View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Dagstuhl Research Online Publication Server

Improved Pseudorandom Generators from Pseudorandom Multi-Switching Lemmas

Rocco A. Servedio
Department of Computer Science, Columbia University, New York, NY, USA
http://www.cs.columbia.edu/~rocco
rocco@cs.columbia.edu

Li-Yang Tan
Department of Computer Science, Stanford University, Palo Alto, CA, USA
liyang@cs.stanford.edu

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: improving on the seed lengths of either PRG would require breakthrough progress on longstanding and notorious circuit lower bounds.

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 Paturi [39] and Håstad [35]. 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.

2012 ACM Subject Classification Theory of computation → Pseudorandomness and derandomization

Keywords and phrases pseudorandom generators, switching lemmas, circuit complexity, unconditional derandomization

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2019.45

Category RANDOM

Related Version A full version of the paper is available at https://arxiv.org/abs/1801.03590.

Funding *Rocco A. Servedio*: Supported by NSF grants CCF-1420349 and CCF-1563155.
Li-Yang Tan: Supported by NSF grant CCF-1563122; part of this research was done during a visit to Columbia University.

Acknowledgements We thank Prahladh Harsha and Srikanth Srinivasan for helpful discussions.

1 Introduction

Switching lemmas. Switching lemmas, first established in a series of breakthrough works in the 1980s [4, 29, 71, 34], 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


© Rocco A. Servedio and Li-Yang Tan;
licensed under Creative Commons License CC-BY
Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019).
Editors: Dimitris Achlioptas and László A. Végh; Article No. 45; pp. 45:1–45:23

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

are at the heart of several near-optimal bounds on AC^0 circuits, such as essentially optimal correlation bounds against the PARITY function [39, 35] and the worst-case and average-case depth hierarchy theorems of [34, 59, 36]. 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 [12, 57], lower bounds on the small-depth circuit complexity of ST-CONNECTIVITY [25], and lower bounds against AC^0 formulas [58]. Beyond the immediate arena of circuit lower bounds, switching lemmas are also important tools in diverse areas including propositional proof complexity [54, 43, 55, 37], computational learning theory [44], the design of circuit satisfiability algorithms [13, 39], and coding theory [23, 10].

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* [39, 35] 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 a coordinatewise-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, 24, 3, 32, 39, 31, 65, 30], 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 [65], known pseudorandom switching lemmas fell short of achieving the parameters of Håstad's influential "full randomness" switching lemma [34]. 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 works 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 works of [4, 29, 71] 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 [31] 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 [65] give a *pseudorandom* switching lemma that is highly randomness efficient and yet achieves the parameters of Håstad's fully random switching lemma (i.e. [65] achieves the same simplification, collapsing to a shallow decision tree, that follows from [34], with the same $*$-parameter p as [34]). The key conceptual ingredient enabling this is a beautiful idea of "fooling the proof" of the 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 [52] 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 for a detailed explanation.)

Via an ingenious extension of the ideas underlying the original switching lemma, Håstad [35] 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 [39] for closely related results). Roughly speaking, [35] 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 [39, 35] to achieve essentially optimal correlation bounds for AC^0 against the Parity function, capping off a long line of work [4, 71, 34, 19, 8, 13]. 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 [13, 39, 60].

[1] The first published version of the switching lemma with a decision tree conclusion is due to Cai [19]; several authors subsequently noted that Håstad's argument also yields such a conclusion.

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 14, 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 14 we "fool the proof" of Håstad's multi-switching lemma [35], exploiting its "computational simplicity". This enables us to achieve optimal parameters in the same sense as [65], namely, that it establishes the same dramatic simplification – now of the family $\mathcal{F}$ of depth-two circuits – as [35], and while only requiring the same $*$-probability p as [35]. 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 14, which we now 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: improving on the seed lengths of either PRG would require breakthrough progress on longstanding and notorious circuit lower bounds. In this sense, our results are in the same spirit as those of Imagliazzo, Meka, and Zuckerman [40], 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 [40].

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, 45, 52, 53, 50, 49, 41, 64, 66, 11, 56, 18, 42, 26, 2, 1, 62, 47, 28, 32, 31, 65, 30, 63, 33, 21]. 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 [52], which introduced ideas that enriched the surprising connections between pseudorandomness and computational hardness [14, 70, 53]. 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 work), this strongly motivates the goal of constructing, for every circuit class $\mathcal{C}$, unconditional PRGs for $\mathcal{C}$ that are best possible given the current best lower bounds for $\mathcal{C}$. In other words, this is the goal of achieving a *quantitatively optimal hardness to randomness conversion* for $\mathcal{C}$, converting "all the hardnesss" in our lower bounds for $\mathcal{C}$ into pseudorandomness for $\mathcal{C}$.

[2] While our focus in this work is on unconditional derandomization, we briefly mention that recent work of Ball et al. [10] establishes a new connection between pseudorandom switching lemmas and *non-malleable codes* in coding theory [27]. Using this connection, [10] are able to leverage the randomness efficiency of [65]'s pseudorandom switching lemma in their 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.

For $\mathcal{C}$ 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 best possible without improving longstanding AC^0 lower bounds that date back to the 1980s [34]. (More precisely, it is well known, see e.g. [65], that achieving seed length say $\log^{d-1.01}(Mn)\log(1/\varepsilon)$ would yield $\exp(\omega(n^{1/(d-1)})$ size lower bounds against depth-d AC^0 circuits, which is a barrier that has stood for over 30 years even in the $d = 3$ case.) We give the first construction of a PRG that achieves this seed length up to an *additive* absolute constant in the exponent of $\log(Mn)$:

▶ **Theorem 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 $\log^{d+O(1)}(Mn)\log(1/\varepsilon)$.*

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 work [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. [6]'s seed length was improved significantly in the celebrated work of Nisan [52], using what is now known as the Nisan–Wigderson framework [53], 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 [34] 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. [65, 33]). This is roughly quadratically worse than the sought-for $\log^{d-1}(Mn)\log(1/\varepsilon)$, the best that can be achieved assuming *only* current AC^0 circuit lower bounds.

