to \{0,1\}$ is defined by $\textsc{Parity}_S(x) = \sum_{i \in S} x_i \mod 2$. We say that a distribution $\mathcal{D}$ over $\{0,1\}^n$ is *k-wise independent with parameter p* if for every $1 \le i_1 < \cdots < i_k \le n$ and every $(b_1,\dots,b_k) \in \{0,1\}^k$, we have

$$\mathop{\mathbf{Pr}}_{x \leftarrow \mathcal{D}} \left[x_{i_1} = b_1 \text{ and } \cdots \text{ and } x_{i_k} = b_k \right] = p^{\sum_{j=1}^k b_j} \cdot (1-p)^{k - \sum_{j=1}^k b_j},$$

i. e., every subset of k coordinates is distributed identically to a product distribution with parameter p.

A *restriction* ρ of variables $x_1,\dots,x_n$ is an element of $\{0,1,*\}^n$. We write $\mathrm{supp}(\rho)$ to denote the set of coordinates that are fixed to 0 or 1 by ρ. Given a function $f(x_1,\dots,x_n)$ and a restriction ρ, we write $f \upharpoonright \rho$ to denote the function obtained by fixing x_i to $\rho(i)$ if $\rho(i) \in \{0,1\}$ and leaving x_i unset if $\rho(i) = *$. For $\rho \in \{0,1\}^S$ with $S \subsetneq [n]$, the notation "$f \upharpoonright \rho$" refers to the restriction of f obtained by "completing" ρ to an element of $\{0,1,*\}^n$ by setting $\rho_i = *$ for all $i \in [n] \setminus S$. For two restrictions $\rho, \rho' \in \{0,1,*\}^n$, their *composition*, denoted $\rho\rho'$ or $\rho \circ \rho'$, is the restriction (element of $\{0,1,*\}^n$) defined by

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

Given a collection $\mathscr{F} = \{f_1,\dots,f_M\}$ of functions and a restriction ρ we write $\mathscr{F} \upharpoonright \rho$ to denote the family $\{f_1 \upharpoonright \rho, \dots, f_M \upharpoonright \rho\}$.

Given an AC^0 circuit, we define its size to include the input variables (along with the number of gates in the circuit). We adopt this convention for notational convenience, since we may then always assume that the size M of an n-variable circuit is always at least n. (We do *not* adopt this convention for $\mathbb{F}_2$ polynomials: as is standard, we define the sparsity of an $\mathbb{F}_2$ polynomial to be the number of monomials in its support.)

Finally, if g is a Boolean function and $\mathscr{C}$ is a class of circuits, we say that g is *computed by a* $(t,\mathscr{C})$*-decision tree* if g is computed by a decision tree of depth t (with single Boolean variables x_i at internal nodes as usual) in which each leaf is labeled by a function from $\mathscr{C}$.

3 Multi-switching lemmas

At the heart of almost all applications of Håstad's original switching lemma [37] is a powerful structural fact about AC^0 circuits: every AC^0 circuit "collapses" (i. e., simplifies dramatically) to a depth-t decision tree with probability at least $1-\varepsilon$ under a random restriction that randomly fixes a $(1-p)$-fraction of coordinates. In the precise quantitative statement of this fact, both t and p depend on ε: as the desired failure probability ε tends to 0, the $*$-probability p tends to 0 (more coordinates are fixed) and t tends to n (the resulting decision tree is of larger depth). It is easy to see that this dependence is inherent given the statement of the [37] switching lemma, and indeed this will be clear from the discussion later in this section.

The recent multi-switching lemma of Håstad [38] (see also [43]) achieves a remarkable strengthening of the above: essentially the same structural fact about AC^0 holds (in terms of the quantitative relation between the decision tree depth t and the failure probability ε) *with the $*$-probability p being independent of ε*. This is the key qualitative difference underlying the optimal AC^0 correlation bounds for PARITY obtained in [43, 38]; likewise, in this work, this is the key qualitative difference underlying the optimal ε-dependence in the seed lengths of our PRGs for AC^0 circuits and sparse $\mathbb{F}_2$ polynomials.

Let $\mathcal{R}_p$ denote the random restriction which independently sets each variable x_i to 0 with probability $(1-p)/2$, to 1 with probability $(1-p)/2$, and to $*$ with probability p. We first recall the original switching lemma from [37]:

Theorem 3.1 (Håstad's switching lemma)**.** *Let F be a k-CNF. Then for all $t \geq 1$, we have that*

$$\Pr_{\rho \leftarrow \mathcal{R}_p}[F \upharpoonright \rho \text{ does not have a decision tree of depth } t] \leq (5pk)^t.$$

In the context of AC^0 circuits the switching lemma is used to achieve *depth reduction* under random restrictions: we apply Theorem 3.1 separately to each of the bottom-layer depth-2 subcircuits, choosing t appropriately so that all of them "switch" to depth-t decision trees with high probability. The following corollary is what is typically used:

Corollary 3.2 (AC^0 depth reduction via Theorem 3.1)**.** *Let $\mathcal{C}$ be a size-M depth-d AC^0 circuit with bottom fan-in k, and let $p = 1/(10k)$. Then for all $\varepsilon > 0$,*

$$\Pr_{\rho \leftarrow \mathcal{R}_p}\left[\mathcal{C} \upharpoonright \rho \text{ is not computed by a depth-}(d-1) \text{ circuit with bottom fan-in } \log(M/\varepsilon)\right] \leq \varepsilon.$$

Proof. This follows from applying Theorem 3.1 with $t = \log(M/\varepsilon)$ to each of the bottom-layer depth-2 subcircuits of $\mathcal{C}$ (at most M of them), along with the basic fact that a depth-t decision tree can be expressed as both a t-DNF as well as a t-CNF. $\square$

The same argument is then repeated again on the $(k = \log(M/\varepsilon))$-DNFs at the bottom two layers of the new circuit (applying the dual form of the switching lemma for k-DNFs rather than k-CNFs) to further reduce the depth to $d-2$. However, observe that in this second application of the switching lemma (and in later applications as well), in order to use Corollary 3.2, the parameter p of the random restriction must now depend on ε, since we must now take $p < 1/(5k) = 1/(5\log(M/\varepsilon))$ in order to get a nontrivial bound in Theorem 3.1. This is why standard applications of the [37] switching lemma (involving $d-1$ iterative applications of Corollary 3.2) show that every size-M depth-d AC^0 circuit collapses to depth-$(t = \log(M/\varepsilon))$ decision tree with high probability, at least $1-\varepsilon$, under a random restriction with $*$-probability $p = \Theta(1/\log^{d-1}(M/\varepsilon))$. Note that t and p both depend on ε.

As alluded to above, the recent multi-switching lemma of [38] shows, remarkably, that essentially the same simplification holds under a random restriction with $*$-probability $p = \Theta(1/\log^{d-1}(M))$, independent of ε. Let us establish some terminology and notation to present these results.

Definition 3.3 (Common partial decision tree)**.** Let $\mathscr{F} = \{F_1, \dots, F_M\}$ be a collection of Boolean functions. We say that a decision tree T is a *common ℓ-partial decision tree for $\mathscr{F}$* if every $F_i \in \mathscr{F}$ can be expressed as T with depth-ℓ decision trees at its leaves. (Equivalently, for every $F_i \in \mathscr{F}$ and root-to-leaf path π in T, we have that $F_i \upharpoonright \pi$ is computed by a depth-ℓ decision tree.)

The multi-switching lemma of [38] is as follows:

Theorem 3.4 (Multi-switching lemma, Lemma 3.8 of [38])**.** *Let $\mathscr{F} = \{F_1, \dots, F_M\}$ be a collection of k-CNFs and $\ell := \log(2M)$. Then for all $t \geq 1$,*

$$\Pr_{\rho \leftarrow \mathcal{R}_p}\left[\mathscr{F} \upharpoonright \rho \text{ does not have a common } \ell\text{-partial DT of depth } t\right] \leq M(24pk)^t.$$

The following corollary should be contrasted with Corollary 3.2:

Corollary 3.5 (AC^0 depth reduction via Theorem 3.4; c.f. Corollary 3.2)**.** *Let $\mathcal{C}$ be a size-M depth-d AC^0 circuit with bottom fan-in k, and let $p = 1/(48k)$. Then for all $\varepsilon > 0$,*

$$\Pr_{\rho \leftarrow \mathcal{R}_p}\left[\mathcal{C} \upharpoonright \rho \text{ is not computed by a } ((\log(M/\varepsilon), \mathsf{AC}^0(\text{depth } d-1, \text{ bottom fan-in } \log(2M))\text{-decision tree})\right] \leq \varepsilon.$$

Proof. This follows by applying Theorem 3.4 with $\mathscr{F}$ being the bottom-layer depth-2 subcircuits of $\mathcal{C}$ and $t = \log(M/\varepsilon)$, along with the fact that a depth-ℓ decision tree can be expressed as both a ℓ-DNF and an ℓ-CNF. $\square$

We highlight a crucial qualitative aspect of Corollary 3.5: while the depth $t = \log(M/\varepsilon)$ of the decision tree does depend on ε, the depth-$(d-1)$ AC^0 circuits at its leaves have bottom fan-in $k = \log(2M)$ which does *not* depend on ε. This means that in successive application of Corollary 3.5, the values of $p = 1/(48k) = \Theta(1/\log M)$ will remain independent of ε. This leads to much better quantitative bounds than can be obtained through repeated applications of Corollary 3.2: $d-1$ iterative applications of Corollary 3.5 imply that every size-M depth-d AC^0 circuit collapses to a depth-$O(2^d \log(M/\varepsilon))$ decision tree with high probability, at least $1-\varepsilon$, under a random restriction with $*$-probability $p = \Theta(1/\log^{d-1} M)$. Note that the overall $*$-probability p is independent of ε.

Multi-switching lemmas and sparse $\mathbb{F}_2$ polynomials. The qualitative advantage of multi-switching lemmas—in particular, the crucial role of a common partial decision tree—can also be seen within the context of $\mathbb{F}_2$ polynomials.

Let P be an S-sparse $\mathbb{F}_2$ polynomial. It is an easy observation that P becomes a low-degree polynomial with high probability when hit with a random restriction: for all $\varepsilon, p \in (0,1)$ and $k \in \mathbb{N}$,

$$\mathop{\mathbf{Pr}}_{\rho \leftarrow \mathcal{R}_{\frac{p}{2}}}[P \upharpoonright \rho \text{ is not a degree-}k\text{ polynomial}] \leq \frac{\varepsilon}{2} + S\binom{w}{k}p^k \quad \text{where } w = \Theta(\log(S/\varepsilon)). \tag{3.1}$$

(The proof follows by considering each monomial of P individually and taking a union bound over all S of them. For a fixed monomial, the probability that more than $\Omega(\log(S/\varepsilon))$ variables survive a random restriction from $\mathcal{R}_{\frac{1}{2}}$ is at most $\varepsilon/(2S)$; next, the probability that at least k variables in a width-w monomial survive a random restriction from $\mathcal{R}_p$ is at most $\binom{w}{k}p^k$.) The failure probability of (3.1) can be made at most ε by choosing p and k appropriately, but note that at least one of p (the $*$-probability) or k (the degree of the resulting polynomial) must depend on ε.

Using a slight extension of the ideas in the multi-switching lemmas of [38], we can instead bound the probability that $P \upharpoonright \rho$ becomes a *depth-t decision tree with degree-k polynomials at its leaves*. While this provides weaker structural information than the simple observation above (cf. Corollary 3.2 vs. Corollary 3.5 in the context of AC^0), the crucial win will come from the fact that p and k can *both* be taken to be independent of the failure probability ε (and only t will depend on ε).

3.1 Canonical common ℓ-partial decision trees

An important concept in the proof of Theorem 3.4 is that of a *canonical* common ℓ-partial decision tree for an ordered collection $\mathscr{F}$ of k-CNFs, which we define in this section.

Given a k-CNF formula F (which we view as an ordered sequence of width-k clauses $C_1 \wedge C_2 \wedge \cdots$), we recall the notion of the *canonical decision tree* for F, denoted $\mathrm{CDT}(F)$. This is a decision tree which computes F and is obtained as follows:

- If any clause C_i is identically 0, then the tree is the constant 0.
- If every clause C_i is identically 1, then the tree is the constant 1.
- Otherwise, let C_{i_1} be the first clause that is not identically 1, and let $\kappa \in [k]$ be the number of variables in C_{i_1}. The first κ levels of $\mathrm{CDT}(F)$ exhaustively query these κ variables. At each of the 2^κ resulting leaves of the tree (each one corresponding to some restriction $\eta \in \{0,1\}^\kappa$ fixing those κ variables), recursively put down the canonical decision tree $\mathrm{CDT}(F \upharpoonright \eta)$.

We observe that the tree $\mathrm{CDT}(F)$ is unique given a fixed ordering $C_1, C_2, \ldots$ of the clauses in F.

Håstad's proof of his original switching lemma (Theorem 3.1) actually shows that if F is a k-CNF, then the canonical decision tree $\mathrm{CDT}(F \upharpoonright \rho)$ is shallow w.h.p. over $\rho \leftarrow \mathcal{R}_p$. This is crucially important for the arguments of Trevisan and Xue [73], who give a *derandomized* version of Håstad's original switching lemma: they construct a pseudorandom distribution over restrictions to take the place of $\mathcal{R}_p$, and show that with high probability a restriction drawn from this pseudorandom distribution causes a k-CNF to collapse to a small-depth decision tree. Their argument uses the structure of a canonical decision tree in an essential way.

Turning to Håstad's multi-switching lemma [38], we observe that analogous to his original switching lemma, the proof of Theorem 3.4 given in [38] implicitly establishes a stronger statement: $\mathscr{F} \upharpoonright \rho$ has a small-depth *canonical* common ℓ-partial decision tree w.h.p. over $\rho \leftarrow \mathcal{R}_p$ (we will define the notion of a canonical common ℓ-partial decision tree below). In fact, we will use the fact that it actually establishes an even stronger statement: w.h.p. over $\rho \leftarrow \mathcal{R}_p$, *every* canonical common ℓ-partial decision tree for $\mathscr{F} \upharpoonright \rho$ is shallow—as we explain below, there is more than one canonical common ℓ-partial decision tree for a sequence $\mathscr{F}$ of CNFs.

Let us explain what a canonical common ℓ-partial decision tree for a sequence of CNFs $\mathscr{F}$ is. We will see that there is a set of canonical common ℓ-partial decision trees for a given $\mathscr{F}$ rather than just one tree; and this is the case even though we assume a fixed ordering $F_1, F_2, \dots$ on the elements of $\mathscr{F}$ as well as on the clauses within each CNF. (Observe the contrast with the case of a canonical decision tree for a single formula F, where we assume a fixed ordering on the clauses of F; in that setting, as explained above there is a single canonical decision tree $\mathrm{CDT}(F)$.)

We need a preliminary definition to handle a technical issue related to the final segment of paths through a canonical decision tree.

Definition 3.6 (Full paths in the CDT)**.** Let $F = C_1 \wedge C_2 \wedge \cdots$ be a k-CNF and consider the canonical decision tree $\mathrm{CDT}(F)$ for F. Every path η in $\mathrm{CDT}(F)$ can be written as the the disjoint union of segments $\eta = \eta^{(1)} \circ \eta^{(2)} \circ \cdots \circ \eta^{(u)}$, where for all $j \in [u]$, the segment $\eta^{(j)}$ is an assignment to the surviving variables in the restricted clause $C_{i_j} \upharpoonright \eta^{(1)} \circ \cdots \circ \eta^{(j-1)}$, and C_{i_j} is the first clause in $F \upharpoonright \eta^{(1)} \circ \cdots \circ \eta^{(j-1)}$ that is not identically 1.

Furthermore, note that for $j \in [u-1]$, the segment $\eta^{(j)}$ is in fact an assignment fixing *all* the surviving variables in $C_{i_j} \upharpoonright \eta^{(1)} \circ \cdots \circ \eta^{(j-1)}$. We say that η is *full* if this is also the case for the final segment: η is *full* if $\eta^{(u)}$ is an assignment fixing all the surviving variables in $C_{i_u} \upharpoonright \eta^{(1)} \circ \cdots \circ \eta^{(u-1)}$.

Observation 1. Let F be a k-CNF and suppose $\mathrm{depth}(\mathrm{CDT}(F)) > \ell$. Then there is a full path η of length $|\eta| \in \{\ell+1, \dots, \ell+k\}$ in $\mathrm{CDT}(F)$.

To help minimize confusion, we will reserve "η" for paths or segments of paths in CDTs, and "π" for paths (or segments of paths) in CCDTs.

We are now ready to define the set of canonical common ℓ-partial decision trees:

Definition 3.7 (Canonical common ℓ-partial DT)**.** Let $\mathscr{F} = (F_1, \dots, F_M)$ be an ordered collection of k-CNFs. The set of all *canonical* common ℓ-partial decision trees for $\mathscr{F}$, which we denote $\mathrm{CCDT}_\ell(\mathscr{F})$, is defined inductively as follows:

0. If $M = 0$ (i. e., $\mathscr{F}$ is an empty collection of k-CNFs) then $\mathrm{CCDT}_\ell(\mathscr{F})$ contains a single tree, the empty tree with no nodes. (Note that otherwise $M \geq 1$, so there is some first formula F_1 in $\mathscr{F}$.)

1. If $\mathrm{depth}(\mathrm{CDT}(F_1)) \leq \ell$, then $\mathrm{CCDT}_\ell(\mathscr{F})$ is simply $\mathrm{CCDT}_\ell(\mathscr{F}')$, where $\mathscr{F}' = (F_2, \dots, F_M)$. (Note that in this case, since inductively each tree in $\mathrm{CCDT}_\ell(\mathscr{F}')$ is a common ℓ-partial DT for $\mathscr{F}'$, each such tree is also a common ℓ-partial DT for $\mathscr{F}$.)

2. Otherwise, since $\mathrm{depth}(\mathrm{CDT}(F_1)) > \ell$ there must be a witnessing *full* path η of length between $\ell+1$ and $\ell+k$ in $\mathrm{CDT}(F_1)$, and there are at most $2^{\ell+k}$ such witnessing full paths. Let P be the set of all such witnessing full paths. For each path $\eta \in P$, let T_η be the tree of depth $|\eta|$ obtained by exhaustively querying all the variables in η in the first $|\eta|$ levels. Recurse at the end of each path in T_η: for each path π in T_η, attach a tree T' from $\mathrm{CCDT}_\ell(\mathscr{F} \upharpoonright \pi)$ at the end of the path. So in this case $\mathrm{CCDT}_\ell(\mathscr{F})$ is the set of all trees that can be obtained in this way (across all possible choices of $\eta \in P$ and all possible choices of a tree $T' \in \mathrm{CCDT}_\ell(\mathscr{F} \upharpoonright \pi)$ for each path $\pi \in T_\eta$).

We write $\mathrm{depth}(\mathrm{CCDT}_\ell(\mathscr{F}))$ to denote the maximum depth of any tree in the set $\mathrm{CCDT}_\ell(\mathscr{F})$.

The following slight variant of Theorem 3.4 can be extracted, with some effort, from a slight modification of the proof given in [38], which we provide in Appendix A:

Theorem 3.8 (Slight variant of Håstad's multi-switching lemma. Theorem 3.4)**.** *Let $\mathscr{F} = (F_1, \dots, F_M)$ be an ordered collection of k-CNFs. Then for all $\ell, t \geq 1$,*

$$\Pr_{\rho \leftarrow \mathcal{R}_p}[\mathrm{depth}(\mathrm{CCDT}_\ell(\mathscr{F} \upharpoonright \rho)) \geq t] \leq M^{\lceil t/\ell \rceil}(32pk)^t.$$

A comparison of Theorem 3.4 (Håstad's multi-switching lemma) and Theorem 3.8 (our variant of it). We emphasize that the differences are technical in nature, and all the ideas in our proof of Theorem 3.8 are from [38]. First, we observe that ℓ is now a free parameter rather than being fixed to $\log(2M)$; this flexibility will be necessary in our PRG construction for sparse $\mathbb{F}_2$ polynomials (where we take $\ell = \Theta(\sqrt{\log M})$). Second, our notion of a canonical common partial decision tree differs slightly from the one that is implicit in [38]: in case 2 of Definition 3.7, we query a witnessing full path of length between $\ell+1$ and $\ell+k$, whereas [38] queries any witnessing path of length greater than ℓ.

4 A pseudorandom multi-switching lemma

As suggested earlier, the crux of our PRG construction is a *derandomization* of the multi-switching lemma (Theorem 3.8): we devise a suitable *pseudorandom* distribution over random restrictions in place of $\mathcal{R}_p$ (the truly random distribution over restrictions) and show that a random restriction ρ drawn from this pseudorandom distribution satisfies a gurantee similar to the one in Theorem 3.8.

Our derandomization of Theorem 3.8 is largely influenced by Trevisan and Xue's [73] ingenious derandomization of Håstad's original switching lemma (Theorem 3.1). Roughly speaking, we will derandomize the multi-switching lemma of Theorem 3.8 by "fooling its proof": we will show that the proof of Theorem 3.8 (given in Appendix A, which we again emphasize is only a slight technical modification of Håstad's proof of his multi-switching lemma, Theorem 3.4) "cannot δ-distinguish" between truly random restrictions and pseudorandom restrictions drawn from $\mathrm{polylog}(n)$-wise independent distributions. Since Theorem 3.8 holds for truly random restrictions, it thus follows that it also holds for pseudorandom restrictions drawn from $\mathrm{polylog}(n)$-wise independent distributions (up to a δ additive loss in the failure probability).

To accomplish this, we exploit the "computational simplicity" of Theorem 3.8's proof: for a fixed family $\mathscr{F}$ of k-CNF formulas, we will show that there is a small AC^0 circuit that takes as input an encoding of a restriction ρ, and returns 1 iff ρ is a bad restriction for the desired conclusion of Theorem 3.8, contributing to its failure probability (i. e., iff $\mathrm{depth}(\mathrm{CCDT}_\ell(\mathscr{F} \upharpoonright \rho)) > t$). As alluded to in Section 3.1, this relies on the fact that Theorem 3.8 does not simply bound the depth of the *optimal* common ℓ-partial decision tree for $\mathscr{F} \upharpoonright \rho$, but instead the depth of any *canonical* common ℓ-partial decision tree for $\mathscr{F} \upharpoonright \rho$. Indeed, this "constructive" aspect of the proof is crucial for our derandomization strategy: it is not at all clear that there is a small circuit for checking if the *optimal* common ℓ-partial decision tree for $\mathscr{F} \upharpoonright \rho$ has depth greater than t.

It will be convenient for us to represent restrictions $\rho \in \{0,1,*\}^n$ as bitstrings $(\rho, y) \in \{0,1\}^{n\times q} \times \{0,1\}^n := \{0,1\}^{Y_q}$, where $q \in \mathbb{N}$ is a parameter and $Y_q = (q+1)n$.

Definition 4.1 (Representing restrictions as bitstrings)**.** We associate with each string $(\rho, y) \in \{0,1\}^{Y_q}$ the restriction $\rho(\rho, y) \in \{0,1,*\}^n$ defined as follows:

$$\rho(\rho,y)_i = \begin{cases} * & \text{if } \rho_{i,1} = \cdots = \rho_{i,q} = 1 \\ y_i & \text{otherwise.} \end{cases}$$

The following observation explains the role of q:

Observation 2. Let (ρ, y) be drawn from the uniform distribution over $\{0,1\}^{Y_q}$. Then the random restriction $\rho(\rho, y) \in \{0,1,*\}^n$ is distributed according to $\mathcal{R}_p$ where $p = 2^{-q}$.

Our main result in this section is a pseudorandom multi-switching lemma:

Theorem 4.2 (Derandomized version of Theorem 3.8)**.** *Let $\mathscr{F} = (F_1, \ldots, F_M)$ be an ordered list of Q-clause k-CNFs. Let $\delta, p \in (0,1)$ and define $q = \log(1/p)$. Let $\ell \geq k$ and $t \in \mathbb{N}$, and $\mathcal{D}$ be any distribution over $\{0,1\}^{Y_q}$ that $(\delta/(M^{\lceil t/\ell \rceil} n^{O(t)}))$-fools the class of depth-4 circuits of size $M(n^{O(\ell)} Q^{O(\ell)} 2^{O(kq)})$. Then*

$$\Pr_{(\eta,z)\leftarrow\mathcal{D}}\left[\mathrm{depth}(\mathrm{CCDT}_\ell(\mathscr{F} \upharpoonright \rho(\eta, z))) \geq t\right] \leq 16^{t+\ell} M^{\lceil t/\ell \rceil}(32pk)^t + \delta.$$

4.1 Bad restrictions and the structure of witnessing paths

Fix $\mathscr{F} = (F_1, \ldots, F_M)$. We say that a restriction $\rho \in \{0,1,*\}^n$ is *bad* if

$$\mathrm{depth}(\mathrm{CCDT}_\ell(\mathscr{F} \upharpoonright \rho)) \geq t.$$

Fix ρ to be a bad restriction. Recalling our definition of the set of canonical common partial decision trees (Definition 3.7), there exists a tree $T \in \mathrm{CCDT}_\ell(\mathscr{F} \upharpoonright \rho)$ and a path Π of length exactly t through T. Furthermore, we have that

1. There exist indices $1 \leq i_1 \leq i_2 \leq \cdots \leq i_u \leq M$ where $u \leq \lceil t/\ell \rceil$, and
2. $\Pi = \pi^{(1)} \circ \cdots \circ \pi^{(u)}$, where for all $j \in [u]$, we have that $\mathrm{supp}(\pi^{(j)}) = \mathrm{supp}(\eta^{(j)})$ where $\eta^{(j)}$ is a path through the canonical decision tree

   $$\mathrm{CDT}(F_{i_j} \upharpoonright \rho \circ \pi^{(1)} \circ \cdots \circ \pi^{(j-1)}).$$

   Furthermore, for every $j \in [u-1]$ we have that $\eta^{(j)}$ is a full path of length between $\ell+1$ and $\ell+k$ through the CDT, and $\eta^{(u)}$ is a path of length exactly $t - \sum_{j=1}^{u-1} |\mathrm{supp}(\eta^{(j)})|$. (Note that $\eta^{(u)}$ is not necessarily a full path.)

(Note that by (2), these subpaths $\pi^{(j)}$ of Π are supported on mutually disjoint sets of coordinates.) With this structure of Π in mind, we make the following definition:

Definition 4.3 ($\mathscr{F}$-traversal)**.** Let $\mathscr{F} = (F_1, \dots, F_M)$ be an ordered list of CNFs. An *ℓ-segmented $\mathscr{F}$-traversal of length t* is a tuple $P = (\mathscr{I}, (S_1, \dots, S_u), \Pi, \mathrm{H})$ comprising:

1. An ordered list of indices $\mathscr{I} = (i_1, \dots, i_u)$ where $1 \le i_1 \le \cdots \le i_u \le M$ and $u \le \lceil t/\ell \rceil$,
2. For each index $i_j \in \mathscr{I}$, a subset $S_j \subseteq [n]$ such that
   (a) These sets are mutually disjoint: $S_j \cap S_{j'} = \emptyset$ for all $j \neq j'$.
   (b) For $1 \le j \le u-1$, each S_j has size between $\ell+1$ and $\ell+k$, and S_u has size exactly $t - \sum_{j=1}^{u-1} |\mathrm{supp}(\eta^{(j)})|$.

   (Consequently $|S_1 \cup \cdots \cup S_u| = t$.)
3. An assignment $\Pi = \pi^{(1)} \circ \cdots \circ \pi^{(u)}$ to the variables in $S_1 \cup \cdots \cup S_u$, where
$$\pi^{(j)} : \{0,1\}^{S_j} \to \{0,1\} \qquad \text{for } 1 \le j \le u.$$
4. An assignment $\mathrm{H} = \eta^{(1)} \circ \cdots \circ \eta^{(u)}$ to the variables in $S_1 \cup \cdots \cup S_u$, where again
$$\eta^{(j)} : \{0,1\}^{S_j} \to \{0,1\} \qquad \text{for } 1 \le j \le u.$$

By our discussion above, for any restriction $\rho \in \{0,1,*\}^n$ and any tree $T \in \mathrm{CCDT}_\ell(\mathscr{F} \upharpoonright \rho)$, every path Π of length t through $\mathrm{CCDT}_\ell(\mathscr{F} \upharpoonright \rho)$ uniquely induces an ℓ-segmented $\mathscr{F}$-traversal P of length t. We say that P *occurs in* $\mathrm{CCDT}_\ell(\mathscr{F} \upharpoonright \rho)$ if it is induced by some path Π of length t through T for some $T \in \mathrm{CCDT}_\ell(\mathscr{F} \upharpoonright \rho)$.

Definition 4.3 immediately yields the following:

Proposition 4.4 (Number of $\mathscr{F}$-traversals)**.** *Fix an ordered list $\mathscr{F} = (F_1, \dots, F_M)$ of k-CNFs, and let $\mathcal{P}_{\mathscr{F},\ell,t}$ denote the collection of all ℓ-segmented $\mathscr{F}$-traversals of length t. Then*
$$|\mathcal{P}_{\mathscr{F},\ell,t}| \le M^{\lceil t/\ell \rceil} n^{O(t)}.$$

4.2 A small AC^0 circuit for recognizing bad restrictions

We begin by showing that for every $\mathscr{F}$-traversal $P = (\mathscr{I}, (S_1, \dots, S_u), \Pi, \mathrm{H})$, there is a small circuit $\mathcal{C}_P$ over $\{0,1\}^{Y_q}$ that returns 1 on input $(\rho, y) \in \{0,1\}^{Y_q}$ iff P occurs in $\mathrm{CCDT}_\ell(\mathscr{F} \upharpoonright \rho(\rho, y))$. Since

$$\begin{aligned} \rho(\rho,y) \text{ is bad} &\iff \mathrm{depth}(\mathrm{CCDT}_\ell(\mathscr{F} \upharpoonright \rho(\rho,y))) \ge t \\ &\iff \exists\ \ell\text{-segmented } \mathscr{F}\text{-traversal } P \text{ of length } t \text{ occurring in } \mathrm{CCDT}_\ell(\mathscr{F} \upharpoonright \rho(\rho,y)), \end{aligned}$$

by considering
$$\mathcal{C}_{\mathscr{F},\ell,t}(\rho,y) := \bigvee_{P \in \mathcal{P}_{\mathscr{F},\ell,t}} \mathcal{C}_P(\rho,y) \tag{4.1}$$
we have that
$$\rho(\rho,y) \text{ is bad} \iff \mathcal{C}_{\mathscr{F},\ell,t}(\rho,y) = 1.$$

Claim 4.5 (Circuit for a single $\mathscr{F}$-traversal)**.** *Let $P = (\mathscr{I}, (S_1, \ldots, S_u), \Pi, \mathrm{H})$ be an ℓ-segmented $\mathscr{F}$-traversal of length t. There is a depth-4 AND-OR-AND-OR circuit $\mathcal{C}_P : \{0,1\}^{Y_q} \to \{0,1\}$ of size $M(n^{O(\ell)} Q^{O(\ell)} 2^{O(kq)})$ such that*

$$\forall (\rho, y) \in \{0,1\}^{Y_q}: \quad \mathcal{C}_P(\rho, y) = 1 \Longleftrightarrow P \textit{ occurs in } \mathrm{CCDT}_\ell(\mathscr{F} \upharpoonright \rho(\rho, y))$$

Proof. Our circuit $\mathcal{C}_P$ will be the AND of M many depth-3 subcircuits, one for each k-CNF $F \in \mathscr{F}$. As we will explain later, each of these subcircuits is one of two types. We first describe these two types of "candidate subcircuits," and then explain precisely which M subcircuits of each type are AND-ed together to give $\mathcal{C}_P$. (Both these types of circuits are implicit in the work of [73].)

1. **First type: Circuits checking that a particular restriction η is a segment of a path in a particular CDT.** We claim that for any Q-clause k-CNF $F' = C_1 \wedge \cdots \wedge C_Q$ and restriction η, there is a depth-3 OR-AND-OR circuit G over $\{0,1\}^{Y_q}$ with fan-in sequence $(\binom{Q}{\le|\eta|}, Q2^{O(kq)}, O(kq))$ that returns 1 on input (ρ, y) iff η is the initial segment of a path in $\mathrm{CDT}(F' \upharpoonright \rho(\rho, y))$.

   For each $i \in [Q]$, we write Fixed_i to denote the set

   $$\{j \in [n]\colon\ j \in \eta^{-1}(\{0,1\}) \text{ and } x_j \text{ occurs in } C_i\}$$

   of all variables that are fixed by η and occur in C_i, and we write $\eta^{(i)} \in \{0,1\}^{\mathrm{Fixed}_i}$ to denote η restricted to the coordinates in Fixed_i. We note that η is a path in $\mathrm{CDT}(F' \upharpoonright \rho(\rho, y))$ if and only if there is a set $I \subseteq [Q]$ of indices (indices of the clauses that contribute to η in $\mathrm{CDT}(F' \upharpoonright \rho(\rho, y))$) where $|I| \le |\eta|$, such that for all $i \in Q$:

   (a) If $i \notin I$, the clause C_i is satisfied by the restriction $\rho \circ \eta^{(<i)}$, where $\eta^{(<i)}$ denotes the composition of $\eta^{(i')}$ for all $i' \in I$ such that $i' < i$. (This clause does not contribute to $\mathrm{CDT}(F' \upharpoonright \rho(\rho, y))$; it is "skipped" in the canonical decision tree construction process);

   (b) Otherwise, if $i \in I$ and $i < \max(I)$,

      (i) for every $j \in [n]$ such that x_j occurs in C_i,

      $$\rho(\rho, y)_j \begin{cases} = * & \text{if } j \in \mathrm{Fixed}_i \\ \in \{0,1\} & \text{otherwise;} \end{cases}$$

      (ii) $\rho \circ \eta^{(<i)}$ does not satisfy any literal in C_i (and hence this clauses contributes to $\mathrm{CDT}(F' \upharpoonright \rho(\rho, y))$);

      (iii) $\rho \circ \eta^{(i)}$ satisfies C_i.

   (c) Finally, if $i = \max(I)$,

      (i) for every $j \in [n]$ such that x_j occurs in C_i, we have $\rho(\rho, y)_j = *$ if $j \in \mathrm{Fixed}_i$;

      (ii) $\rho \circ \eta^{(<i)}$ does not satisfy any literal in C_i.