Bounded independence fools AC^0. Nisan's seed length for AC^0 circuits stood unmatched 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 [45]. Braverman's breakthrough result [18] showed that polylog(n)-wise independence fools AC^0, which (along with standard constructions of k-wise independent distributions) 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 [63]. Recently, Harsha and Srinivasan [33] 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 [65] 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 Nisan's PRG the dependence on ε is suboptimal: $\log^{d+O(1)}(1/\varepsilon)$ 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, [65] adopts the strategy employed in the early work of Ajtai and Wigderson [6]. We describe this strategy in detail in the full version of this paper, 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 [34]. (And as we will soon discuss, in this work we obtain our PRG by instantiating the [6] framework with our derandomization of the [35] multi-switching lemmas.)

PRGs via polarizing random walks. Finally, in recent exciting work Chattopadhyay, Hatami, Hosseini, and Lovett [21] 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. [21] show that if a class $\mathcal{C}$ is closed under restrictions and has sufficiently strong Fourier concentration on low-degree coefficients, then almost k-wise independence suffice to yield a fractional PRG, which their random walk approach can then convert into a standard PRG against $\mathcal{C}$. Using Tal's sharp bounds [63] 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 [65], which has seed length $\log^{d+O(1)}(Mn/\varepsilon)$; Harsha and Srinivasan's improvement of Braverman's generator [33], which has seed length $\log^{3d+O(1)}(Mn)\log(1/\varepsilon)$; and the [21] PRG, which has seed length $O(\log(n/\varepsilon)(\log(\log(n)/\varepsilon))\cdot \log^{2d-2} M)$, i.e. essentially $\log^{2d-1}(Mn)\log^2(1/\varepsilon)$.

Theorem 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.

Our approach. Our approach draws on and unifies ideas in the works of [6, 65, 33] discussed above, which we use in conjunction with our derandomization of the [35] multi-switching lemma to obtain our PRG.

At a high level, we adopt the overall conceptual strategy of Ajtai and Wigderson [6] and Trevisan and Xue [65], 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 [39, 35] optimal correlation bounds for AC^0 against Parity. Our pseudorandom multi-switching lemma improves both the pseudorandom switching lemma of [65] (a derandomization of Håstad's

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

Reference	Seed length	Techniques
[6]	$n^{o(1)}$ for $M = \mathrm{poly}(n)$	derandomize [4] switching lemma
[52]	$\log^{2d+O(1)}(Mn/\varepsilon)$	[53] framework, [34] correlation bounds
[18]	$\log^{O(d^2)}(Mn/\varepsilon)$	bounded independence
[65]	$\log^{d+O(1)}(Mn/\varepsilon)$	[6] framework, derandomize [34] switching lemma
[63]	$\log^{3d+O(1)}(Mn/\varepsilon)$	bounded independence
[33]	$\log^{3d+O(1)}(Mn)\log(1/\varepsilon)$	bounded independence
[21]	(essentially) $\log^{2d-1}(Mn)\log^2(1/\varepsilon)$	almost bounded independence, fractional PRGs, polarizing random walks
This work	$\log^{d+O(1)}(Mn)\log(1/\varepsilon)$	[6] framework, derandomize [35] multi-switching lemma, bounded independence

switching lemma [34] 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 [35] multi-switching lemma is largely influenced by Trevisan and Xue's derandomization of the Håstad's original switching lemma [34]. We describe our approach in detail in Section 4, but highlight here the simple but ingenious new idea underlying [65]'s argument. Very roughly speaking, they derandomize the [34] 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 outputs 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 $F \restriction \rho$ has large depth). In similar spirit, our derandomization of the [35] multi-switching lemma also exploits the "computational simplicity" of its proof. In our case, for a fixed family $\mathcal{F}$ 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 $d = 3$ case of Harsha and Srinivasan's strengthening of Braverman's generator [33] to fool this depth-4 circuit, and hence show that [35]'s proofs of the multi-switching lemmas "cannot distinguish" between truly random and pseudorandom restrictions. The fact that [33] achieves an optimal $\log(1/\varepsilon)$ seed length dependence 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 [51, 7, 50, 15, 66, 46, 68, 16, 47, 48, 22].

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 [9, 17, 20, 69], and breaking this "degree $\log n$ barrier" would constitute a significant technical breakthrough[3]. See Open Question 1 of Viola's excellent survey [67] for a detailed discussion of this important open problem and its relationship with 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** (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 [50], 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 [53] together with multi-party number-on-the-forehead (NOF) communication complexity lower bounds from the seminal work of Babai, Nisan, and Szegedy [9]. Viola [66] 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 [61], the authors have improved the seed length dependence on ε of [50, 66] to $2^{O(\sqrt{\log(S)})} + \mathrm{polylog}(1/\varepsilon)$. We discuss the relation between our techniques and those of [61] 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 [51] 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 [15, 16] and culminated in the breakthroughs of Lovett [46] and Viola [68] which are described in more detail below. It is interesting to note that prior to our work, the underlying techniques used for the sparse case (multi-party communication complexity) are completely different from the techniques used for the low-degree case (Fourier analysis).

Our PRG and approach. Theorem 2 gives an exponential and optimal improvement of the PRG of [50] 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 $\log(n)$-degree $\mathbb{F}_2$ polynomial has at most $n^{\log n}$ monomials, it can be shown (using the simple Proposition 3.1 of [68]) that a PRG with seed length $2^{o(\sqrt{\log S})}\log(1/\varepsilon)$ would break the degree $\log n$ barrier.

[3] Breaking this "degree $\log n$ barrier" is also well-known (via a simple and beautiful observation of Håstad and Goldmann [38]) to be a prerequisite for breaking the notorious "$\log n$ party barrier" in multi-party communication complexity [9], a longstanding open problem that has resisted attack for over two decades.

Our techniques for Theorem 2 are substantially different from the techniques of [61, 66]. As summarized in Table 2, the basic approach of [61], like [66] and [50], is via the Nisan-Wigderson paradigm using multi-party communication complexity bounds; the main point of departure between [61] and [66] is that [61] leverages Håstad's multi-switching lemma from [35] in place of his earlier [34] switching lemma which was used in [66]. (We note that similar 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 [35] multi-switching lemma. Indeed, our approach to obtaining Theorem 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 degree-$\sqrt{\log S}$ polynomials. This allows us to leverage the result of Viola [68] (building on the work of Lovett [46]), which gives PRGs for n-variable degree-k $\mathbb{F}_2$ polynomials with seed length

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

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.