   We now argue that for any fixed set $I \subseteq [Q]$, the above conditions can be checked by a $Q2^{O(kq)}$-clause $O(kq)$-width CNF, from which our overall claim about circuits of the first type follows by taking an OR over all $\binom{Q}{\le|\eta|}$ possibilities for I. All three conditions (a), (b), and (c) depend only on the coordinates of $\rho(\rho,y)$ that occur in C_i; there are at most k such coordinates since C_i has width at most k, and hence these conditions depend on at most $k(q+1)$ coordinates of $(\rho,y)\in\{0,1\}^{Y_q}$. Consequently it is clear that all three conditions can be checked by a $2^{O(kq)}$-clause $O(kq)$-CNF over $\{0,1\}^{Y_q}$. The CNF is simply the AND of all Q many of these CNFs, one for each clause C_i of F', and hence is itself a $Q2^{O(kq)}$-clause $O(kq)$-width CNF.

2. **Second type: Circuits checking that a particular CDT has depth at most ℓ.** Next, we claim that for every Q-clause k-CNF F', there is a depth-3 AND-OR-AND circuit with fan-in sequence $((2n)^{\ell+1}Q^{O(\ell)}, Q2^{O(kq)}, O(kq))$ that returns 1 on input (ρ,y) iff $\mathrm{depth}(\mathrm{CDT}(F'\upharpoonright\rho(\rho,y)))\le\ell$.

   We establish this by showing that there is a depth-3 OR-AND-OR circuit Σ with the claimed fan-in sequence that returns 1 on input (ρ,y) if $\mathrm{depth}(\mathrm{CDT}(F'\upharpoonright\rho(\rho,y)))>\ell$; given such a circuit Σ, the desired AND-OR-AND circuit is obtained by negating Σ and using de Morgan's law. Certainly $\mathrm{depth}(\mathrm{CDT}(F'\upharpoonright\rho(\rho,y)))>\ell$ iff there is a path η of length $\ell+1$ in $\mathrm{CDT}(F'\upharpoonright\rho(\rho,y))$. There are at most $(2n)^{\ell+1}$ many possible paths of length $\ell+1$ (every path is simply an ordered list of literals), and as argued in (1) above, for every such path η there is a depth-3 OR-AND-OR circuit over $\{0,1\}^{Y_q}$ with fan-in sequence $(\binom{Q}{\le\ell+1}, Q2^{O(kq)}, O(kq))$ that checks if η is a path in $\mathrm{CDT}(F'\upharpoonright\rho(\rho,y))$. The overall circuit Σ is simply the OR of at most $(2n)^{\ell+1}$ such circuits, one for each path η.

With these two types of circuits in hand the overall circuit $\mathcal{C}_P$ is now easy to describe. $\mathcal{C}_P$ is the AND of depth-3 subcircuits, one for each k-CNF $F\in\mathscr{F}$:

- For each of the indices $i_j\in\mathscr{I}$, a circuit of the first type that checks that $\eta^{(j)}$ is a path in $\mathrm{CDT}(F_{i_j}\upharpoonright\rho(\rho,y)\circ\pi^{(1)}\circ\cdots\circ\pi^{(j-1)})$ (recall from Definition 4.3 that $\eta^{(j)}$ is H restricted to the variables in S_j). If $i<u$, we also check that $\eta^{(j)}$ is a full path.
- For all other indices $i\in[M]\setminus\mathscr{I}$, defining $i^-=\max\{j\in[u]\colon i_j<i\}$, if $i^-<u$ we include a circuit of the second type that checks that $\mathrm{depth}(\mathrm{CDT}(F_i\upharpoonright\rho(\rho,y)\circ\pi^{(1)}\circ\cdots\circ\pi^{(i^-)}))\le\ell$, where $i^-=\max\{j\in[u]\colon i_j<i\}$.

The bound on the size of this overall circuit follows from a union bound over the sizes of the subcircuits given in (1) and (2) above. $\square$

4.3 Putting the pieces together: Proof of Theorem 4.2

Recalling the definition (4.1) of $\mathcal{C}_{\mathscr{F},\ell,t}$,

$$\mathcal{C}_{\mathscr{F},\ell,t}(\rho,y):=\bigvee_{P\in\mathcal{P}_{\mathscr{F},\ell,t}}\mathcal{C}_P(\rho,y),$$

Proposition 4.4 giving a bound on its top fan-in, and Claim 4.5 giving a bound on the size of its subcircuits, we have shown the following:

Claim 4.6 (Circuit for recognizing bad restrictions)**.** *Let $\mathscr{F} = (F_1, \ldots, F_M)$ be an ordered list of Q-clause k-CNFs, and let $\ell, t \geq 1$. There is a depth-5 circuit $\mathcal{C}_{\mathscr{F},\ell,t}$ over $\{0,1\}^{Y_q}$ such that*

$$\mathcal{C}_{\mathscr{F},\ell,t}(\rho, y) = 1 \quad \Longleftrightarrow \quad \mathrm{depth}(\mathrm{CCDT}_\ell(\mathscr{F} \upharpoonright \rho(\rho, y))) \geq t.$$

This circuit $\mathcal{C}_{\mathscr{F},\ell,t}$ is the OR of $M^{\lceil \ell/t \rceil} n^{O(t)}$ many depth-4 circuits of size $M(n^{O(\ell)} Q^{O(\ell)} 2^{O(kq)})$.

The following observation will be useful for us:

Observation 3. Let $\mathscr{F} = (F_1, \ldots, F_M)$ be an ordered collection of k-CNFs. For $\ell \geq k$, the total number of paths Π such that Π is a path of length exactly t in some tree $T \in \mathrm{CCDT}_\ell(\mathscr{F})$ is at most $(2^{\ell+k} \cdot 2^{\ell+k})^{\lceil t/\ell \rceil} \leq 16^{t+\ell}$. Consequently, if $(\rho, y) \in \{0,1\}^{Y_q}$ is such that $\mathcal{C}_{\mathscr{F},\ell,t}(\rho, y) = 1$, then $\mathcal{C}_P(\rho, y) = 1$ for (at least one) and at most $16^{t+\ell}$ many ℓ-segmented $\mathscr{F}$-traversals P of length t.

Proof. This follows by inspection of the recursive construction of the set $\mathrm{CCDT}_\ell(\mathscr{F})$ of canonical common ℓ-partial decision trees for $\mathscr{F}$. Each time case (2) of the definition is reached, the set P of witnessing full paths has size at most $2^{\ell+k}$, and for each path in P there are at most $2^{\ell+k}$ possible assignments to the variables on the path. Finally, there are at most $\lceil t/\ell \rceil$ levels of recursive calls. □

With Claim 4.6 and Observation 3 in hand, we are now ready to prove our main result of this section (Theorem 4.2), a derandomized version of the multi-switching lemma (Theorem 3.8). We restate Theorem 4.2 here for the reader's convenience:

Theorem 4.7. *Let $\mathscr{F} = (F_1, \ldots, F_M)$ be an ordered list of Q-clause k-CNFs. Let $\delta, p \in (0,1)$ and define $q = \log(1/p)$. Let $\ell \geq k$ and $t \in \mathbb{N}$, and $\mathcal{D}$ be any distribution over $\{0,1\}^{Y_q}$ that $(\delta/(M^{\lceil t/\ell \rceil} n^{O(t)}))$-fools the class of depth-4 circuits of size $M(n^{O(\ell)} Q^{O(\ell)} 2^{O(kq)})$. Then*

$$\Pr_{(\eta,z) \leftarrow \mathcal{D}} \left[\mathrm{depth}(\mathrm{CCDT}_\ell(\mathscr{F} \upharpoonright \rho(\eta, z))) \geq t \right] \leq 16^{t+\ell} M^{\lceil t/\ell \rceil} (32pk)^t + \delta.$$

Proof.

$$
\begin{aligned}
&\mathop{\mathbf{Pr}}_{(\eta,z)\leftarrow\mathcal{D}}\left[\mathrm{depth}(\mathrm{CCDT}_\ell(\mathscr{F}\upharpoonright\rho(\eta,z)))\geq t\right] \\
&= \mathop{\mathbf{E}}_{(\eta,z)\leftarrow\mathcal{D}}\left[\mathcal{C}_{\mathscr{F},\ell,t}(\eta,z)\right] && \text{(Claim 4.6)}\\
&\leq \sum_{P\in\mathcal{P}_{\mathscr{F},\ell,t}} \mathop{\mathbf{E}}_{(\eta,z)\leftarrow\mathcal{D}}\left[\mathcal{C}_P(\eta,z)\right] && \text{(union bound)}\\
&\leq \sum_{P\in\mathcal{P}_{\mathscr{F},\ell,t}} \left(\mathop{\mathbf{E}}_{(\rho,y)\leftarrow\mathcal{U}}[\mathcal{C}_P(\rho,y)] + \frac{\delta}{M^{\lceil t/\ell\rceil}n^{O(t)}}\right) && (\mathcal{D}\ (\delta/(M^{\lceil t/\ell\rceil}n^{O(t)}))\text{-fools } \mathcal{C}_P)\\
&\leq \delta + \mathop{\mathbf{E}}_{(\rho,y)\leftarrow\mathcal{U}}\left[\sum_{P\in\mathcal{P}_{\mathscr{F},\ell,t}}\mathcal{C}_P(\rho,y)\right] && \text{(Proposition 4.4)}\\
&\leq \delta + 16^{t+\ell}\mathop{\mathbf{E}}_{(\rho,y)\leftarrow\mathcal{U}}[\mathcal{C}_{\mathscr{F},\ell,t}(\rho,y)] && \text{(Observation 3)}\\
&= \delta + 16^{t+\ell}\mathop{\mathbf{Pr}}_{(\rho,y)\leftarrow\mathcal{U}}\left[\mathrm{depth}(\mathrm{CCDT}_\ell(\mathscr{F}\upharpoonright\rho(\rho,y)))\geq t\right] && \text{(Claim 4.6)}\\
&= \delta + 16^{t+\ell}\mathop{\mathbf{Pr}}_{\rho\leftarrow\mathcal{R}_p}\left[\mathrm{depth}(\mathrm{CCDT}_\ell(\mathscr{F}\upharpoonright\rho))\geq t\right] && \text{(Observation 2)}\\
&\leq \delta + 16^{t+\ell}M^{\lceil t/\ell\rceil}(32pk)^t. && \text{(Theorem 3.8)}
\end{aligned}
$$

□

5 Applying our pseudorandom multi-switching lemma: the Ajtai–Wigderson framework for PRG constructions

Implicit in the early work of Ajtai–Wigderson [6] giving the first PRG for AC^0 circuits is a powerful, generic framework for constructing PRGs from "pseudorandom simplification lemmas." In this section we give an explicit description of their framework in general terms. Our work shows that this framework is fairly versatile: both our PRGs, for AC^0 circuits and sparse $\mathbb{F}_2$ polynomials, are obtained within it (albeit with specialized pseudorandom simplification lemmas for each class). Variants of these ideas from [6] are also present in the more recent PRG constructions of [35, 44, 63, 73, 55].

- Let $\mathscr{C}$ be the function class of interest, the class for which we would like to design a PRG. For us $\mathscr{C}$ will either be the class of size-M depth-d AC^0 circuits, or the class of S-sparse $\mathbb{F}_2$ polynomials. (Our analysis will assume that $\mathscr{C}$ is closed under restrictions, which holds for natural function classes including our two classes of interest.)
- Let $\mathscr{C}_{\mathrm{simple}}$ be a class of "simple" functions. We will describe the relationship between $\mathscr{C}$ and $\mathscr{C}_{\mathrm{simple}}$ in detail shortly, but we mention here that this approach relies on the simplicity of the functions in $\mathscr{C}_{\mathrm{simple}}$ enabling PRGs of short seed length. For us, when $\mathscr{C}$ is the class of AC^0 circuits, $\mathscr{C}_{\mathrm{simple}}$ will be the class of small-depth decision trees; when $\mathscr{C}$ is the class of sparse $\mathbb{F}_2$ polynomials, $\mathscr{C}_{\mathrm{simple}}$ will be the class of small-depth decision trees with low-degree $\mathbb{F}_2$ polynomials at its leaves. (Note that we do not require that $\mathscr{C}_{\mathrm{simple}}$ be a subclass of $\mathscr{C}$.)

At a high level, the plan is to give a *randomness-efficient* reduction from the task of fooling $\mathscr{C}$ to that of fooling $\mathscr{C}_{\text{simple}}$; we obtain a pseudorandom distribution $\mathcal{D}$ over $\{0,1\}^n$ that fools $\mathscr{C}$ by "pseudorandomly stitching together" independent copies of a pseudorandom distribution $\mathcal{D}_{\text{simple}}$ over $\{0,1\}^{n'}$ that fools $\mathscr{C}_{\text{simple}}$ (for some $n' \le n$). In more detail, the plan is to fool $\mathscr{C}$ recursively in stages, where in each stage we employ two pseudorandom constructs:

1. A PRG for $\mathscr{C}_{\text{simple}}$, and
2. A "pseudorandom $\mathscr{C}$-to-$\mathscr{C}_{\text{simple}}$ simplification lemma."

   Roughly speaking, such a simplification lemma says the following: there is a pseudorandom distribution $\mathcal{R}$ over restrictions such that for all $\mathcal{C} \in \mathscr{C}$, with high probability over $\rho \leftarrow \mathcal{R}$ the randomly restricted function $\mathcal{C} \upharpoonright \rho$ belongs to $\mathscr{C}_{\text{simple}}$. This pseudorandom distribution $\mathcal{R}$ over the space of restrictions $\{0,1,*\}^n$ should have the following structure:

   (a) The set of "live" positions $L \subseteq [n]$ (i. e., the set of $*$'s) can be sampled with seed length s_{SL}. We write $L \leftarrow \mathcal{R}_{\text{stars}}$ to denote a draw from this pseudorandom distribution over subsets of $[n]$.

   (b) Non-live positions $[n] \setminus L$ are filled in independently and uniformly with $\{0,1\}$, and do not count against the seed length s_{SL}. We write $\rho \leftarrow \{0,1\}^{[n]\setminus L}$ to denote a draw of such a restriction.

   We will require each subset $L \in \text{supp}(\mathcal{R}_{\text{stars}})$ to have size at least pn for some not-too-small $p \in (0,1)$ (equivalently, we will require $\mathcal{R}$ to be supported on restrictions that leave at least a p fraction of coordinates unfixed); as we will soon see, this ensures that we "make good progress" in each stage.

   The guarantee that we will require of this pseudorandom $\mathscr{C}$-to-$\mathscr{C}_{\text{simple}}$ simplification lemma is as follows: for every $\mathcal{C} \in \mathscr{C}$,

$$\mathop{\mathbf{E}}_{L \leftarrow \mathcal{R}_{\text{stars}}}\left[\mathop{\mathbf{Pr}}_{\rho \leftarrow \{0,1\}^{[n]\setminus L}}\left[(\mathcal{C} \upharpoonright \rho) \notin \mathscr{C}_{\text{simple}}\right]\right] \le \delta_{\text{SL}}, \tag{5.1}$$

   where the failure probability δ_{SL} is as small as possible.

An aside about applying Theorem 4.2 within this framework. The astute reader may have noticed that our pseudorandom multi-switching lemma (Thereom 4.2) from the previous section is established for a distribution over restrictions that does *not* have the structure prescribed above: rather than a pseudorandom choice of live variables $L \subseteq [n]$ and a fully random choice of bits as values for the non-live variables $[n] \setminus L$, Theorem 4.2 is established for a distribution over restrictions where both choices are pseudorandom. (Recalling Definition 4.1, we see that η in the statement of Theorem 4.2 corresponds to the choice of $L \subseteq [n]$, and z to the choice of bits for the coordinates in $[n] \setminus L$; in the proof of Theorem 4.2 this pair (η, z) is sampled from a single pseudorandom distribution over Y_q.) However, this suggests that Theorem 4.2 is "stronger than it has to be," since it is more randomness efficient than necessary for this application. Indeed, in Proposition 6.2 we formalize this intuition, showing that our proof of Theorem 4.2 also extends to hold for distributions over restrictions with the prescribed structure.

One stage of the PRG construction. Going back to the general framework, we next describe how the two pseudorandom constructs described above—a PRG for $\mathscr{C}_{\text{simple}}$ and a pseudorandom $\mathscr{C}$-to-$\mathscr{C}_{\text{simple}}$ simplification lemma—are employed together within a single stage of the PRG construction for $\mathscr{C}$.