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

Reference/ Class	Seed length	Techniques
[50] S sparse	$2^{O(\sqrt{\log(S/\varepsilon)})}$	[53] framework, [9] multi-party NOF communication complexity
[61] S sparse	$2^{O(\sqrt{\log S})} + (\log(1/\varepsilon))^{4.01}$	[53] framework, [9] multi-party NOF communication complexity, [35] multi-switching lemma
[46] degree k	$O(2^k \log n + 4^k \log(1/\varepsilon))$	Fourier analysis
[68] degree k	$O(k \log n + k2^k \log(1/\varepsilon))$	Fourier analysis
This work S sparse	$2^{O(\sqrt{\log S})} \log(1/\varepsilon)$	[6] framework, derandomize [35] multi-switching lemma, Fourier analysis, bounded independence

2.3 Organization

Section 2.4 recalls some basic preliminaries from unconditional pseudorandomness. We describe and contrast the original Håstad switching lemma [34] versus the [35] multi-switching lemma in Section 3. Section 3.1 establishes some infrastructure towards derandomizing the [35] switching lemma, and the actual derandomization result (the pseudorandom multi-switching lemma, Theorem 14) is stated in Section 4 and proved in Appendix A. In the full version we describe a general framework for constructing pseudorandom generators that is implicit in the work of Ajtai and Wigderson [6], and explain how our derandomized multi-switching lemma from Section 4 can be used (along with other ingredients) within this framework to establish the PRGs for AC^0 and for sparse $\mathbb{F}_2$ polynomials that are our main PRG results.

2.4 Preliminaries

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 $z^{(1)}, \ldots, z^{(s)}$ of strings from $\{0,1\}^n$ where $s \in [\frac{1}{\mathrm{poly}(n)} \cdot 2^r, 2^r]$ 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 outputs a string drawn from $\mathcal{D}$.

For $\delta > 0$ and a class $\mathcal{C}$ of functions from $\{0,1\}^n$ to $\{0,1\}$, we say that a distribution $\mathcal{D}$ over $\{0,1\}^n$ *δ-fools $\mathcal{C}$ 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 $f \in \mathcal{C}$, we have

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

Equivalently, we say that $\mathrm{Gen}_{\mathcal{D}}$ is a *δ-PRG for $\mathcal{C}$ 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 $1 \leq i_1 < \cdots < i_k \leq n$ and every $(b_1, \ldots, b_k) \in \{0,1\}^k$, we have

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

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

A *restriction* ρ of variables $x_1, \ldots, x_n$ 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 $f(x_1, \ldots, x_n)$ and a restriction ρ, we write $f \restriction \rho$ 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 two restrictions $\rho, \rho' \in \{0,1,*\}^n$, their *composition*, denoted $\rho\rho' \in \{0,1,*\}^n$, is the restriction 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 $\mathcal{F} = \{f_1, \ldots, f_M\}$ of functions and a restriction ρ we write $\mathcal{F} \restriction \rho$ to denote the family $\{f_1 \restriction \rho, \ldots, f_M \restriction \rho\}$.

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 $\mathcal{C}$ is a class of circuits, we say that g is *computed by a* $(t, \mathcal{C})$*-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 $\mathcal{C}$.

3 Multi-switching lemmas

At the heart of almost all applications of Håstad's original switching lemma [34] 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 high 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 [34] switching lemma, and indeed this will be clear from the discussion later in this section.

The recent multi-switching lemma of Håstad [35] (see also [39]) 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 [39, 35]; 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 [34]:

▶ **Theorem 3** (Håstad's switching lemma). *Let F be a k-CNF. Then for all $t \geq 1$, we have that*

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

In the context of AC^0 circuits the switching lemma is used to achieve *depth reduction* under random restrictions: we apply Theorem 3 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 4** (AC^0 depth reduction via Theorem 3). *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 $\varepsilon > 0$,*

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

Proof. This follows from applying Theorem 3 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. ◀

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 4, 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. This is why standard applications of the [34] switching lemma (involving $d-1$ iterative applications of Corollary 4) 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 [35] 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 5** (Common partial decision tree). *Let $\mathcal{F} = \{F_1, \dots, F_M\}$ be a collection of Boolean functions. We say that a decision tree T is a* common ℓ-partial decision tree for $\mathcal{F}$ *if every $F_i \in \mathcal{F}$ can be expressed as T with depth-ℓ decision trees at its its leaves. (Equivalently, for every $F_i \in \mathcal{F}$ and root-to-leaf path π in T, we have that $F_i \restriction \pi$ is computed by a depth-ℓ decision tree.)*

The multi-switching lemma of [35] is as follows:

▶ **Theorem 6** (Multi-switching lemma, Lemma 3.8 of [35]). *Let $\mathcal{F} = \{F_1, \dots, F_M\}$ be a collection of k-CNFs and $\ell := \log(2M)$. Then for all $t \geq 1$,*

$$\Pr_{\boldsymbol{\rho} \leftarrow \mathcal{R}_p}[\mathcal{F} \restriction \boldsymbol{\rho} \text{ does not have a common } \ell := \log(2M)\text{-partial DT of depth } t] \leq M(24pk)^t.$$

The following corollary should be contrasted with Corollary 4:

▶ **Corollary 7** (AC^0 depth reduction via Theorem 6; c.f. Corollary 4). *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 $\varepsilon > 0$, the probability (over $\boldsymbol{\rho} \leftarrow \mathcal{R}_p$) that $\mathcal{C} \restriction \boldsymbol{\rho}$ is not computed by a $((\log(M/\varepsilon), \mathsf{AC}^0(\text{depth } d-1, \text{ bottom fan-in } \log(2M))$-decision tree is at most ε.*

Proof. This follows by applying Theorem 6 with $\mathcal{F}$ 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. ◀

We highlight a crucial qualitative aspect of Corollary 7: while the depth $t = \log(M/\varepsilon)$ of the decision tree whose existence it asserts 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 7, 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 4: $d-1$ iterative applications of Corollary 7 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 $k \in \mathbb{N}$,

$$\mathop{\mathbf{Pr}}_{\boldsymbol{\rho} \leftarrow \mathcal{R}_{\frac{p}{2}}} [\, P \restriction \boldsymbol{\rho} \text{ is not a degree-}k \text{ polynomial} \,] \leq \frac{\varepsilon}{2} + S\binom{w}{k} p^k \quad \text{where } w = \Theta(\log(S/\varepsilon)). \quad (1)$$

(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 (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 [35], we can instead bound the probability that $P \restriction \boldsymbol{\rho}$ 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 4 vs. Corollary 7 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 6 is that of a *canonical* common ℓ-partial decision tree for an ordered collection $\mathcal{F}$ 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 $\mathrm{CDT}(F \restriction \eta)$.