For $L \subseteq [n]$ let us write $\delta(L)$ to denote the probability $\mathbf{Pr}_{\rho \leftarrow \{0,1\}^{[n]\setminus L}}[(\mathcal{C} \upharpoonright \rho) \notin \mathscr{C}_{\text{simple}}]$; by (5.1) we have that $\mathbf{E}_{L \leftarrow \mathcal{R}_{\text{stars}}}[\delta(L)] \leq \delta_{\text{SL}}$. Fix an $L \subseteq [n]$. Let $\mathcal{D}_{\text{simple}}$ be a distribution that δ_{PRG}-fools $\mathscr{C}_{\text{simple}}$, and suppose $\mathcal{D}_{\text{simple}}$ can be sampled with s_{PRG} many random bits. A simple but crucial fact from [6] is the following: the distribution over $\{0,1\}^n$ where

1. The coordinates in $[n] \setminus L$ are filled in with uniform random bits;

2. The coordinates in L are filled in according to the pseudorandom distribution $\mathcal{D}_{\text{simple}}$,

$(\delta(L) + \delta_{\text{PRG}})$-fools $\mathscr{C}$. That is, for all $\mathcal{C} \in \mathscr{C}$,

$$\mathop{\mathbf{E}}_{\substack{x \leftarrow \mathcal{U} \\ y \leftarrow \mathcal{D}_{\text{simple}}}} \left[\mathcal{C}(x_{[n]\setminus L}, y_L)\right] = \mathop{\mathbf{E}}_{x \leftarrow \mathcal{U}} \left[\mathcal{C}(x)\right] \pm (\delta(L) + \delta_{\text{PRG}}).$$

Taking expectations over $L \leftarrow \mathcal{R}_{\text{stars}}$ and using (5.1), we get that

$$\mathop{\mathbf{E}}_{L \leftarrow \mathcal{R}_{\text{stars}}} \left[\mathop{\mathbf{E}}_{\substack{x \leftarrow \mathcal{U} \\ y \leftarrow \mathcal{D}_{\text{simple}}}} \left[\mathcal{C}(x_{[n]\setminus L}, y_L)\right] \right] = \mathop{\mathbf{E}}_{x \leftarrow \mathcal{U}} \left[\mathcal{C}(x)\right] \pm (\delta_{\text{SL}} + \delta_{\text{PRG}}). \tag{5.2}$$

Consider the distribution $\mathcal{R}_{\text{gentle}}$ over the space of restrictions $\{0,1,*\}^n$ defined as follows: to make a draw $\pi \leftarrow \mathcal{R}_{\text{gentle}}$, first make draws $L \leftarrow \mathcal{R}_{\text{stars}}$ and $y \leftarrow \mathcal{D}_{\text{simple}}$, and then output the restriction $\pi \in \{0,1,*\}^n$ where

$$\pi_i = \begin{cases} y_i & \text{if } i \in L \\ * & \text{otherwise.} \end{cases} \qquad \text{for all } i \in [n].$$

In words, π is the restriction that fixes the coordinates in L according to y. With this definition of $\mathcal{R}_{\text{gentle}}$ in hand, we can rewrite (5.2) as

$$\mathop{\mathbf{E}}_{\pi \leftarrow \mathcal{R}_{\text{gentle}}} \left[\mathop{\mathbf{E}}_{x \leftarrow \mathcal{U}} \left[(\mathcal{C} \upharpoonright \pi)(x)\right] \right] = \mathop{\mathbf{E}}_{x \leftarrow \mathcal{U}} \left[\mathcal{C}(x)\right] \pm (\delta_{\text{SL}} + \delta_{\text{PRG}}). \tag{5.3}$$

Note that a draw $\pi \leftarrow \mathcal{R}_{\text{gentle}}$ can be sampled with $s_{\text{SL}} + s_{\text{PRG}}$ random bits. (We need s_{SL} random bits to make a draw $L \leftarrow \mathcal{R}_{\text{stars}}$, and s_{PRG} random bits to make a draw $y \leftarrow \mathcal{D}_{\text{simple}}$.)

We emphasize that the restriction π is supported on L (i. e., $\pi^{-1}(\{0,1\}) = L$), rather than $[n] \setminus L$. For this reason we may view $\mathcal{R}_{\text{gentle}}$ as being "dual" to the distribution $\mathcal{R}$ that yields a $\mathscr{C}$-to-$\mathscr{C}_{\text{simple}}$ simplification lemma: while $\mathcal{R}$ is supported on restrictions that leave at least a p fraction of coordinates unfixed, $\mathcal{R}_{\text{gentle}}$ is supported on restrictions that fix at least a p fraction of coordinates. This explains why, as alluded to above, we require the pseudorandom simplification lemma to be such that every $L \in \text{supp}(\mathcal{R}_{\text{stars}})$ has size at least pn for some not-too-small $p \in (0,1)$.

Fooling $\mathscr{C}$ recursively: the overall PRG construction and its analysis. We have sketched the construction of a distribution $\mathcal{R}_{\text{gentle}}$ over restrictions in $\{0,1,*\}^n$ that preserves $\mathcal{C}$'s bias up to an error of $(\delta_{\text{SL}}+\delta_{\text{PRG}})$ in the sense of (5.3); furthermore, $\mathcal{R}_{\text{gentle}}$ is supported on restrictions that fix at least a p fraction of coordinates. Since an ε-PRG is simply a distribution over assignments in $\{0,1\}^n$ that preserves $\mathcal{C}$'s bias up to an error of ε, we see that we have made a "p-fraction of progress" towards a PRG, while incurring $(\delta_{\text{SL}}+\delta_{\text{PRG}})$ out of the total ε amount of error allowed.

Our PRG construction will recurse on $\mathcal{C}\upharpoonright\pi$ for all $\pi\in\text{supp}(\mathcal{R}_{\text{gentle}})$, all of which are functions over at most $(1-p)n$ variables. (Since $\mathscr{C}$ is closed under restrictions, we note that $\mathcal{C}\upharpoonright\pi$ belongs to $\mathscr{C}$ and so we can indeed apply the same argument recursively.) By fixing at least a p fraction of the remaining coordinates in each stage, we ensure that there are at most $p^{-1}\ln n$ stages in total, after which n coordinates will have been fixed. Hence, as long as

$$\delta_{\text{SL}}+\delta_{\text{PRG}}\le\frac{\varepsilon}{p^{-1}\ln n},$$

i. e., the total error incurred across all stages is at most ε, we will have that the final distribution over $\{0,1\}^n$ does indeed ε-fool $\mathcal{C}$.

As noted above, the seed length required to sample from $\mathcal{R}_{\text{gentle}}$ in each stage is $s_{\text{SL}}+s_{\text{PRG}}$. Since there are at most $p^{-1}\ln n$ stages in total, the overall seed length of this PRG construction is

$$(s_{\text{SL}}+s_{\text{PRG}})\cdot p^{-1}\ln n.$$

The following theorem summarizes the upshot of our discussion in this section:

Theorem 5.1 (PRGs from pseudorandom simplification lemmas; implicit in [6])**.** *Let $\mathscr{C}$ and $\mathscr{C}_{\text{simple}}$ be two function classes over $\{0,1\}^n$, and suppose we have*

1. *A δ_{PRG}-PRG for $\mathscr{C}_{\text{simple}}$ with seed length $s_{\text{PRG}}(\delta_{\text{PRG}})$ for all $\delta_{\text{PRG}}>0$, and*
2. *A pseudorandom $\mathscr{C}$-to-$\mathscr{C}_{\text{simple}}$ simplification lemma with the following parameters: for all $\delta_{\text{SL}}>0$, there is a distribution $\mathcal{R}_{\text{stars}}$ over subsets of $[n]$ such that*
   (a) *A draw $L\leftarrow\mathcal{R}_{\text{stars}}$ can be sampled with $s_{\text{SL}}(\delta_{\text{SL}})$ random bits.*
   (b) *Every $L\in\text{supp}(\mathcal{R}_{\text{stars}})$ satisfies $|L|\ge pn$ for some $p\in(0,1)$.*
   (c) *For all $\mathcal{C}\in\mathscr{C}$, we have that*

$$\mathop{\mathbf{E}}_{L\leftarrow\mathcal{R}_{\text{stars}}}\left[\mathop{\mathbf{Pr}}_{\rho\leftarrow\{0,1\}^{[n]\setminus L}}\left[(\mathcal{C}\upharpoonright\rho)\notin\mathscr{C}_{\text{simple}}\right]\right]\le\delta_{\text{SL}}.$$

Then for all $\varepsilon>0$, there is an ε-PRG for $\mathscr{C}$ with seed length

$$\left(s_{\text{SL}}\left(\frac{\varepsilon p}{2\ln n}\right)+s_{\text{PRG}}\left(\frac{\varepsilon p}{2\ln n}\right)\right)\cdot p^{-1}\ln n.$$

6 Pseudorandom simplification lemmas for AC^0 circuits and sparse $\mathbb{F}_2$ polynomials

In order to apply Theorem 4.2, we need a PRG that can fool depth-4 circuits (to play the role of $\mathcal{D}$ in that theorem). We recall a very recent result of Harsha and Srinivasan giving the first PRG for fooling AC^0 with a seed length whose ε-dependence is $\log(1/\varepsilon)$; we state this result, specialized to the notation of Section 4, below.

Theorem 6.1 ([36])**.** *The class of size-S depth-d circuits over $\{0,1\}^{Y_q}$ is δ-fooled by r_{HS}-wise independence where*

$$r_{\mathrm{HS}}(S,d,\delta) = \log^{3d+O(1)}(S)\cdot\log(1/\delta).$$

We will need an elementary fact that states, roughly speaking, that if $\mathcal{D}$ is a distribution that fools a class $\mathscr{F}$ that is closed under restrictions, then the distribution obtained by replacing a subset of its coordinates with fully random bits also fools $\mathscr{F}$. Specialized to our context, we state this fact as follows:

Proposition 6.2. *Let $\mathcal{D}_{r\text{-wise}}$ be an r_{HS}-wise independent distribution over $\{0,1\}^{Y_q}$ where $r_{\mathrm{HS}}(S,d,\delta)$ is as defined in Theorem 6.1. Consider the distribution $\mathcal{D}_{\mathrm{mix}}$ over $\{0,1\}^{Y_q}$ where a draw from $\mathcal{D}_{\mathrm{mix}}$ is $(\eta,y)\in\{0,1\}^{n\times q}\times\{0,1\}^n$ where*

1. *(Pseudorandom stars) η is drawn from the marginal distribution of $\mathcal{D}_{r\text{-wise}}$ on $\{0,1\}^{n\times q}$, and*
2. *(Non-stars filled in fully randomly) y is an independent uniform string drawn from $\{0,1\}^n$.*

Then like $\mathcal{D}_{r\text{-wise}}$, this distribution $\mathcal{D}_{\mathrm{mix}}$ also δ-fools the class of size-S depth-d circuits over $\{0,1\}^{Y_q}$.

Proof. This follows from the observation that $\mathcal{D}_{\mathrm{mix}}$, like $\mathcal{D}_{r\text{-wise}}$, is r_{HS}-wise independent (or from the simple argument that gives Fact 9 of [73]). □

We can now state the pseudorandom multi-switching lemma that we will use for both our pseudorandom simplification lemmas (for AC^0 circuits and for sparse $\mathbb{F}_2$ polynomials):

Lemma 6.3 (Stars chosen pseudorandomly, non-stars filled in fully randomly)**.** *Let $\mathscr{F}=(F_1,\dots,F_M)$ be an ordered list of Q-clause k-CNFs. Let $\ell\ge k$, $t\in\mathbb{N}$ and $\delta,p\in(0,1)$, and define $q=\log(1/p)$. There is a distribution $\mathcal{R}_{\mathrm{stars}}$ over subsets of $[n]$ such that the following hold:*

1. *A draw $L\leftarrow\mathcal{R}_{\mathrm{stars}}$ can be sampled with $O(r\log(n))$ random bits, where*
$$r = r_{\mathrm{HS}}\left(M\big(n^{O(\ell)}Q^{O(\ell)}2^{O(kq)}\big),4,\frac{\delta}{M^{\lceil t/\ell\rceil}n^{O(t)}}\right)$$
and $r_{\mathrm{HS}}(\cdot,\cdot,\cdot)$ is as defined in Theorem 6.1.
2. *$\mathcal{R}_{\mathrm{stars}}$ is p-regular: $\mathbf{Pr}_{L\leftarrow\mathcal{R}_{\mathrm{stars}}}\big[i\in L\big]=p$ for all $i\in[n]$.*
3. *A multi-switching lemma holds with respect to $\mathcal{R}_{\mathrm{stars}}$:*
$$\mathop{\mathbf{Pr}}_{\substack{L\leftarrow\mathcal{R}_{\mathrm{stars}}\\ \rho\leftarrow\{0,1\}^{[n]\setminus L}}}\big[\mathrm{depth}(\mathrm{CCDT}_\ell(\mathscr{F}\upharpoonright\rho))\ge t\big]\le 16^{t+\ell}M^{\lceil t/\ell\rceil}(32pk)^t+\delta. \tag{6.1}$$

Proof. Let $\mathcal{D}$ be an r-wise independent distribution over $\{0,1\}^{Y_q}$; standard constructions [7] show that $\mathcal{D}$ can be sampled with $O(r\log|Y_q|) = O(r\log(n))$ random bits. The marginal of $\mathcal{D}$ on $\{0,1\}^{n\times q}$ naturally induces a distribution $\mathcal{R}_{\text{stars}}$ over subsets of $[n]$ via Definition 4.1, where a draw $L \leftarrow \mathcal{R}_{\text{stars}}$ is defined to be $\rho(\rho,z)^{-1}(*)$ (i. e., for all coordinates $i \in [n]$, $i \in L$ iff $\rho_{i,1} = \rho_{i,2} = \cdots = \rho_{i,q} = 1$). Since $\mathcal{D}$ is r-wise independent for $r \gg q$, we have that

$$\Pr_{L\leftarrow\mathcal{R}_{\text{stars}}}\left[i \in L\right] = \Pr_{(\rho,z)\leftarrow\mathcal{D}}\left[\rho_{i,1} = \rho_{i,2} = \cdots = \rho_{i,q} = 1\right] = 2^{-q} = p,$$

which establishes the second claim. The third claim follows by combining Theorem 4.2, Theorem 6.1, and Proposition 6.2. $\square$

6.1 Pseudorandom simplification lemma for AC^0 circuits

We will use the following instantiation of Lemma 6.3 in our construction of a PRG for AC^0 circuits:

Corollary 6.4. *There is a universal constant $c > 0$ such that the following holds. Let $\mathscr{F} = (F_1, \dots, F_M)$ be an ordered list of Q-clause k-CNFs with $\log M \geq k$ and $\varepsilon_0 \in (0,1)$. There is a distribution $\mathcal{R}_{\text{stars}}$ over subsets of $[n]$ such that:*

1. *A draw $L \leftarrow \mathcal{R}_{\text{stars}}$ can be sampled with $s = \log^c(MQ)\log(1/\varepsilon_0)$ random bits.*
2. *$\mathcal{R}_{\text{stars}}$ is p-regular for $p = \Omega(1/k)$.*
3. *A multi-switching lemma holds with respect to $\mathcal{R}_{\text{stars}}$:*

$$\Pr_{\substack{L\leftarrow\mathcal{R}_{\text{stars}}\\ \rho\leftarrow\{0,1\}^{[n]\setminus L}}}\left[\operatorname{depth}(\mathrm{CCDT}_{\log M}(\mathscr{F}\upharpoonright\rho)) \geq \log(2M^5/\varepsilon_0)\right] \leq \varepsilon_0.$$

Proof. Applying Lemma 6.3 with $\ell = \log M$, we see that the failure probability (6.1) can be bounded by

$$16^{t+\ell}M^{\lceil t/\ell\rceil}(32pk)^t + \delta \leq 16^t M^4 M^{\lceil t/\ell\rceil}(64)^{-t} + \delta < M^5 2^{-t} + \delta$$

by choosing $p = \Omega(1/k)$. We make this at most ε_0 by choosing $t = \log(2M^5/\varepsilon_0)$ and $\delta = \varepsilon_0/2$. The bound on s follows from the $d = 3$ case of Theorem 6.1 and our setting of parameters, and this completes the proof. $\square$

Following the standard bottom-up approach to AC^0 circuit lower bounds, we compose $d-1$ iterative applications of the pseudorandom multi-switching lemma of Corollary 6.4 to obtain our pseudorandom simplification lemma for AC^0:

Lemma 6.5 (Pseudorandom simplification lemma for AC^0)**.** *There is a universal constant $C > 0$ such that the following holds. Let $\mathcal{C}$ be a size-M depth-d Boolean circuit over $\{0,1\}^n$ (so recall that $M \geq n$) and $\varepsilon_1 \in (0,1)$. There is a distribution $\mathcal{R}_{\text{stars}}$ over subsets of $[n]$ such that*

1. *A draw $L \leftarrow \mathcal{R}_{\text{stars}}$ can be sampled with $s = O(2^d\log^C(M)\log(1/\varepsilon_1))$ random bits.*

2. *$\mathcal{R}_{\text{stars}}$ is p-regular for $p = \Omega(1/\log^{d-1}(M))$.*

3. *The following simplification lemma holds with respect to $\mathcal{R}_{\text{stars}}$:*

$$\Pr_{\substack{L \leftarrow \mathcal{R}_{\text{stars}} \\ \rho \leftarrow \{0,1\}^{[n]\setminus L}}} \left[\mathcal{C} \upharpoonright \rho \textit{ is not a decision tree of depth } O(2^d \log(M/\varepsilon_1)) \right] \le \varepsilon_1.$$

Proof. Fix $t := \log(2dM^5/\varepsilon_1)$.

Preprocessing stage: We begin with a zeroth stage of preprocessing to trim the bottom fan-in of $\mathcal{C}$: applying Corollary 6.4 with $\mathscr{F}$ being the bottom layer gates of $\mathcal{C}$ (viewed as depth-2 circuits of size $Q \le n$ and bottom fan-in $k = 1$) and $\varepsilon_0 = \varepsilon_1/d$, we get that there is a distribution $\mathcal{R}^{(0)}_{\text{stars}}$ such that $\mathcal{R}^{(0)}_{\text{stars}}$ can be sampled with $s_0 := \log^c(Mn)\log(d/\varepsilon_1)$ random bits (where c is the universal constant from Corollary 6.4), $\mathcal{R}^{(0)}_{\text{stars}}$ is p_0-regular for $p_0 = \Omega(1)$, and

$$\Pr_{\substack{L \leftarrow \mathcal{R}^{(0)}_{\text{stars}} \\ \rho^{(0)} \leftarrow \{0,1\}^{[n]\setminus L}}} \left[\mathcal{C} \upharpoonright \rho^{(0)} \text{ is not a } (t, \mathsf{AC}^0(\text{depth } d, \text{ bottom fan-in } \log M))\text{-decision tree} \right] \le \frac{\varepsilon_1}{d}.$$

First stage: Let $T^{(0)}$ be any good outcome of the zeroth stage above, a $(t, \mathsf{AC}^0(\text{depth } d$, bottom fan-in $\log M))$-decision tree. Note that there are at most 2^t many $\mathsf{AC}^0(\text{depth } d$, bottom fan-in $\log M)$ circuits at the leaves of this depth-t decision tree $T^{(0)}$, each of size at most M. Fix any such circuit $\mathcal{C}'$. Applying Corollary 6.4 to $\mathcal{C}'$, with $\mathscr{F}$ being all its bottom layer depth-2 subcircuits of bottom fan-in $\log M$ (so $Q \le M$) and $\varepsilon_0 = \varepsilon_1/(d2^t)$, we get that there is a distribution $\mathcal{R}^{(1)}_{\text{stars}}$ such that $\mathcal{R}^{(1)}_{\text{stars}}$ can be sampled with $s_1 := \log^c(M^2)\log(d2^t/\varepsilon_1)$ random bits, $\mathcal{R}^{(1)}_{\text{stars}}$ is p_1-regular for $p_1 = \Omega(1/\log M)$, and

$$\Pr_{\substack{L \leftarrow \mathcal{R}^{(1)}_{\text{stars}} \\ \rho^{(1)} \leftarrow \{0,1\}^{[n]\setminus L}}} \left[\mathcal{C}' \upharpoonright \rho^{(1)} \text{ is not a } (2t, \mathsf{AC}^0(\text{depth } d-1, \text{ bottom fan-in } \log M))\text{-decision tree} \right] \le \frac{\varepsilon_1}{d\,2^t}.$$

Taking a union bound over all the circuits at the leaves of $T^{(0)}$ (at most 2^t of them), we get that

$$\Pr_{\substack{L \leftarrow \mathcal{R}^{(1)}_{\text{stars}} \\ \rho^{(1)} \leftarrow \{0,1\}^{[n]\setminus L}}} \left[T^{(0)} \upharpoonright \rho^{(1)} \text{ is not a } (t+2t, \mathsf{AC}^0(\text{depth } d-1, \text{ bottom fan-in } \log M))\text{-decision tree} \right] \le \frac{\varepsilon_1}{d}.$$

Let $T^{(1)}$ be any good outcome of the above, and consider any circuit $\mathcal{C}''$ at a leaf of this depth-$3t$ decision tree. We note a subtlety at this point (this same subtlety is present in applications of the standard switching lemma): while $\mathcal{C}''$ has at most M gates in total from levels 1 to $d-2$ (indeed, its number of gates in those layers is at most that of $\mathcal{C}$), each of its bottom layer depth-2 subcircuits may have size as large as M^2. This is because the M-way AND of depth-$(\log M)$ decision trees, when expressed as depth-2 circuit, can have size as large as $M \cdot 2^{\log M} = M^2$. (And of course the same is true for the M-way OR.) Therefore from the second stage onwards, we will always apply Corollary 6.4 with $\mathscr{F}$ being a family of M many M^2-clause $(\log M)$-CNFs (or DNFs), and so $Q = M^2$.

The i-th stage: We repeat for $d-2$ more stages, where in the i-th stage we consider a good outcome $T^{(i-1)}$ of the previous stage, a $((2^i-1)t, \mathsf{AC}^0(\text{depth } d-i+1, \text{bottom fan-in } \log M))$-decision tree. Fix any subcircuit $\mathcal{C}'''$ of at a leaf of this depth-$((2^i-1)t)$ decision tree $T^{(i-1)}$. Applying Corollary 6.4 to $\mathcal{C}'''$, with $\mathscr{F}$ being all its bottom layer depth-2 subcircuits of bottom fan-in $\log M$ (as noted above, we take $Q = M^2$) and

$$\varepsilon_0 = \frac{\varepsilon_1}{d\, 2^{(2^i-1)t}},$$

we get that there is a distribution $\mathcal{D}^{(i)}_{\text{stars}}$ such that $\mathcal{R}^{(i)}_{\text{stars}}$ can be sampled with

$$s_i := \log^c(M^3)\log(1/\varepsilon_0) = 2^i \cdot O(t\log^c(M))$$

random bits, $\mathcal{R}^{(i)}_{\text{stars}}$ is p_i-regular for $p_i = \Omega(1/\log M)$, and

$$\mathop{\mathbf{Pr}}_{\substack{L\leftarrow\mathcal{R}^{(i)}_{\text{stars}} \\ \rho^{(i)}\leftarrow\{0,1\}^{[n]\setminus L}}}\left[\mathcal{C}''' \upharpoonright \rho^{(i)} \text{ is not a } (2^i t, \mathsf{AC}^0(\text{depth } d-i, \text{ bottom fan-in } \log M))\text{-decision tree}\right] \le \frac{\varepsilon_1}{d\, 2^{(2^i-1)t}}.$$

(We have used the fact that $\log(2M^5/\varepsilon_0) = (2^i-1)t + \log(2dM^5/\varepsilon_1) = 2^i t$.) Taking a union bound over all the circuits at the leaves of $T^{(i-1)}$ (at most $2^{(2^i-1)t}$ of them), we get that

$$\mathop{\mathbf{Pr}}_{\substack{L\leftarrow\mathcal{R}^{(i)}_{\text{stars}} \\ \rho^{(i)}\leftarrow\{0,1\}^{[n]\setminus L}}}\left[T^{(i-1)} \upharpoonright \rho^{(i)} \text{ is not a } ((2^{(i+1)}-1)t, \mathsf{AC}^0(\text{depth } d-i, \text{ bottom fan-in } \log M))\text{-decision tree}\right] \le \frac{\varepsilon_1}{d}.$$

The overall distribution. Composing all d stages described above (including the zeroth preprocessing stage), we get an overall distribution $\mathcal{R}_{\text{stars}}$ where a draw $L \leftarrow \mathcal{R}_{\text{stars}}$ is simply

$$L = L^{(0)} \cap L^{(1)} \cap \cdots \cap L^{(d-1)}, \quad L^{(i)} \leftarrow \mathcal{R}^{(i)}_{\text{stars}} \text{ for all } 0 \le i \le d-1.$$

This distribution $\mathcal{R}_{\text{stars}}$ can be sampled with

$$\sum_{i=0}^{d-1} s_i = O(2^d \log^C(M)\log(1/\varepsilon_1))$$

random bits for some constant $C > 0$, $\mathcal{R}_{\text{stars}}$ is p-regular for

$$p = \prod_{i=0}^{d-1} p_i = (\Omega(1/\log(M)))^{d-1},$$

and by a union bound over the d many failure probabilities of ε_1/d from each of the d stages, we have that indeed

$$\mathop{\mathbf{Pr}}_{\substack{L\leftarrow\mathcal{R}_{\text{stars}} \\ \rho\leftarrow\{0,1\}^{[n]\setminus L}}}\left[\mathcal{C} \upharpoonright \rho \text{ is not a depth-}((2^d-1)t) \text{ decision tree}\right] \le \varepsilon_1.$$

Since $(2^d-1)t = O(2^d \log(M/\varepsilon_1))$ (using $d \le M$ so $\log(2dM^5/\varepsilon_1) \le \log(2M^6/\varepsilon_1)$), this completes the proof. $\square$

6.2 Pseudorandom simplification lemma for sparse $\mathbb{F}_2$ polynomials

To motivate the parameter settings used in this subsection, we recall the discussion about multi-switching lemmas and sparse $\mathbb{F}_2$ polynomials right before Section 3.1; observe that both the $*$-probability p and the degree of the $\mathbb{F}_2$ polynomials obtained below are independent of the failure probability ε_2.

Lemma 6.6 (Pseudorandom simplification lemma for sparse $\mathbb{F}_2$ polynomials)**.** *There is a universal constant $C > 0$ such that the following holds. Let P be an S-sparse $\mathbb{F}_2$ polynomial and $\varepsilon_2 \in (0,1)$. There is a distribution $\mathcal{R}_{\text{stars}}$ over subsets of $[n]$ such that*

1. *A draw $L \leftarrow \mathcal{R}_{\text{stars}}$ can be sampled with $s = \log^C(Sn)\log(1/\varepsilon_2)$ random bits.*
2. *$\mathcal{R}_{\text{stars}}$ is p-regular for $p = 2^{-O(\sqrt{\log S})}$.*
3. *The following simplification lemma holds with respect to $\mathcal{R}_{\text{stars}}$:*

$$\Pr_{\substack{L\leftarrow\mathcal{R}_{\text{stars}}\\ \rho\leftarrow\{0,1\}^{[n]\setminus L}}}\left[P\restriction\rho \textit{ is not a } \left(O(\sqrt{\log S})+\log(2/\varepsilon_2), \mathbb{F}_2(\textit{degree } \sqrt{\log S})\right)\textit{-decision tree}\right] \le \varepsilon_2.$$

Proof. We observe that an S-sparse $\mathbb{F}_2$ polynomial is simply a $\mathsf{PAR}\circ\mathsf{AND}$ circuit with S many bottom layer gates of unbounded fan-in. With this point of view in mind, we apply Lemma 6.3 with $\mathscr{F}$ being this family of S many AND gates (viewed as depth-2 circuits of size $Q \le n$ and bottom fan-in $k = 1$) and $\ell = \sqrt{\log S}$. By choosing $t = A\cdot\sqrt{\log S}+\log(2/\varepsilon_2)$, $p = 2^{-B\sqrt{\log S}}$, and $\delta = \varepsilon_2/2$, we get that the failure probability (6.1) can be bounded by

$$\begin{aligned}16^{t+\ell}S^{\lceil t/\ell\rceil}(32pk)^t+\delta &= 16^{(A+1)\sqrt{\log S}+\log(2/\varepsilon_2)}\cdot S^{1+A}\cdot 2^{\sqrt{\log S}\cdot\log(2/\varepsilon_2)}\cdot\frac{32^{A\cdot\sqrt{\log S}+\log(2/\varepsilon_2)}}{S^{AB}\cdot 2^{B\sqrt{\log S}\cdot\log(2/\varepsilon_2)}}+\frac{\varepsilon_2}{2}\\ &< \varepsilon_2,\end{aligned}$$

where the inequality holds for a suitable choice of absolute constant values A,B. The bound on s follows from the $d = 4$ case of Theorem 6.1 and our setting of parameters, and this completes the proof. ☐

7 PRGs for AC^0 and sparse $\mathbb{F}_2$ polynomials from pseudorandom simplification lemmas

We will need the following easy fact for both our PRG constructions: we can derive from a p-regular distribution $\mathcal{R}_{\text{stars}}$ satisfying a pseudorandom simplification lemma (in the sense of our main results in the previous section, Lemmas 6.5 and 6.6) a distribution $\mathcal{R}'_{\text{stars}}$ supported entirely on sets of size at least $(pn)/2$, such that $\mathcal{R}'_{\text{stars}}$ also satisfies a pseudorandom simplification lemma with only a slightly worse failure probability. More precisely, and in more generality:

Proposition 7.1 (Condition on having sufficiently many stars)**.** *Fix any property $\Phi : \{0,1,*\}^n \to \{0,1\}$ of restrictions. Let $\mathcal{R}_{\text{stars}}$ be a p-regular distribution over subsets of $[n]$ and suppose*

$$\Pr_{\substack{L\leftarrow\mathcal{R}_{\text{stars}}\\ \rho\leftarrow\{0,1\}^{[n]\setminus L}}}\left[\Phi(\rho)=1\right] \le \tau$$

for some $\tau > 0$. Let $\mathcal{R}'_{\text{stars}}$ be the distribution of $L \leftarrow \mathcal{R}_{\text{stars}}$ conditioned on L satisfying $|L| \geq (pn)/2$. Then

$$\Pr_{\substack{L \leftarrow \mathcal{R}'_{\text{stars}} \\ \rho \leftarrow \{0,1\}^{[n] \setminus L}}} \left[\Phi(\rho) = 1\right] \leq \frac{2\tau}{p}.$$

Proof. Since $\mathcal{R}_{\text{stars}}$ is p-regular we have that $\mathbf{E}_{L \leftarrow \mathcal{R}_{\text{stars}}}\left[|L|\right] = pn$, and so

$$\Pr_{L \leftarrow \mathcal{R}_{\text{stars}}}\left[L \in \text{supp}(\mathcal{R}'_{\text{stars}})\right] = \Pr_{L \leftarrow \mathcal{R}_{\text{stars}}}\left[|L| \geq \frac{pn}{2}\right] \geq \frac{p}{2}.$$

Hence

$$\begin{aligned} \Pr_{\substack{L \leftarrow \mathcal{R}'_{\text{stars}} \\ \rho \leftarrow \{0,1\}^{[n] \setminus L}}} \left[\Phi(\rho) = 1\right] &= \Pr_{\substack{L \leftarrow \mathcal{R}_{\text{stars}} \\ \rho \leftarrow \{0,1\}^{[n] \setminus L}}} \left[\Phi(\rho) = 1 \mid L \in \text{supp}(\mathcal{R}'_{\text{stars}})\right] \\ &\leq \Pr_{\substack{L \leftarrow \mathcal{R}_{\text{stars}} \\ \rho \leftarrow \{0,1\}^{[n] \setminus L}}} \left[\Phi(\rho) = 1\right] \cdot \frac{1}{\mathbf{Pr}_{L \leftarrow \mathcal{R}_{\text{stars}}}[L \in \text{supp}(\mathcal{R}'_{\text{stars}})]} \leq \frac{2\tau}{p}. \end{aligned}$$

□

7.1 PRGs for AC^0 circuits

Theorem 7.2. *For every $d \geq 2$, $M \geq n$, and $\varepsilon > 0$, there is an ε-PRG for the class $\mathscr{C}$ of n-variable size-M depth-d circuits with seed length $2^d \log^{d+O(1)}(M) \log(1/\varepsilon)$.*

Proof. Applying Proposition 7.1 to the pseudorandom simplification lemma of Lemma 6.5, we get that for all $\varepsilon_1 > 0$, there is a distribution $\mathcal{R}'_{\text{stars}}$ over subsets of $[n]$ such that

1. A draw $L \leftarrow \mathcal{R}'_{\text{stars}}$ can be sampled with $s_{\text{SL}} = O(2^d \log^C(M) \log(1/\varepsilon_1)$ random bits, where $C > 0$ is the universal constant from Lemma 6.5.