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) actually shows that if F is a k-CNF, then the canonical decision tree $\mathrm{CDT}(F \restriction \boldsymbol{\rho})$ is shallow w.h.p. over $\boldsymbol{\rho} \leftarrow \mathcal{R}_p$. This is crucially important for the arguments of Trevisan and Xue [65], 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 [35], we observe that analogous to his original switching lemma, the proof of Theorem 6 given in [35] implicitly establishes a stronger statement: $\mathcal{F} \restriction \boldsymbol{\rho}$ has a small-depth *canonical* common ℓ-partial decision tree w.h.p. over $\boldsymbol{\rho} \leftarrow \mathcal{R}_p$. In fact, we will use the fact that it actually establishes an even stronger statement: w.h.p. over $\boldsymbol{\rho} \leftarrow \mathcal{R}_p$, *every* canonical common ℓ-partial decision tree for $\mathcal{F} \restriction \boldsymbol{\rho}$ is shallow – as we explain below, there is more than one canonical common ℓ-partial decision tree for a sequence $\mathcal{F}$ of CNFs.

Let us explain what a canonical common ℓ-partial decision tree for a sequence of CNFs $\mathcal{F}$ is. We will see that there is a set of canonical common ℓ-partial decision trees for a given $\mathcal{F}$ rather than just one tree; note that this is the case even though we assume a fixed ordering $F_1, F_2, \ldots$ on the elements of $\mathcal{F}$ 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 8** (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 $C_{i_j} \restriction \eta^{(1)} \circ \cdots \circ \eta^{(j-1)}$, and C_{i_j} is the first clause in $F \restriction \eta^{(1)} \circ \cdots \circ \eta^{(j-1)}$ 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 $C_{i_j} \restriction \eta^{(1)} \circ \cdots \circ \eta^{(j-1)}$. 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 $C_{i_u} \restriction \eta^{(1)} \circ \cdots \circ \eta^{(u-1)}$.*

▶ **Observation 9.** *Let F be a k-CNF and suppose $\mathrm{depth}(\mathrm{CDT}(F)) > \ell$. Then there is a full path η of length $|\eta| \in \{\ell+1, \ldots, \ell+k\}$ 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 10** (Canonical common ℓ-partial DT). *Let $\mathcal{F} = (F_1, \ldots, F_M)$ be an ordered collection of k-CNFs. The set of all* canonical *common ℓ-partial decision trees for $\mathcal{F}$, which we denote $\mathrm{CCDT}_\ell(\mathcal{F})$, is defined inductively as follows:*

0. *If $M = 0$ (i.e. $\mathcal{F}$ is an empty collection of k-CNFs) then $\mathrm{CCDT}_\ell(\mathcal{F})$ 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 $\mathcal{F}$.)*
1. *If $\mathrm{CDT}(F_1) \leq \ell$, then $\mathrm{CCDT}_\ell(\mathcal{F})$ is simply $\mathrm{CCDT}_\ell(\mathcal{F}')$, where $\mathcal{F}' = (F_2, \ldots, F_M)$. (Note that in this case, since inductively each tree in $\mathrm{CCDT}_\ell(\mathcal{F}')$ is a common ℓ-partial DT for $\mathcal{F}'$, each such tree is also a common ℓ-partial DT for $\mathcal{F}$.)*
2. *Otherwise, since $\mathrm{CDT}(F_1) > \ell$ 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 $\mathrm{CCDT}_\ell(\mathcal{F} \restriction \pi)$ at the end of the path. So in this case $\mathrm{CCDT}_\ell(\mathcal{F})$ 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 $T' \in \mathrm{CCDT}_\ell(\mathcal{F} \restriction \pi)$ for each path $\pi \in T_\eta$).*

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

The following slight variant of Theorem 6 can be extracted, with some effort, from a slight modification of the proof given in [35], which we provide in the full version:

▶ **Theorem 11** (Slight variant of Håstad's multi-switching lemma. Theorem 6). *Let $\mathcal{F} = (F_1, \ldots, F_M)$ be an ordered collection of k-CNFs. Then for all $\ell, t \geq 1$,*

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

A comparison of Theorem 6 (Håstad's multi-switching lemma) and Theorem 11 (our variant of it). We emphasize that the differences are technical in nature, and all the ideas in our proof of Theorem 11 are from [35]. 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 [35]: in case 2 of Definition 10, we query a witnessing full path of length between $\ell+1$ and $\ell+k$, whereas [35] 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 of Theorem 11: 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 $\boldsymbol{\rho}$ drawn from this pseudorandom distribution satisfies a similar guarantee to Theorem 11.

Our derandomization of Theorem 11 is largely influenced by Trevisan and Xue's [65] ingenious derandomization of Håstad's original switching lemma (Theorem 3). Roughly speaking, we will derandomize the multi-switching lemma of Theorem 11 by "fooling its proof": we will show that the proof of Theorem 11 (given in the full version, which we again emphasize is only a slight technical modification of Håstad's proof of his multi-switching lemma, Theorem 6) "cannot δ-distinguish" between truly random restrictions and pseudorandom restrictions drawn from polylog(n)-wise independent distributions. Since Theorem 11 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, we exploit the "computational simplicity" of Theorem 11's proof: for a fixed family $\mathcal{F}$ 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 outputs 1 iff ρ is a bad restriction for the desired conclusion of Theorem 11, contributing to its failure probability (i.e. iff $\mathrm{depth}(\mathrm{CCDT}_\ell(\mathcal{F} \restriction \rho)) > t$). As alluded to in Section 3.1, this relies on the fact that Theorem 11 does not simply bound the depth of the *optimal* common ℓ-partial decision tree for $\mathcal{F} \restriction \boldsymbol{\rho}$, but instead the depth of any *canonical* common ℓ-partial decision tree for $\mathcal{F} \restriction \boldsymbol{\rho}$. 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 $\mathcal{F} \restriction \rho$ has depth greater than t.

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

▶ **Definition 12** (Representing restrictions as bitstrings). *We associate with each string $(\varrho, y) \in \{0,1\}^{Y_q}$ the restriction $\rho(\varrho, y) \in \{0,1,*\}^n$ defined as follows:*

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

The following observation explains the role of q:

▶ **Observation 13.** *Let $(\boldsymbol{\varrho}, \boldsymbol{y})$ be drawn from the uniform distribution over $\{0,1\}^{Y_q}$. Then the random restriction $\rho(\boldsymbol{\varrho}, \boldsymbol{y}) \in \{0,1,*\}^n$ is distributed according to $\mathcal{R}_p$ where $p = 2^{-q}$.*

Now we are ready to state our pseudorandom multi-switching lemma:

▶ **Theorem 14** (Derandomized version of Theorem 11). *Let $\mathcal{F} = (F_1, \ldots, F_M)$ be an ordered list of Q-clause k-CNFs. Let $\delta, p \in (0,1)$ and define $q = \log(1/p)$. Let $\mathcal{D}$ be any distribution over $\{0,1\}^{Y_q}$ that $(\delta/(M^{\lceil t/\ell \rceil} n^{O(t)}))$-fools the class of depth-3 circuits of size $M(n^{O(\ell)} + Q2^{O(kq)})$. Then for all $\ell \geq k$ and all $t \in \mathbb{N}$,*

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

In the full version of the paper we prove this lemma and show how it, along with other ingredients, yields our circuit complexity derandomization results.

References

1 Scott Aaronson. A Counterexample to the Generalized Linial–Nisan Conjecture. *Electronic Colloquium on Computational Complexity*, 17:109, 2010.

2 Scott Aaronson. BQP and the polynomial hierarchy. In *Proceedings of the 42nd ACM Symposium on Theory of Computing*, pages 141–150, 2010.

3 Manindra Agrawal, Eric Allender, Russell Impagliazzo, Toniann Pitassi, and Steven Rudich. Reducing the complexity of reductions. *Comput. Complexity*, 10(2):117–138, 2001.

4 Miklós Ajtai. Σ^1_1-formulae on finite structures. *Annals of Pure and Applied Logic*, 24(1):1–48, 1983.

5 Miklós Ajtai. Geometric properties of sets defined by constant depth circuits. In *Combinatorics, Paul Erdős is eighty, Vol. 1*, Bolyai Soc. Math. Stud., pages 19–31. János Bolyai Math. Soc., Budapest, 1993.

6 Miklós Ajtai and Avi Wigderson. Deterministic Simulation of Probabilistic Constant Depth Circuits. In *Proceedings of the 26th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 11–19, 1985.

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

8 László Babai. Random oracles separate PSPACE from the polynomial-time hierarchy. *Information Processing Letters*, 26(1):51–53, 1987.

9 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.

10 Marshall Ball, Dana Dachman-Soled, Siyao Guo, Tal Malkin, and Li-Yang Tan. Non-malleable codes for small-depth circuits. In *Proceedings of the 59th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2018. To appear.

11 Louay Bazzi. Polylogarithmic independence can fool DNF formulas. *SIAM Journal on Computing*, 38(6):2220–2272, 2009.

12 Paul Beame. Lower bounds for recognizing small cliques on CRCW PRAM's. *Discrete Applied Mathematics*, 29(1):3–20, 1990.

13 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 *Proceedings of the 27th IEEE Conference on Computational Complexity (CCC)*, pages 117–125, 2012.

14 Manuel Blum and Silvio Micali. How to Generate Cryptographically Strong Sequences of Pseudo Random Bits. In *Proceedings of the 23rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 112–117, 1982.

15 Andrej Bogdanov. Pseudorandom generators for low degree polynomials. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing (STOC)*, pages 21–30. SIAM, 2005. doi:10.1145/1060590.1060594.

16 Andrej Bogdanov and Emanuele Viola. Pseudorandom bits for polynomials. *SIAM J. Comput.*, 39(6):2464–2486, 2010. doi:10.1137/070712109.

17 Jean Bourgain. Estimation of certain exponential sums arising in complexity theory. *Comptes Rendus Mathematique*, 340(9):627–631, 2005. doi:10.1016/j.crma.2005.03.008.

18 Mark Braverman. Polylogarithmic independence fools AC^0 circuits. *Journal of the ACM*, 57(5):28, 2010.

19 Jin-Yi Cai. With Probability One, a Random Oracle Separates PSPACE from the Polynomial-Time Hierarchy. In *Proceedings of the 18th Annual ACM Symposium on Theory of Computing (STOC)*, pages 21–29, 1986.

20 Arkadev Chattopadhyay. Discrepancy and the power of bottom fan-in in depth-three circuits. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 449–458, 2007.

21 Eshan Chattopadhyay, Pooya Hatami, Kaave Hosseini, and Shachar Lovett. Pseudorandom Generators from Polarizing Random Walks. In *33rd Computational Complexity Conference, CCC*, pages 1:1–1:21, 2018.

22 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 *10th Innovations in Theoretical Computer Science Conference (ITCS)*, pages 22:1–22:15, 2019.

23 Eshan Chattopadhyay and Xin Li. Non-malleable codes and extractors for small-depth circuits, and affine functions. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 1171–1184. ACM, 2017.

24 Shiva Chaudhuri and Jaikumar Radhakrishnan. Deterministic restrictions in circuit complexity. In *Proceedings of the 28th Annual ACM Symposium on Theory of Computing (STOC)*, pages 30–36, 1996.

25 Xi Chen, Igor Carboni Oliveira, Rocco A. Servedio, and Li-Yang Tan. Near-optimal small-depth lower bounds for small distance connectivity. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing (STOC)*, pages 612–625, 2016.

26 Anindya De, Omid Etesami, Luca Trevisan, and Madhur Tulsiani. Improved pseudorandom generators for depth 2 circuits. In *Proceedings of the 13th International Workshop on Randomization and Computation (RANDOM)*, pages 504–517, 2010.