2. Every $L \in \text{supp}(\mathcal{R}'_{\text{stars}})$ satisfies $|L| \geq pn$ where $p = \Omega(1/\log^{d-1}(M))$.

3. For all $\mathcal{C} \in \mathscr{C}$,

$$\Pr_{\substack{L \leftarrow \mathcal{R}'_{\text{stars}} \\ \rho \leftarrow \{0,1\}^{[n] \setminus L}}} \left[\, \mathcal{C} \upharpoonright \rho \text{ is not a decision tree of depth } O(2^d \log(M/\varepsilon_1)) \,\right] \leq \frac{\varepsilon_1}{p}.$$

Setting $\varepsilon_1 = \varepsilon p^2/(2 \ln n)$ and taking $\mathscr{C}_{\text{simple}}$ to be the class of depth-t decision trees where

$$t = O(2^d \log(M/\varepsilon_1)) = O(d\, 2^d \log(M/\varepsilon)),$$

we get that a draw $L \leftarrow \mathcal{R}'_{\text{stars}}$ can be sampled with

$$s_{\text{SL}} = O(2^d \log^C(M) \log(1/\varepsilon_1)) = O(d 2^d \log^C(M) \log((\log M)/\varepsilon))$$

random bits, and $\mathcal{R}'_{\text{stars}}$ satisfies

$$\mathop{\mathbf{E}}_{L\leftarrow\mathcal{R}'_{\text{stars}}}\left[\mathop{\mathbf{Pr}}_{\rho\leftarrow\{0,1\}^{[n]\setminus L}}\left[(\mathcal{C}\upharpoonright\rho)\notin\mathscr{C}_{\text{simple}}\right]\right]\leq\frac{\varepsilon p}{2\ln n}$$

for all $\mathcal{C}\in\mathscr{C}$. Since $\mathscr{C}_{\text{simple}}$ is 0-fooled by any t-wise independent distribution, we get from Theorem 5.1 that there is an ε-PRG for $\mathscr{C}$ with seed length

$$O(s_{\text{SL}}+t\log(n))\cdot p^{-1}\ln n=2^d\log^{d+O(1)}(M)\log(1/\varepsilon),$$

and this completes the proof. □

7.2 PRGs for sparse $\mathbb{F}_2$ polynomials

Theorem 7.3. *For every $S=2^{\omega(\log\log(n))^2}$ and $\varepsilon>0$ there is an ε-PRG for the class $\mathscr{C}$ of n-variable S-sparse $\mathbb{F}_2$ polynomials with seed length $2^{O(\sqrt{\log S})}\log(1/\varepsilon)$.*

Proof. Applying Proposition 7.1 to the pseudorandom simplification lemma of Lemma 6.6, we get that for all $\varepsilon_2>0$, there is a distribution $\mathcal{R}'_{\text{stars}}$ over subsets of $[n]$ such that

1. A draw $L\leftarrow\mathcal{R}'_{\text{stars}}$ can be sampled with $s_{\text{SL}}=\log^C(Sn)\log(1/\varepsilon_2)$ random bits, where $C>0$ is the universal constant from Lemma 6.6.
2. Every $L\in\text{supp}(\mathcal{R}'_{\text{stars}})$ satisfies $|L|\geq pn$ where $p=2^{-O(\sqrt{\log S})}$.
3. For all $P\in\mathscr{C}$,

$$\mathop{\mathbf{Pr}}_{\substack{L\leftarrow\mathcal{R}'_{\text{stars}}\\ \rho\leftarrow\{0,1\}^{[n]\setminus L}}}\left[P\upharpoonright\rho\text{ is not a }\big(O(\sqrt{\log S})+\log(2/\varepsilon_2),\mathbb{F}_2(\text{degree }\sqrt{\log S})\big)\text{-decision tree}\right]\leq\frac{\varepsilon_2}{p}.$$

Setting $\varepsilon_2=\varepsilon p^2/(2\ln n)$ and taking $\mathscr{C}_{\text{simple}}$ to be the class of $(t,\mathbb{F}_2(\text{degree }\sqrt{\log S}))$-decision trees where

$$t=O(\sqrt{\log S})+\log(2/\varepsilon_2)=O(\sqrt{\log S})+\log(1/\varepsilon)$$

(where the second equality uses $S=2^{\omega(\log\log(n))^2}$), we get that a draw $L\leftarrow\mathcal{R}'_{\text{stars}}$ can be sampled with

$$s_{\text{SL}}=\log^C(Sn)\log(1/\varepsilon_2)=O\big(\log^{C+\frac{1}{2}}(Sn)\log(1/\varepsilon)\big)$$

random bits, and $\mathcal{R}'_{\text{stars}}$ satisfies

$$\mathop{\mathbf{E}}_{L\leftarrow\mathcal{R}_{\text{stars}}}\left[\mathop{\mathbf{Pr}}_{\rho\leftarrow\{0,1\}^{[n]\setminus L}}\left[(P\upharpoonright\rho)\notin\mathscr{C}_{\text{simple}}\right]\right]\leq\frac{\varepsilon p}{2\ln n}$$

for all $P\in\mathscr{C}$.

We claim that the class of $(t,\mathbb{F}_2(\text{degree }k))$-decision trees can be δ-fooled with seed length

$$s_{\text{PRG}}(\delta)=k\cdot O(t+2^k\log(1/\delta))+O(t\log(n));$$

we defer the proof of this claim to the next subsection (see Lemma 7.8). Recalling our definition of $\mathscr{C}_{\text{simple}}$ where $t = O(\sqrt{\log S}) + \log(1/\varepsilon)$ and $k = \sqrt{\log S}$, it follows from this claim that $\mathscr{C}_{\text{simple}}$ can be $(\varepsilon p/(2\ln n))$-fooled with seed length

$$\begin{aligned} s_{\text{PRG}} &= 2^{O(\sqrt{\log S})}\log(1/\varepsilon) + O(t\log(n)) = 2^{O(\sqrt{\log S})}\log(1/\varepsilon) + O(\sqrt{\log S}\log(n)) + \log(1/\varepsilon)\log(n) \\ &= 2^{O(\sqrt{\log S})}\log(1/\varepsilon) \end{aligned}$$

(where we have again used $S = 2^{\omega(\log\log(n))^2}$). Now applying Theorem 5.1, we get that there is an ε-PRG for $\mathscr{C}$ with seed length

$$\begin{aligned} (s_{\text{SL}} + s_{\text{PRG}}) \cdot p^{-1}\ln n &= \Big(O\big(\log^{C+\frac{1}{2}}(Sn)\log(1/\varepsilon)\big) + 2^{O(\sqrt{\log S})}\log(1/\varepsilon)\Big) \cdot 2^{O(\sqrt{\log S})}\ln n \\ &= 2^{O(\sqrt{\log S})}\log(1/\varepsilon) \end{aligned}$$

(where the last equality yet again uses $S = 2^{\omega(\log\log(n))^2}$), and this completes the proof. □

7.2.1 Fooling depth-t decision trees with degree-k $\mathbb{F}_2$ polynomials at its leaves

We recall the following well-known result of Viola:

Theorem 7.4 ([76]). *The sum of k independent $(\frac{1}{16}\delta^{2^{k-1}})$-biased distributions δ-fools the class of degree-k $\mathbb{F}_2$ polynomials.*

Earlier work of Lovett [50] proved the weaker statement with 2^k independent copies instead of k. We note that Lovett's result suffices for our purposes.

We will need a few simple facts about distributions. Recall that a distribution $\mathcal{D}$ is a *mixture* of *component* distributions $\mathcal{D}^{(1)}, \ldots, \mathcal{D}^{(\ell)}$ if there exist non-negative weights $w_1, \ldots, w_\ell$ summing to 1 such that making a draw from $\mathcal{D}$ corresponds to first drawing $i \in [\ell]$ with probability w_i and then making a draw from $\mathcal{D}^{(i)}$.

Fact 7.5. *Let $\mathscr{C}$ be a class of functions and suppose that distributions $\mathcal{D}^{(1)}, \ldots, \mathcal{D}^{(\ell)}$ each δ-fool $\mathscr{C}$. Then any mixture $\mathcal{D}$ of distributions $\mathcal{D}^{(1)}, \ldots, \mathcal{D}^{(\ell)}$ also δ-fools $\mathscr{C}$.*

We say that a class $\mathscr{C}$ of Boolean functions is *closed under shifts* if for all $f \in \mathscr{C}$ and $y \in \{0,1\}^n$, the function $g(x) := f(x+y)$ is also in $\mathscr{C}$ (where addition is coordinatewise over $\mathbb{F}_2$). An easy consequence of Fact 7.5 is the following:

Fact 7.6. *Let $\mathscr{C}$ be a class of functions, closed under shifts, that is δ-fooled by a distribution $\mathcal{D}$. Let $\mathcal{D}'$ be any other independent distribution. Then the distribution $\mathcal{D} + \mathcal{D}'$, where a draw from $\mathcal{D} + \mathcal{D}'$ is $x + y$ where $x \leftarrow \mathcal{D}$ and $y \leftarrow \mathcal{D}'$, also δ-fools $\mathscr{C}$.*

Finally we recall the following which is an easy consequence of the definition of a δ-biased distribution:

Fact 7.7 (Conditioning a δ-biased distribution). *Let $\mathcal{D}$ be a δ-biased distribution over $\{0,1\}^n$. Fix $i \in [n]$ and $b \in \{0,1\}$, and let $\mathcal{D}'$ denote the distribution of $x \leftarrow \mathcal{D}$ conditioned on $x_i = b$. Then the marginal distribution of $\mathcal{D}'$ on the coordinates in $[n] \setminus \{i\}$ is $2\delta/(1-\delta) \le 4\delta$ biased.*

Lemma 7.8 (Fooling decision trees with low-degree polynomials at leaves)**.** *Let $\mathcal{D}^{(1)}_{\delta'\text{-biased}}, \ldots, \mathcal{D}^{(k)}_{\delta'\text{-biased}}$ be k independent δ'-biased distribution where $\delta' = \frac{1}{16}\delta^{2^{k-1}} \cdot 4^{-t}$. Let $\mathcal{D}_{t\text{-wise}}$ be an independent t-wise independent distribution. Then the sum*

$$\mathcal{D} := \mathcal{D}^{(1)}_{\delta'\text{-biased}} + \cdots + \mathcal{D}^{(k)}_{\delta'\text{-biased}} + \mathcal{D}_{t\text{-wise}}$$

δ-fools the class of depth-t decision trees with degree-k polynomials at its leaves. Since δ'-biased distributions can be generated with seed length $O(\log(n) + \log(1/\delta'))$, and t-wise independent distributions with seed length $O(t\log(n))$, we get that we can sample from $\mathcal{D}$ using

$$k \cdot O(t + 2^k \log(1/\delta)) + O(t\log(n))$$

random bits.

The intuition underlying Lemma 7.8 is as follows:

1. $\mathcal{D}_{t\text{-wise}}$ ensures that every branch of the decision tree is taken with the right probability.
2. By Fact 7.7, each $\mathcal{D}_{\delta'\text{-biased}}$ remains $(\frac{1}{16}\delta^{2^{k-1}} 4^{-t}) \cdot 4^t = \frac{1}{16}\delta^{2^{k-1}}$-biased even when conditioned on a length-t branch. By Theorem 7.4, their sum δ-fools the degree-k polynomial at the leaf.

Proof. Let F be computed by a depth-t decision tree T with degree-k polynomials at its leaves. We begin by noting that every branch π of T is taken with the right probability under a random draw from $\mathcal{D}$:

$$\begin{aligned}
\mathop{\mathbf{E}}_{y \leftarrow \mathcal{D}}\big[F(y) = 1\big] &= \sum_{\pi \in T} \mathop{\mathbf{Pr}}_{y \leftarrow \mathcal{D}}[\, y \text{ follows } \pi \,] \cdot \mathop{\mathbf{E}}_{y \leftarrow \mathcal{D}}\big[(F \upharpoonright \pi)(y) \mid y \text{ follows } \pi\big] \\
&= \sum_{\pi \in T} \mathop{\mathbf{Pr}}_{x \leftarrow \mathcal{U}}[\, x \text{ follows } \pi \,] \cdot \mathop{\mathbf{E}}_{y \leftarrow \mathcal{D}}\big[(F \upharpoonright \pi)(y) \mid y \text{ follows } \pi\big], \\
&\qquad\qquad\qquad\qquad\qquad (\text{since } \mathcal{D} \text{ is } t\text{-wise independent and } |\pi| \le t)
\end{aligned}$$

so it remains to show that

$$\mathop{\mathbf{E}}_{y \leftarrow \mathcal{D}}\big[(F \upharpoonright \pi)(y) \mid y \text{ follows } \pi\big] = \mathop{\mathbf{E}}_{x \leftarrow \mathcal{U}}\big[(F \upharpoonright \pi)(x)\big] \pm \delta \qquad \text{for all } \pi \in T.$$

Since for all $\pi \in T$ $F \upharpoonright \pi$ is a degree-k polynomial over the coordinates in $[n] \setminus \mathrm{supp}(\pi)$, it suffices to show that $\mathcal{D} \upharpoonright \pi$, the distribution of $y \leftarrow \mathcal{D}$ conditioned on y following π, δ-fools the class of degree-k polynomials over the coordinates in $[n] \setminus \mathrm{supp}(\pi)$.

Fix $\pi \in T$ and let S denote $\mathrm{supp}(\pi)$. We will express $\mathcal{D} \upharpoonright \pi$ as a mixture of distributions, and argue that each component distribution in the mixture δ-fools the class of degree-k polynomials over the coordinates in $[n] \setminus S$. Recall that $\mathcal{D}$ is the sum of $k+1$ many independent distributions

$$\mathcal{D} = \mathcal{D}^{(1)}_{\delta'\text{-biased}} + \cdots + \mathcal{D}^{(k)}_{\delta'\text{-biased}} + \mathcal{D}_{t\text{-wise}},$$

and so a draw $y = z^{(1)} + \cdots + z^{(k+1)}$ is consistent with π iff

$$z^{(1)}_S + \cdots + z^{(k+1)}_S = \pi_S.$$

Therefore, $\mathcal{D} \upharpoonright \pi$ is a mixture of component distributions each of which is the sum of $k+1$ independent distributions. Each component distribution is specified by a $(k+1)$-tuple $(\pi^{(1)}, \ldots, \pi^{(k+1)})$ where $\operatorname{supp}(\pi^{(i)}) = S$ for all $i \in [k+1]$ and

$$\bigoplus_{i \in [k+1]} \pi_S^{(i)} = \pi_S.$$

Given such a $(k+1)$-tuple $(\pi^{(1)}, \ldots, \pi^{(k+1)})$, the corresponding component distribution is

$$\left(\mathcal{D}^{(1)}_{\delta'\text{-biased}} \upharpoonright \pi^{(1)}\right) + \cdots + \left(\mathcal{D}^{(k)}_{\delta'\text{-biased}} \upharpoonright \pi^{(k)}\right) + \left(\mathcal{D}_{t\text{-wise}} \upharpoonright \pi^{(k+1)}\right). \tag{7.1}$$

(The values of the mixing weights for the components are By Fact 7.7, the marginal distribution of each $\mathcal{D}^{(i)}_{\delta\text{-biased}} \upharpoonright \pi^{(i)}$ on the coordinates in $[n] \setminus S$ is

$$\delta' \cdot 4^{|\pi^{(i)}|} \le \left(\frac{1}{16}\delta^{2^{k-1}} 4^{-t}\right) \cdot 4^t = \frac{1}{16}\delta^{2^{k-1}}$$

biased, and hence by Viola's theorem (Theorem 7.4) their sum

$$\left(\mathcal{D}^{(1)}_{\delta'\text{-biased}} \upharpoonright \pi^{(1)}\right) + \cdots + \left(\mathcal{D}^{(k)}_{\delta'\text{-biased}} \upharpoonright \pi^{(k)}\right)$$

δ-fools the class of degree-k polynomials over the coordinates in $[n] \setminus S$. By Fact 7.6, so does the distribution in (7.1). By Fact 7.5 the mixture distribution $\mathcal{D} \upharpoonright \pi$ likewise δ-fools the class of degree-k polynomials over the coordinates in $[n] \setminus S$, and the proof is complete. $\square$

A Proof sketch of Theorem 3.8

We sketch a proof of the following:

Theorem A.1. *Let $\mathscr{F} = (F_1, \ldots, F_M)$ be an ordered collection of k-CNFs. Then for all $t, \ell \in \mathbb{N}$,*

$$\Pr_{\rho \leftarrow \mathcal{R}_p}[\operatorname{depth}(\mathrm{CCDT}_\ell(\mathscr{F} \upharpoonright \rho) \ge t] \le M^{\lceil t/\ell \rceil}(32pk)^t.$$

Our proof sketch of Theorem A.1 is carried out in the "encoding-decoding" framework of Razborov's alternative proof [61] of Håstad's original switching lemma [37], Theorem 3.1. (For a detailed exposition of Razborov's proof technique see [15, 71] and Chapter §14 of [9].) We emphasize that the ideas in our proof of Theorem A.1 are all from [38], but in our view the encoding–decoding presentation is more amenable to the derandomization that we ultimately require than the conditioning-based inductive argument given in [38]. We also note that a similar proof based on the encoding–decoding framework, achieving a very similar bound, appears in [70, Section 7].

A.1 Bad restrictions and the structure of witnessing paths

Fix $\mathscr{F} = (F_1, \dots, F_M)$ and consider the set $\mathcal{B} \subseteq \{0,1,*\}^n$ of all *bad* restrictions ρ, namely the ones such that

$$\mathrm{depth}(\mathrm{CCDT}_\ell(\mathscr{F} \upharpoonright \rho)) \geq t.$$

Fix any bad restriction $\rho \in \mathcal{B}$. Recalling our definition of the set of canonical common partial decision trees (Definition 3.7), there exists a canonical common ℓ-partial decision tree $T \in \mathrm{CCDT}_\ell(\mathscr{F} \upharpoonright \rho)$ and a path Π of length exactly t through T. Furthermore, we have that

1. There exist indices $1 \leq i_1 \leq i_2 \leq \cdots \leq i_u \leq M$ where $u \leq \lceil t/\ell \rceil$, and

2. $\Pi = \pi^{(1)} \circ \cdots \circ \pi^{(u)}$, where for all $j \in [u]$, we have that $\mathrm{supp}(\pi^{(j)}) = \mathrm{supp}(\eta^{(j)})$ where $\eta^{(j)}$ is a path through the canonical decision tree

$$\mathrm{CDT}(F_{i_j} \upharpoonright \rho \circ \pi^{(1)} \circ \cdots \circ \pi^{(j-1)}).$$

   Furthermore, for every $j \in [u-1]$ we have that $\eta^{(j)}$ is a full path of length between $\ell+1$ and $\ell+k$ through the CDT, and $\eta^{(u)}$ is a path of length exactly $t - \sum_{j=1}^{u-1} |\mathrm{supp}(\eta^{(j)})|$. (Note that $\eta^{(u)}$ is not necessarily a full path.)

(Note that by (2), these restrictions $\pi^{(j)}$ are supported on mutually disjoint sets of coordinates.)