27 Stefan Dziembowski, Krzysztof Pietrzak, and Daniel Wichs. Non-Malleable Codes. *Journal of the ACM (JACM)*, 65(4):20, 2018.

28 Bill Fefferman, Ronen Shaltiel, Christopher Umans, and Emanuele Viola. On beating the hybrid argument. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference (ITCS)*, pages 468–483. ACM, 2012.

29 Merrick Furst, James Saxe, and Michael Sipser. Parity, circuits, and the polynomial-time hierarchy. *Mathematical Systems Theory*, 17(1):13–27, 1984.

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

31 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.

32 Parikshit Gopalan, Raghu Meka, Omer Reingold, Luca Trevisan, and Salil P. Vadhan. Better Pseudorandom Generators from Milder Pseudorandom Restrictions. In *Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 120–129, 2012.

33 Prahladh Harsha and Srikanth Srinivasan. On Polynomial Approximations to AC^0. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2016*, pages 32:1–32:14, 2016.

34 Johan Håstad. Almost optimal lower bounds for small depth circuits. In *Proceedings of the 18th Annual ACM Symposium on Theory of Computing (STOC)*, pages 6–20, 1986.

35 Johan Håstad. On the Correlation of Parity and Small-Depth Circuits. *SIAM Journal on Computing*, 43(5):1699–1708, 2014.

36 Johan Håstad. An Average-Case Depth Hierarchy Theorem for Higher Depths. In *Proceedings of the 57th Annual Symposium on Foundations of Computer Science (FOCS)*, 2016.

37 Johan Håstad. On small-depth Frege proofs for Tseitin for grids. In *Proceedings of the 58th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 97–108. IEEE Computer Society, 2017.

38 Johan Håstad and Mikael Goldmann. On the power of small-depth threshold circuits. *Comput. Complexity*, 1(2):113–129, 1991. doi:10.1007/BF01272517.

39 Russell Impagliazzo, William Matthews, and Ramamohan Paturi. A satisfiability algorithm for AC^0. In *Proceedings of the 23rd Annual Symposium on Discrete Algorithms (SODA)*, pages 961–972, 2012.

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

41 Adam Klivans. On the Derandomization of Constant Depth Circuits. In *Proceedings of 5th International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM)*, pages 249–260, 2001.

42 Adam Klivans, Homin Lee, and Andrew Wan. Mansour's Conjecture is True for Random DNF Formulas. In *Proceedings of the 23rd Conference on Learning Theory (COLT)*, pages 368–380, 2010.

43 Jan Krajíček, Pavel Pudlák, and Alan Woods. An exponential lower bound to the size of bounded depth Frege proofs of the pigeonhole principle. *Random Structures & Algorithms*, 7(1):15–39, 1995.

44 Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, Fourier transform and learnability. *Journal of the ACM*, 40(3):607–620, 1993.

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

46 Shachar Lovett. Unconditional pseudorandom generators for low-degree polynomials. *Theory Comput.*, 5:69–82, 2009. doi:10.4086/toc.2009.v005a003.

47 Shachar Lovett and Srikanth Srinivasan. Correlation bounds for poly-size AC^0 circuits with $n^{1-o(1)}$ symmetric gates. In *Approximation, randomization, and combinatorial optimization*, volume 6845 of *Lecture Notes in Comput. Sci.*, pages 640–651. Springer, Heidelberg, 2011. doi:10.1007/978-3-642-22935-0_54.

48 Chi-Jen Lu. Hitting set generators for sparse polynomials over any finite fields. In *Proceedings of the 27th IEEE Conference on Computational Complexity (CCC)*, pages 280–286, 2012. doi:10.1109/CCC.2012.20.

49 Michael Luby and Boban Veličković. On deterministic approximation of DNF. *Algorithmica*, 16(4-5):415–433, 1996. doi:10.1007/s004539900054.

50 Michael Luby, Boban Veličković, and Avi Wigderson. Deterministic approximate counting of depth-2 circuits. In *Proceedings of the 2nd ISTCS*, pages 18–24, 1993.

51 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.

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

53 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.

54 Toniann Pitassi, Paul Beame, and Russell Impagliazzo. Exponential lower bounds for the pigeonhole principle. *Computational complexity*, 3(2):97–140, 1993.

55 Toniann Pitassi, Benjamin Rossman, Rocco A. Servedio, and Li-Yang Tan. Poly-logarithmic Frege depth lower bounds via an expander switching lemma. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing (STOC)*, pages 644–657, 2016.

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

57 Benjamin Rossman. On the constant-depth complexity of k-clique. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing (STOC)*, pages 721–730, 2008.

58 Benjamin Rossman. The Average Sensitivity of Bounded-Depth Formulas. In *Proceedings of the 56th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 424–430, 2015.

59 Benjamin Rossman, Rocco A. Servedio, and Li-Yang Tan. An Average-Case Depth Hierarchy Theorem for Boolean Circuits. In *Proceedings of the 56th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1030–1048, 2015.

60 Rocco A. Servedio and Li-Yang Tan. What circuit classes can be learned with nontrivial savings? In *Proceedings of the 8th Innovations in Theoretical Computer Science Conference (ITCS)*, 2017.

61 Rocco A. Servedio and Li-Yang Tan. Luby–Veličković–Wigderson revisited: Improved correlation bounds and pseudorandom generators for depth-two circuits. In *Proceedings of the 22nd International Workshop on Randomization and Computation (RANDOM)*, pages 56:1–56:20, 2018.

62 Jirí Síma and Stanislav Zák. A Polynomial Time Construction of a Hitting Set for Read-Once Branching Programs of Width 3. *Electronic Colloquium on Computational Complexity (ECCC)*, 17:88, 2010.

63 Avishay Tal. Tight Bounds on the Fourier Spectrum of AC^0. In *Proceedings of the 32nd Computational Complexity Conference (CCC)*, pages 15:1–15:31, 2017. doi:10.4230/LIPIcs.CCC.2017.15.

64 Luca Trevisan. A note on approximate counting for k-DNF. In *Proceedings of the 8th International Workshop on Randomization and Computation (RANDOM)*, pages 417–426, 2004.

65 Luca Trevisan and Tongke Xue. A derandomized switching lemma and an improved derandomization of AC^0 . In *Proceedings of the 28th IEEE Conference on Computational Complexity (CCC)*, pages 242–247, 2013.