A.2 Encoding bad restrictions ρ

Recalling the statement of Theorem A.1, our goal is to bound $\mathbf{Pr}_{\rho \leftarrow \mathcal{R}_p}[\rho \in \mathcal{B}]$, the weight of the set $\mathcal{B}$ of bad restrictions under $\mathcal{R}_p$. To do so, we define an encoding of each bad restriction $\rho \in \mathcal{B}$ as a different restriction $\rho' \in \{0,1,*\}^n$ and a small amount (say at most m bits) of "auxiliary information":

$$\mathrm{encode} : \mathcal{B} \to \{0,1,*\}^n \times \{0,1\}^m$$
$$\mathrm{encode}(\rho) = (\rho', \text{auxiliary information})$$

This encoding should satisfy two key properties. First, it should be uniquely decodable, meaning that one is always able to recover ρ given ρ' and the auxiliary information; equivalently, the function $\mathrm{encode}(\cdot)$ is an injection. Second, ρ' should extend ρ by exactly t bits, meaning that $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\rho')$ and $|\mathrm{supp}(\rho') \setminus \mathrm{supp}(\rho)| = t$. From this second property we get that

$$\frac{\mathbf{Pr}_{\rho \leftarrow \mathcal{R}_p}[\rho = \rho']}{\mathbf{Pr}_{\rho \leftarrow \mathcal{R}_p}[\rho = \rho]} = \left(\frac{1-p}{2p}\right)^t,$$

i. e., that the weight of ρ' under $\mathcal{R}_p$ is larger than that of ρ by a $O(p)^{-t}$ multiplicative factor. It is not hard to see that together, these two properties imply that the total weight of all bad restrictions with the *same* auxiliary information is at most $O(p)^t$. To complete the proof of Theorem A.1, we then bound the overall weight of $\mathcal{B}$ via a union bound over all 2^m possible strings of auxiliary information, giving us a failure probability of

$$2^m \cdot \left(\frac{2p}{1-p}\right)^t. \tag{A.1}$$

We now describe the encoding in more detail. Given a bad restriction $\rho \in \mathcal{B}$, the extension ρ' of ρ will be

$$\rho' = \rho \circ \sigma^{(1)} \circ \cdots \circ \sigma^{(u)}, \qquad u \leq \lceil t/\ell \rceil \tag{A.2}$$

where $\sigma^{(j)}$ is a restriction that is supported on the same coordinates as $\pi^{(j)}$ for all $j \in [u]$. (Hence these restrictions $\sigma^{(j)}$'s are supported on mutually disjoint sets of coordinates, every $\sigma^{(j)}$ has length between $\ell+1$ and $\ell+k$, except $\sigma^{(u)}$ which has length $t - \sum_{j=1}^{u-1} |\mathrm{supp}(\sigma^{(j)})|$ and is not necessarily a full path.)

We now define these restrictions $\sigma^{(j)}$. Recall that $\mathrm{supp}(\pi^{(j)}) = \mathrm{supp}(\eta^{(j)})$ where $\eta^{(j)}$ is a full path of length between $\ell+1$ and $\ell+k$ through the canonical decision tree $\mathrm{CDT}(F_{i_j} \upharpoonright \rho \circ \pi^{(1)} \circ \cdots \circ \pi^{(j-1)})$, a full path witnessing that fact that $\mathrm{CDT}(F_{i_j} \upharpoonright \rho \circ \pi^{(1)} \circ \cdots \circ \pi^{(j-1)}) > \ell$. That is, $\eta^{(j)}$ is a full path witnessing the fact that $\rho \circ \pi^{(1)} \circ \cdots \circ \pi^{(j-1)}$ is a bad restriction for the usual switching lemma, and $\pi^{(j)}$ is an assignment to the variables in $\mathrm{supp}(\eta^{(j)})$. (Once again this is with the possible exception of the segment $\pi^{(u)}$ of Π, which has length $t - \sum_{j=1}^{u-1} |\mathrm{supp}(\pi^{(j)})|$ and is not necessarily a full path.) Razborov's encoding–decoding proof of the usual switching lemma defines an encoding of this bad restriction $\rho \circ \pi^{(1)} \circ \cdots \circ \pi^{(j-1)}$ to an extension

$$\rho^{(j)} := \rho \circ \pi^{(1)} \circ \cdots \circ \pi^{(j-1)} \circ \sigma^{(j)}$$

where $\sigma^{(j)}$ is supported on the same ℓ coordinates as $\eta^{(j)}$ (and hence $\pi^{(j)}$ as well). Razborov's proof hinges on the fact that given the k-CNF F, this encoding $\rho^{(j)}$, and a small amount of auxiliary information, one is able to recover the bad restriction $\rho \circ \pi^{(1)} \circ \cdots \circ \pi^{(j-1)}$; that is, one is able to "undo" $\sigma^{(j)}$ in $\rho^{(j)}$, flipping the coordinates in $\mathrm{supp}(\sigma^{(j)})$ from $\{0,1\}$ back to $*$. This restriction $\sigma^{(j)}$ as defined in Razborov's proof is precisely the $\sigma^{(j)}$ we will use in our encoding (A.2).

We summarize the discussion above in the following fact:

Fact A.2 (Main lemma in encoding–decoding proof of the usual switching lemma, notation specialized to our current context)**.** *Let F_{i_j} be a k-CNF, $\rho \circ \pi^{(1)} \circ \cdots \pi^{(j-1)}$ be a restriction, and $\eta^{(j)}$ be a path in* $\mathrm{CDT}(F_{i_j} \upharpoonright \rho \circ \pi^{(1)} \circ \cdots \circ \pi^{(j-1)})$*. There is a restriction $\sigma^{(j)}$ to the coordinates in* $\mathrm{supp}(\eta^{(j)})$ *such that given*

1. *The k-CNF F_{i_j},*
2. *The restriction $\rho^{(j)} = \rho \circ \pi^{(1)} \circ \cdots \pi^{(j-1)} \circ \sigma^{(j)}$,*
3. $|\mathrm{supp}(\eta^{(j)})| \cdot (2 + \log k)$ *bits of auxiliary information $\iota(\rho \circ \pi^{(1)} \circ \cdots \circ \pi^{(j-1)}, F_{i_j}, \eta_j)$,*

a decoder is able to recover the restriction $\pi^{(1)} \circ \cdots \circ \pi^{(j-1)}$.

Furthermore, if $\eta^{(j)}$ is a full *path in* $\mathrm{CDT}(F_{i_j} \upharpoonright \rho \circ \pi^{(1)} \circ \cdots \circ \pi^{(j-1)})$ *(recall the definition of a full path given in Definition 3.6) then given*

1. *The k-CNF F_{i_j},*
2. Any extension $\rho^{(j)}$ of *the restriction $\rho^{(j)} = \rho \circ \pi^{(1)} \circ \cdots \circ \pi^{(j-1)} \circ \sigma^{(j)}$,*
3. $|\mathrm{supp}(\eta^{(j)})| \cdot (2 + \log k)$ *bits of auxiliary information $\iota(\rho \circ \pi^{(1)} \circ \cdots \circ \pi^{(j-1)}, F_{i_j}, \eta_j)$,*

a decoder is able to "undo" $\sigma^{(j)}$ in $\rho^{(j)}$, by which we mean that she is able to recover the restriction $\bar{\rho}^{(j)}$ where

$$\bar{\rho}_i^{(j)} = \begin{cases} * & \text{if } i \in \operatorname{supp}(\sigma^{(j)}) \\ \rho_i^{(j)} & \text{otherwise.} \end{cases}$$

A.3 Our auxiliary information

We will provide the decoder with

1. $u \log M$ bits of information specifying the u indices $i_1, \ldots, i_u \in [M]$.
2. The auxiliary information $\iota(\rho \circ \pi^{(1)} \circ \cdots \circ \pi^{(j-1)}, F_{i_j}, \eta_j)$ for all $j \in [u]$ (as defined in Fact A.2), a total of
$$\sum_{j=1}^{u} |\operatorname{supp}(\eta^{(j)})| \cdot (2 + \log k) = t \cdot (2 + \log k)$$
many bits.
3. t bits of information specifying the length-t path $\Pi = \pi^{(1)} \circ \cdots \circ \pi^{(u)}$ through $\mathrm{CCDT}_\ell(\mathscr{F} \upharpoonright \rho)$.

This is a total of
$$m := u \log M + t \log k + 3t$$
bits of auxiliary information; recalling equation (A.1) and the preceding discussion, to establish Theorem A.1 it remains to argue that the map $\operatorname{encode}(\rho) = (\rho', \text{auxiliary information})$ is indeed invertible.

A.4 Decoding

Fix $\mathscr{F} = (F_1, \ldots, F_M)$ and consider a bad restriction $\rho \in \mathcal{B}$, one such that
$$\operatorname{depth}(\mathrm{CCDT}_\ell(\mathscr{F} \upharpoonright \rho)) \geq t.$$

Let $\Pi = \pi^{(1)} \circ \cdots \circ \pi^{(u)}$ be a path of length t through a canonical common ℓ-partial decision tree $T \in \mathrm{CCDT}_\ell(\mathscr{F} \upharpoonright \rho)$ that witnesses the badness of ρ. We claim that for all $j \in [u]$, given

1. The family of k-CNFs $\mathscr{F}$,
2. The "hybrid" restriction $\rho^{(j)} := \rho \circ \pi^{(1)} \circ \cdots \circ \pi^{(j-1)} \circ \sigma^{(j)} \circ \cdots \circ \sigma^{(u)}$,
3. The auxiliary information described in Section A.3,

the decoder can recover the "next" hybrid restriction $\rho^{(j+1)} := \rho \circ \pi^{(1)} \circ \cdots \pi^{(j)} \circ \sigma^{(j+1)} \circ \cdots \circ \sigma^{(u)}$. Before justifying this claim, we note that from this claim we get that the map

$$\operatorname{encode}(\rho) = (\rho', \text{auxiliary information}) \tag{A.3}$$

is indeed invertible, i. e., that given ρ' as defined in (A.2) and the auxiliary information described above, we can recover ρ (this would complete our proof of Theorem A.1). To see this, we first observe that ρ'

is simply $\rho^{(1)}$. Applying the claim u times the decoder is able to iteratively recover $\rho^{(2)}, \ldots, \rho^{(u+1)} = \rho \circ \pi^{(1)} \circ \cdots \circ \pi^{(u)}$, and having done so she will have identified $\mathrm{supp}(\pi^{(1)} \circ \cdots \circ \pi^{(u)})$. With this information she is then able to recover ρ from $\rho^{(u+1)}$ (simply by flipping the bits in $\mathrm{supp}(\pi^{(1)} \circ \cdots \circ \pi^{(u)})$ back to $*$'s).

We now show how the decoder obtains $\rho^{(j+1)}$ from $\rho^{(j)}$ for all $j \in [u]$. First, since the auxiliary information specifies $i_j \in [M]$ she is able to identify F_{i_j} within $\mathscr{F}$. Next,

- For $j \in [u-1]$, we recall that $\eta^{(j)}$ is a full path in $\mathrm{CDT}(F_{i_j} \upharpoonright \rho \circ \pi^{(1)} \circ \cdots \circ \pi^{(j-1)})$ (whose length is between $\ell+1$ and $\ell+k$ and hence is "approximately known" to the decoder), and hence we may apply the "Furthermore" part of Fact A.2 to "undo" $\sigma^{(j)}$ in $\rho^{(j)}$ and obtain the restriction $\rho \circ \pi^{(1)} \circ \cdots \pi^{(j-1)} \circ \sigma^{(j+1)} \circ \cdots \circ \sigma^{(u)}$.

- For $j = u$, while $\eta^{(u)}$ is not necessarily a full path in $\mathrm{CDT}(F_{i_u} \upharpoonright \rho \circ \pi^{(1)} \circ \cdots \circ \pi^{(u-1)})$ we observe that $\rho^{(u)}$ is simply $\rho^{(u)}$, and hence we may apply the first part of Fact A.2 to obtain the restriction $\rho \circ \pi^{(1)} \circ \cdots \circ \pi^{(u-1)}$.

In either case, since our auxiliary information to the decoder specifies the values of $\pi^{(j)}$ on $\mathrm{supp}(\sigma^{(j)})$, the decoder is able to fill in these coordinates accordingly to obtain $\rho^{(j+1)}$.

Acknowledgements

We thank Prahladh Harsha and Srikanth Srinivasan for helpful discussions.

References

[1] SCOTT AARONSON: A counterexample to the generalized Linial–Nisan conjecture. *Electron. Colloq. Comput. Complexity*, TR10-109, 2010. [ECCC] 5

[2] SCOTT AARONSON: BQP and the polynomial hierarchy. In *Proc. 42nd STOC*, pp. 141–150. ACM Press, 2010. [doi:10.1145/1806689.1806711] 5

[3] MANINDRA AGRAWAL, ERIC ALLENDER, RUSSELL IMPAGLIAZZO, TONIANN PITASSI, AND STEVEN RUDICH: Reducing the complexity of reductions. *Comput. Complexity*, 10(2):117–138, 2001. [doi:10.1007/s00037-001-8191-1] 3

[4] MIKLÓS AJTAI: Σ_1^1-formulae on finite structures. *Ann. Pure Appl. Logic*, 24(1):1–48, 1983. [doi:10.1016/0168-0072(83)90038-6] 2, 3, 4, 6, 7, 8

[5] MIKLÓS AJTAI: Geometric properties of sets defined by constant depth circuits. In *Combinatorics, Paul Erdős is Eighty, Vol. 1*, Bolyai Soc. Math. Studies, pp. 19–31. J. Bolyai Math. Soc., Budapest, 1993. 3

[6] MIKLÓS AJTAI AND AVI WIGDERSON: Deterministic simulation of probabilistic constant depth circuits. *Adv. Comput. Res.*, 5:199–222, 1989. Preliminary version in FOCS'85. 2, 3, 5, 6, 7, 8, 9, 10, 11, 22, 24, 25

[7] NOGA ALON, LÁSZLÓ BABAI, AND ALON ITAI: A fast and simple randomized algorithm for the maximal independent set problem. *J. Algorithms*, 7(4):567–583, 1986. [doi:10.1016/0196-6774(86)90019-2] 27

[8] NOGA ALON, ODED GOLDREICH, JOHAN HÅSTAD, AND RENÉ PERALTA: Simple constructions of almost k-wise independent random variables. *Random Struct. Algor.*, 3(3):289–304, 1992. Preliminary version in FOCS'90. [doi:10.1002/rsa.3240030308] 8

[9] SANJEEV ARORA AND BOAZ BARAK: *Computational Complexity: A modern approach*. Cambridge Univ. Press, 2009. [doi:10.1017/CBO9780511804090] 35

[10] LÁSZLÓ BABAI: Random oracles separate PSPACE from the polynomial-time hierarchy. *Inform. Process. Lett.*, 26(1):51–53, 1987. [doi:10.1016/0020-0190(87)90036-6] 4

[11] LÁSZLÓ BABAI, NOAM NISAN, AND MÁRIÓ SZEGEDY: Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *J. Comput. System Sci.*, 45(2):204–232, 1992. [doi:10.1016/0022-0000(92)90047-M] 8, 9, 10

[12] MARSHALL BALL, DANA DACHMAN-SOLED, SIYAO GUO, TAL MALKIN, AND LI-YANG TAN: Non-malleable codes for small-depth circuits. In *Proc. 59th FOCS*, pp. 826–837. IEEE Comp. Soc., 2018. [doi:10.1109/FOCS.2018.00083] 2, 4

[13] LOUAY M. J. BAZZI: Polylogarithmic independence can fool DNF formulas. *SIAM J. Comput.*, 38(6):2220–2272, 2009. [doi:10.1137/070691954] 5, 6

[14] PAUL BEAME: Lower bounds for recognizing small cliques on CRCW PRAM's. *Discr. Appl. Math.*, 29(1):3–20, 1990. [doi:10.1016/0166-218X(90)90079-R] 2

[15] PAUL BEAME: A switching lemma primer. Technical Report UW-CSE-95-07-01, Univ. Washington, 1994. LINK. 35

[16] PAUL BEAME, RUSSELL IMPAGLIAZZO, AND SRIKANTH SRINIVASAN: Approximating AC^0 by small height decision trees and a deterministic algorithm for $\#\mathsf{AC}^0$-SAT. In *Proc. 27th IEEE Conf. on Comput. Complexity (CCC'12)*, pp. 117–125. IEEE Comp. Soc., 2012. [doi:10.1109/CCC.2012.40] 2, 4

[17] MANUEL BLUM AND SILVIO MICALI: How to construct cryptographically strong sequences of pseudorandom bits. *SIAM J. Comput.*, 13(4):850–864, 1984. Preliminary version in FOCS'82. [doi:10.1137/0213053] 5

[18] ANDREJ BOGDANOV: Pseudorandom generators for low degree polynomials. In *Proc. 37th STOC*, pp. 21–30. ACM Press, 2005. [doi:10.1145/1060590.1060594] 8, 9

[19] ANDREJ BOGDANOV AND EMANUELE VIOLA: Pseudorandom bits for polynomials. *SIAM J. Comput.*, 39(6):2464–2486, 2010. [doi:10.1137/070712109] 8, 9