66 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.

67 Emanuele Viola. *On the power of small-depth computation*. Now Publishers Inc, 2009.

68 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.

69 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.

70 Andrew Yao. Theory and applications of trapdoor functions. In *Proceedings of the 23rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 80–91, 1982.

71 Andrew Yao. Separating the polynomial-time hierarchy by oracles. In *Proceedings of the 26th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1–10, 1985.

A Proof of Theorem 14

A.1 Bad restrictions and the structure of witnessing paths

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

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

Fix ρ to be a bad restriction. Recalling our definition of the set of canonical common partial decision trees (Definition 10), there exists a tree $T \in \mathrm{CCDT}_\ell(\mathcal{F} \restriction \rho)$ and a path Π of length exactly t through T. Furthermore, we have that

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

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

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

1. *An ordered list of indices $\mathcal{I} = (i_1, \ldots, i_u)$ where $1 \le i_1 \le \cdots \le i_u \le M$ and $u \le \lceil t/\ell \rceil$,*
2. *For each index $i_j \in \mathcal{I}$, a subset $S_j \subseteq [n]$ such that*
   a. *These sets are mutually disjoint: $S_j \cap S_{j'} = \emptyset$ for all $j \ne j'$.*
   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 $T \in \mathrm{CCDT}_\ell(\mathcal{F} \restriction \rho)$, every path Π of length t through $\mathrm{CCDT}_\ell(\mathcal{F} \restriction \rho)$ uniquely induces an ℓ-segmented $\mathcal{F}$-traversal P of length t. We say that P *occurs in* $\mathrm{CCDT}_\ell(\mathcal{F} \restriction \rho)$ if it is induced by some path Π of length t through T for some $T \in \mathrm{CCDT}_\ell(\mathcal{F} \restriction \rho)$.

Definition 15 immediately yields the following:

▶ **Proposition 16** (Number of $\mathcal{F}$-traversals). *Fix an ordered list $\mathcal{F} = (F_1, \ldots, F_M)$ of k-CNFs, and let $\mathcal{P}_{\mathcal{F},\ell,t}$ denote the collection of all ℓ-segmented $\mathcal{F}$-traversals of length t. Then*

$$|\mathcal{P}_{\mathcal{F},\ell,t}| \le M^{\lceil t/\ell \rceil} n^{O(t)}.$$

A.2 A small AC^0 circuit for recognizing bad restrictions

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

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

by considering

$$\mathcal{C}_{\mathcal{F},\ell,t}(\varrho, y) := \bigvee_{P \in \mathcal{P}_{\mathcal{F},\ell,t}} \mathcal{C}_P(\varrho, y) \tag{2}$$

we have that

$$\rho(\varrho, y) \text{ is bad} \iff \mathcal{C}_{\mathcal{F},\ell,t}(\varrho, y) = 1.$$

▷ Claim 17 (Circuit for a single $\mathcal{F}$-traversal). Let $P = (\mathrm{I}, \{S_1, \ldots, S_u\}, \Pi, \mathrm{H})$ be an ℓ-segmented $\mathcal{F}$-traversal of length t. There is a depth-3 AND-OR-AND circuit $\mathcal{C}_P : \{0,1\}^{Y_q} \to \{0,1\}$ of size $M(n^{O(\ell)} + Q2^{O(kq)})$ such that

$$\forall\, (\varrho, y) \in \{0,1\}^{Y_q} : \quad \mathcal{C}_P(\varrho, y) = 1 \iff P \text{ occurs in } \mathrm{CCDT}_\ell(\mathcal{F} \restriction \rho(\varrho, y))$$

Proof. Our circuit $\mathcal{C}_P$ will be the AND of M many depth-3 subcircuits of size $n^{O(\ell)}$, one for each k-CNF $F \in \mathcal{F}$. 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 [65].)

1. **First type: Circuits checking that a particular restriction η is 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 $Q2^{O(kq)}$-clause $O(kq)$-CNF G over $\{0,1\}^{Y_q}$ that outputs 1 on input (ϱ, y) iff η is a path in $\mathrm{CDT}(F' \restriction \rho(\varrho, y))$.
   For each $i \in [Q]$, we write Fixed_i to denote the set

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

   of all variables that are fixed by η and occur in C_i. We write $\sigma^{(i)} \in \{0,1\}^{\mathrm{Fixed}_i}$ to denote η restricted to the coordinates in Fixed_i. It is straightforward to verify that η is a path in $\mathrm{CDT}(F' \restriction \rho(\varrho, y))$ iff for all $i \in [Q]$ such that $\mathrm{Fixed}_1 \cup \cdots \cup \mathrm{Fixed}_{i-1} \subsetneq \mathrm{supp}(\eta)$,
   - **a.** If $\mathrm{Fixed}_i \setminus (\mathrm{Fixed}_1 \cup \cdots \cup \mathrm{Fixed}_{i-1}) = \emptyset$ then the clause C_i is satisfied by $\rho(\varrho, y) \circ \sigma^{(1)} \circ \cdots \circ \sigma^{(i-1)}$. (Hence this clause does not contribute to $\mathrm{CDT}(F' \restriction \rho(\varrho, y))$; it is "skipped" in the canonical decision tree construction process.)
   - **b.** Otherwise, writing $\mathrm{Fixed}'_i := \mathrm{Fixed}_i \setminus (\mathrm{Fixed}_1 \cup \cdots \cup \mathrm{Fixed}_{i-1})$,
     - **i.** $\rho(\varrho, y)_j = *$ for all $j \in \mathrm{Fixed}'_i$, and
     - **ii.** $\rho(\varrho, y) \circ \sigma_1 \circ \cdots \circ \sigma_{i-1}$ falsifies all the remaining literals in C_i and are not in Fixed'_i.
     In other words, the clause

     $$C_i \restriction \rho(\varrho, y) \circ \sigma^{(1)} \circ \cdots \circ \sigma^{(i-1)}$$

     is not satisfied and its surviving variables are precisely those in Fixed'_i. (Hence the variables in Fixed'_i are exactly those queried by the canonical decision tree construction process when it reaches C_i.)