[20] JEAN BOURGAIN: Estimation of certain exponential sums arising in complexity theory. *C. R. Math. Acad. Sci. Paris*, 340(9):627–631, 2005. [doi:10.1016/j.crma.2005.03.008] 8

[21] MARK BRAVERMAN: Polylogarithmic independence fools AC^0 circuits. *J. ACM*, 57(5):1–10, 2010. [doi:10.1145/1754399.1754401] 5, 7

[22] JIN-YI CAI: With probability one, a random oracle separates PSPACE from the polynomial-time hierarchy. *J. Comput. System Sci.*, 38(1):68–85, 1989. Preliminary version in STOC'86. [doi:10.1016/0022-0000(89)90033-0] 3, 4

[23] ARKADEV CHATTOPADHYAY: Discrepancy and the power of bottom fan-in in depth-three circuits. In *Proc. 48th FOCS*, pp. 449–458. IEEE Comp. Soc., 2007. [doi:10.1109/FOCS.2007.30, ECCC:TR07-050] 8

[24] ESHAN CHATTOPADHYAY, POOYA HATAMI, KAAVE HOSSEINI, AND SHACHAR LOVETT: Pseudorandom generators from polarizing random walks. *Theory of Computing*, 15(10):1–26, 2019. Preliminary version in CCC'18. [doi:10.4086/toc.2019.v015a010] 5, 6, 7

[25] ESHAN CHATTOPADHYAY, POOYA HATAMI, SHACHAR LOVETT, AND AVISHAY TAL: Pseudorandom generators from the second Fourier level and applications to AC^0 with parity gates. In *Proc. 10th Innovations in Theoret. Comp. Sci. conf. (ITCS'19)*, pp. 22:1–15. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019. [doi:10.4230/LIPIcs.ITCS.2019.22] 8

[26] ESHAN CHATTOPADHYAY AND XIN LI: Non-malleable codes and extractors for small-depth circuits, and affine functions. In *Proc. 49th STOC*, pp. 1171–1184. ACM Press, 2017. [doi:10.1145/3055399.3055483] 2

[27] SHIVA CHAUDHURI AND JAIKUMAR RADHAKRISHNAN: Deterministic restrictions in circuit complexity. In *Proc. 28th STOC*, pp. 30–36. ACM Press, 1996. [doi:10.1145/237814.237824, ECCC:TR96-004] 3

[28] XI CHEN, IGOR CARBONI OLIVEIRA, ROCCO A. SERVEDIO, AND LI-YANG TAN: Near-optimal small-depth lower bounds for small distance connectivity. In *Proc. 48th STOC*, pp. 612–625. ACM Press, 2016. [doi:10.1145/2897518.2897534] 2

[29] ANINDYA DE, OMID ETESAMI, LUCA TREVISAN, AND MADHUR TULSIANI: Improved pseudorandom generators for depth 2 circuits. In *Proc. 14th Internat. Workshop on Randomization and Computation (RANDOM'10)*, pp. 504–517. Springer, 2010. [doi:10.1007/978-3-642-15369-3_38] 5

[30] STEFAN DZIEMBOWSKI, KRZYSZTOF PIETRZAK, AND DANIEL WICHS: Non-malleable codes. *J. ACM*, 65(4):20:1–32, 2018. [doi:10.1145/3178432] 4

[31] BILL FEFFERMAN, RONEN SHALTIEL, CHRISTOPHER UMANS, AND EMANUELE VIOLA: On beating the hybrid argument. *Theory of Computing*, 9(26):809–843, 2013. Preliminary version in ITCS'12. [doi:10.4086/toc.2013.v009a026, ECCC:TR10-186] 5

[32] MERRICK FURST, JAMES SAXE, AND MICHAEL SIPSER: Parity, circuits, and the polynomial-time hierarchy. *Math. Sys. Theory*, 17(1):13–27, 1984. [doi:10.1007/BF01744431] 2, 3

[33] ODED GOLDREICH AND AVI WIDGERSON: On derandomizing algorithms that err extremely rarely. In *Proc. 46th STOC*, pp. 109–118. ACM Press, 2014. [doi:10.1145/2591796.2591808] 3, 5

[34] PARIKSHIT GOPALAN, RAGHU MEKA, AND OMER REINGOLD: DNF sparsification and a faster deterministic counting algorithm. *Comput. Complexity*, 22(2):275–310, 2013. [doi:10.1007/s00037-013-0068-6] 3, 5

[35] PARIKSHIT GOPALAN, RAGHU MEKA, OMER REINGOLD, LUCA TREVISAN, AND SALIL P. VADHAN: Better pseudorandom generators from milder pseudorandom restrictions. In *Proc. 53rd FOCS*, pp. 120–129. IEEE Comp. Soc., 2012. [doi:10.1109/FOCS.2012.77, arXiv:1210.0049, ECCC:TR12-123] 3, 5, 22

[36] PRAHLADH HARSHA AND SRIKANTH SRINIVASAN: On polynomial approximations to AC^0. *Random Struct. Algor.*, 54(2):289–303, 2019. Preliminary version in RANDOM'16. [doi:10.1002/rsa.20786] 5, 6, 7, 8, 26

[37] JOHAN HÅSTAD: Almost optimal lower bounds for small depth circuits. In *Proc. 18th STOC*, pp. 6–20. ACM Press, 1986. [doi:10.1145/12130.12132] 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 35

[38] JOHAN HÅSTAD: On the correlation of parity and small-depth circuits. *SIAM J. Comput.*, 43(5):1699–1708, 2014. [doi:10.1137/120897432, ECCC:TR12-137] 2, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 35

[39] JOHAN HÅSTAD: An average-case depth hierarchy theorem for higher depths. In *Proc. 57th FOCS*, pp. 79–88. IEEE Comp. Soc., 2016. [doi:10.1109/FOCS.2016.18, ECCC:TR16-041] 2

[40] JOHAN HÅSTAD: On small-depth Frege proofs for Tseitin for grids. *J. ACM*, 68(1):1–31, 2020. Preliminary version in FOCS'17. [doi:10.1145/3425606, ECCC:TR17-142] 2

[41] JOHAN HÅSTAD AND MIKAEL GOLDMANN: On the power of small-depth threshold circuits. *Comput. Complexity*, 1(2):113–129, 1991. [doi:10.1007/BF01272517] 8

[42] JOHAN HÅSTAD, BENJAMIN ROSSMAN, ROCCO A. SERVEDIO, AND LI-YANG TAN: An average-case depth hierarchy theorem for Boolean circuits. *J. ACM*, 64(5):1–27, 2017. Preliminary version in FOCS'15. [doi:10.1145/3095799, ECCC:TR15-065] 2

[43] RUSSELL IMPAGLIAZZO, WILLIAM MATTHEWS, AND RAMAMOHAN PATURI: A satisfiability algorithm for AC^0. In *Proc. 23rd Ann. ACM–SIAM Symp. on Discrete Algorithms (SODA'12)*, pp. 961–972. SIAM, 2012. [doi:10.1137/1.9781611973099.77] 2, 3, 4, 8, 12

[44] RUSSELL IMPAGLIAZZO, RAGHU MEKA, AND DAVID ZUCKERMAN: Pseudorandomness from shrinkage. *J. ACM*, 66(2):11:1–16, 2019. [doi:10.1145/3230630] 4, 22

[45] ADAM KLIVANS: On the derandomization of constant depth circuits. In *Proc. 5th Internat. Workshop on Randomization and Computation (RANDOM'01)*, pp. 249–260. Springer, 2001. [doi:10.1007/3-540-44666-4_28] 5

[46] ADAM KLIVANS, HOMIN LEE, AND ANDREW WAN: Mansour's conjecture is true for random DNF formulas. In *Proc. 23rd Ann. Conf. on Learning Theory (COLT'10)*, pp. 368–380. Springer, 2010. [ECCC:TR10-023] 5

[47] JAN KRAJÍČEK, PAVEL PUDLÁK, AND ALAN WOODS: An exponential lower bound to the size of bounded depth Frege proofs of the pigeonhole principle. *Random Struct. Algor.*, 7(1):15–39, 1995. [doi:10.1002/rsa.3240070103, ECCC:TR94-018] 2

[48] NATHAN LINIAL, YISHAY MANSOUR, AND NOAM NISAN: Constant depth circuits, Fourier transform and learnability. *J. ACM*, 40(3):607–620, 1993. Preliminary version in FOCS'89. [doi:10.1145/174130.174138] 2

[49] NATHAN LINIAL AND NOAM NISAN: Approximate inclusion-exclusion. *Combinatorica*, 10(4):349–365, 1990. Preliminary version in STOC'90. [doi:10.1007/BF02128670] 5, 6

[50] SHACHAR LOVETT: Unconditional pseudorandom generators for low-degree polynomials. *Theory of Computing*, 5(3):69–82, 2009. [doi:10.4086/toc.2009.v005a003] 8, 9, 10, 33

[51] SHACHAR LOVETT AND SRIKANTH SRINIVASAN: Correlation bounds for poly-size AC^0 circuits with $n^{1-o(1)}$ symmetric gates. In *Proc. 15th Internat. Workshop on Randomization and Computation (RANDOM'11)*, pp. 640–651. Springer, 2011. [doi:10.1007/978-3-642-22935-0_54] 5, 8

[52] CHI-JEN LU: Hitting set generators for sparse polynomials over any finite fields. In *Proc. 27th IEEE Conf. on Comput. Complexity (CCC'12)*, pp. 280–286. IEEE Comp. Soc., 2012. [doi:10.1109/CCC.2012.20] 8

[53] MICHAEL LUBY AND BOBAN VELIČKOVIĆ: On deterministic approximation of DNF. *Algorithmica*, 16(4–5):415–433, 1996. Preliminary version in STOC'91. [doi:10.1007/BF01940873] 5

[54] MICHAEL LUBY, BOBAN VELIČKOVIĆ, AND AVI WIGDERSON: Deterministic approximate counting of depth-2 circuits. In *Proc. 2nd Isr. Symp. Theory Comp. Sys. (ISTCS'93)*, pp. 18–24. IEEE Comp. Soc., 1993. [doi:10.1109/ISTCS.1993.253488] 5, 8, 9, 10

[55] RAGHU MEKA, OMER REINGOLD, AND AVISHAY TAL: Pseudorandom generators for width-3 branching programs. In *Proc. 51st STOC*, pp. 626–637. ACM Press, 2019. [doi:10.1145/3313276.3316319, arXiv:1806.04256] 22

[56] JOSEPH NAOR AND MONI NAOR: Small-bias probability spaces: Efficient constructions and applications. *SIAM J. Comput.*, 22(4):838–856, 1993. [doi:10.1137/0222053] 8, 9

[57] NOAM NISAN: Pseudorandom bits for constant depth circuits. *Combinatorica*, 11(1):63–70, 1991. [doi:10.1007/BF01375474] 3, 5, 7

[58] NOAM NISAN AND AVI WIGDERSON: Hardness vs. randomness. *J. Comput. System Sci.*, 49(2):149–167, 1994. [doi:10.1016/S0022-0000(05)80043-1] 5, 7, 9, 10

[59] TONIANN PITASSI, PAUL BEAME, AND RUSSELL IMPAGLIAZZO: Exponential lower bounds for the pigeonhole principle. *Comput. Complexity*, 3(2):97–140, 1993. [doi:10.1007/BF01200117] 2

[60] TONIANN PITASSI, BENJAMIN ROSSMAN, ROCCO A. SERVEDIO, AND LI-YANG TAN: Polylogarithmic Frege depth lower bounds via an expander switching lemma. In *Proc. 48th STOC*, pp. 644–657. ACM Press, 2016. [doi:10.1145/2897518.2897637] 2

[61] ALEXANDER A. RAZBOROV: Bounded arithmetic and lower bounds in Boolean complexity. In *Feasible Mathematics II*, pp. 344–386. Springer, 1995. [doi:10.1007/978-1-4612-2566-9_12] 35

[62] ALEXANDER A. RAZBOROV: A simple proof of Bazzi's theorem. *ACM Trans. Comput. Theory*, 1(1):3:1–5, 2009. [doi:10.1145/1490270.1490273] 5, 6

[63] OMER REINGOLD, THOMAS STEINKE, AND SALIL VADHAN: Pseudorandomness for regular branching programs via Fourier analysis. In *Proc. 17th Internat. Workshop on Randomization and Computation (RANDOM'13)*, pp. 655–670. Springer, 2013. [doi:10.1007/978-3-642-40328-6_45] 22

[64] BENJAMIN ROSSMAN: On the constant-depth complexity of k-clique. In *Proc. 40th STOC*, pp. 721–730. ACM Press, 2008. [doi:10.1145/1374376.1374480] 2

[65] BENJAMIN ROSSMAN: The average sensitivity of bounded-depth formulas. *Comput. Complexity*, 27(2):209–223, 2018. [doi:10.1007/s00037-017-0156-0] 2

[66] ROCCO A. SERVEDIO AND LI-YANG TAN: What circuit classes can be learned with nontrivial savings? In *Proc. 8th Innovations in Theoret. Comp. Sci. conf. (ITCS'17)*. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2017. [doi:10.4230/LIPIcs.ITCS.2017.30] 4

[67] ROCCO A. SERVEDIO AND LI-YANG TAN: Luby–Veličković–Wigderson revisited: Improved correlation bounds and pseudorandom generators for depth-two circuits. In *Proc. 22nd Internat. Conf. on Randomization and Computation (RANDOM'18)*, pp. 56:1–20. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018. [doi:10.4230/LIPIcs.APPROX-RANDOM.2018.56] 9, 10

[68] ROCCO A. SERVEDIO AND LI-YANG TAN: Improved pseudorandom generators from pseudorandom multi-switching lemmas. In *Proc. 23rd Internat. Conf. on Randomization and Computation (RANDOM'19)*, pp. 45:1–23. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019. [doi:10.4230/LIPIcs.APPROX-RANDOM.2019.45] 1

[69] JIRÍ SÍMA AND STANISLAV ZÁK: A polynomial time construction of a hitting set for read-once branching programs of width 3, 2010/2021. [ECCC:TR10-088, arXiv:2101.01151] 5

[70] AVISHAY TAL: Tight bounds on the Fourier spectrum of AC^0. In *Proc. 32nd Comput. Complexity Conf. (CCC'17)*, pp. 15:1–31. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2017. [doi:10.4230/LIPIcs.CCC.2017.15] 5, 6, 7, 35

[71] NEIL THAPEN: Notes on switching lemmas, 2009. Posted on arXiv in 2022. [arXiv:2202.05651] 35

[72] LUCA TREVISAN: A note on approximate counting for k-DNF. In *Proc. 8th Internat. Workshop on Randomization and Computation (RANDOM'04)*, pp. 417–425. Springer, 2004. [doi:10.1007/978-3-540-27821-4_37] 5

[73] LUCA TREVISAN AND TONGKE XUE: A derandomized switching lemma and an improved derandomization of AC^0 . In *Proc. 28th IEEE Conf. on Comput. Complexity (CCC'13)*, pp. 242–247. IEEE Comp. Soc., 2013. [doi:10.1109/CCC.2013.32, ECCC:TR12-116] 3, 4, 5, 6, 7, 8, 14, 16, 19, 22, 26

[74] EMANUELE VIOLA: Pseudorandom bits for constant-depth circuits with few arbitrary symmetric gates. *SIAM J. Comput.*, 36(5):1387–1403, 2007. [doi:10.1137/050640941] 5, 8, 9

[75] EMANUELE VIOLA: On the power of small-depth computation. *Found. Trends Theor. Comp. Sci.*, 5(1):1–72, 2009. [doi:10.1561/0400000033] 8

[76] EMANUELE VIOLA: The sum of d small-bias generators fools polynomials of degree d. *Comput. Complexity*, 18(2):209–217, 2009. [doi:10.1007/s00037-009-0273-5] 8, 9, 10, 33

[77] EMANUELE VIOLA AND AVI WIGDERSON: Norms, XOR lemmas, and lower bounds for polynomials and protocols. *Theory of Computing*, 4(7):137–168, 2008. [doi:10.4086/toc.2008.v004a007] 8

[78] ANDREW YAO: Theory and applications of trapdoor functions. In *Proc. 23rd FOCS*, pp. 80–91. IEEE Comp. Soc., 1982. [doi:10.1109/SFCS.1982.45] 5

[79] ANDREW YAO: Separating the polynomial-time hierarchy by oracles. In *Proc. 26th FOCS*, pp. 1–10. IEEE Comp. Soc., 1985. [doi:10.1109/SFCS.1985.49] 2, 3, 4

AUTHORS

Rocco A. Servedio
Professor
Columbia University
rocco@cs.columbia.edu
http://www.cs.columbia.edu/~rocco

Li-Yang Tan
Assistant Professor
Stanford University
liyang@cs.stanford.edu
http://theory.stanford.edu/~liyang

ABOUT THE AUTHORS

ROCCO SERVEDIO is a professor in the Department of Computer Science at Columbia University. He graduated from Harvard University in 2001, where his Ph. D. was supervised by Les Valiant. He is interested in computational complexity theory, computational learning theory, randomness in computing, property testing, and other topics.

LI-YANG TAN is an assistant professor of computer science at Stanford University. He received his Ph. D. in Computer Science from Columbia University in 2014, advised by Rocco Servedio. His research is in theoretical computer science, with an emphasis on complexity theory.