   Since both conditions (a) and (b) depend only on the coordinates of $\rho(\varrho, y)$ that occur in C_i (at most k such coordinates since C_i has width at most k), and hence at most $k(q+1)$ coordinates of $(\varrho, y) \in \{0,1\}^{Y_q}$, it is clear that both conditions can be checked by a $2^{O(kq)}$-clause $O(kq)$-CNF over $\{0,1\}^{Y_q}$. The overall CNF G is simply the AND of all Q many of these CNFs, one for each clause C_i of F', and hence G 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 $((2n)^{\ell+1}, Q2^{O(kq)}, O(kq))$ that outputs 1 on input (ϱ, y) iff $\mathrm{depth}(\mathrm{CDT}(F' \restriction \rho(\varrho, y))) \leq \ell$.
   We establish this by showing that there is a depth-3 OR-AND-OR circuit Σ with the claimed fan-in sequence that outputs 1 on input (ϱ, y) if $\mathrm{depth}(\mathrm{CDT}(F' \restriction \rho(\varrho, y))) > \ell$; given such a circuit Σ, the desired AND-OR-AND circuit is obtained by negating Σ and using de Morgan's law. Certainly $\mathrm{depth}(\mathrm{CDT}(F' \restriction \rho(\varrho, y))) > \ell$ iff there is a path η of

length $\ell+1$ in $\mathrm{CDT}(F' \restriction \rho(\varrho, y))$. 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 path η there is a $Q2^{O(kq)}$-clause, $O(kq)$-CNF over $\{0,1\}^{Y_q}$ that checks if η is a path in $\mathrm{CDT}(F' \restriction \rho(\varrho, y))$. 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 M many depth-3 subcircuits, one for each k-CNF $F \in \mathcal{F}$:

- For each of the u indices $i_j \in \mathcal{I}$, a circuit of the first type that checks that $\eta^{(j)}$ is a path in $\mathrm{CDT}(F_{i_j} \restriction \rho(\varrho, y) \circ \pi^{(1)} \circ \cdots \circ \pi^{(j-1)})$ (recall from Definition 15 that $\eta^{(j)}$ is H restricted to the variables in S_j);
- For all $M-u$ other indices $i \in [M] \setminus \mathcal{I}$, a circuit of the second type that checks that $\mathrm{depth}(\mathrm{CDT}(F_i \restriction \rho(\varrho, y) \circ \pi^{(1)} \circ \cdots \circ \pi^{(i^-)})) \le \ell$, 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. ◁

A.3 Putting the pieces together: Proof of Theorem 14

Recalling the definition (2) of $\mathcal{C}_{\mathcal{F},\ell,t}$,

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

Proposition 16 giving a bound on its top fan-in, and Claim 17 giving a bound on the size of its subcircuits, we have shown the following:

▷ Claim 18 (Circuit for recognizing bad restrictions). Let $\mathcal{F} = (F_1, \ldots, F_M)$ be an ordered list of Q-clause k-CNFs, and let $\ell, t \ge 1$. There is a depth-4 circuit $\mathcal{C}_{\mathcal{F},\ell,t}$ over $\{0,1\}^{Y_q}$ such that

$$\mathcal{C}_{\mathcal{F},\ell,t}(\varrho, y) = 1 \iff \mathrm{depth}(\mathrm{CCDT}_\ell(\mathcal{F} \restriction \rho(\varrho, y))) \ge t.$$

This circuit $\mathcal{C}_{\mathcal{F},\ell,t}$ is the OR of $M^u n^{O(t)}$ many depth-3 circuits of size $M(n^{O(\ell)} + Q2^{O(kq)})$.

The following observation will be useful for us:

▶ **Observation 19.** *Let $\mathcal{F} = (F_1, \ldots, F_M)$ be an ordered collection of k-CNFs. For $\ell \ge k$, the total number of paths Π such that Π is a path of length exactly t in some tree $T \in \mathrm{CCDT}_\ell(\mathcal{F})$ is at most $(2^{\ell+k} \cdot 2^{\ell+k})^{\lceil t/\ell \rceil} \le 16^{t+\ell}$. Consequently, if $(\varrho, y) \in \{0,1\}^{Y_q}$ is such that $\mathcal{C}_{\mathcal{F},\ell,t}(\varrho, y) = 1$, then $\mathcal{C}_P(\varrho, y) = 1$ for (at least one) and at most $16^{t+\ell}$ many ℓ-segmented $\mathcal{F}$-traversals P of length t.*

Proof. This follows by inspection of the recursive construction of the set $\mathrm{CCDT}_\ell(\mathcal{F})$ of canonical common ℓ-partial decision trees for $\mathcal{F}$. 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 18 and Observation 19 in hand, we are now ready to prove our main result of this section (Theorem 14), a derandomized version of the multi-switching lemma (Theorem 11). We restate Theorem 14 here for the reader's convenience:

▶ **Theorem 14.** *Let $\mathcal{F} = (F_1, \ldots, F_M)$ be an ordered list of Q-clause k-CNFs. Let $\delta, p \in (0,1)$ and define $q = \log(1/p)$. Let $\mathcal{D}$ be any distribution over $\{0,1\}^{Y_q}$ that $(\delta/(M^{\lceil t/\ell \rceil} n^{O(t)}))$-fools the class of depth-3 circuits of size $M(n^{O(\ell)} + Q2^{O(kq)})$. Then for all $\ell \geq k$ and all $t \in \mathbb{N}$,*

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

Proof.

$$\begin{aligned}
&\Pr_{(\boldsymbol{\eta},\boldsymbol{z})\leftarrow\mathcal{D}}\left[\mathrm{depth}(\mathrm{CCDT}_\ell(\mathcal{F} \restriction \rho(\boldsymbol{\eta},\boldsymbol{z}))) \geq t\right] \\
&= \mathop{\mathbf{E}}_{(\boldsymbol{\eta},\boldsymbol{z})\leftarrow\mathcal{D}}\left[\mathcal{C}_{\mathcal{F},\ell,t}(\boldsymbol{\eta},\boldsymbol{z})\right] && \text{(Claim 18)} \\
&\leq \sum_{P\in\mathcal{P}_{\mathcal{F},\ell,t}} \mathop{\mathbf{E}}_{(\boldsymbol{\eta},\boldsymbol{z})\leftarrow\mathcal{D}}\left[\mathcal{C}_P(\boldsymbol{\eta},\boldsymbol{z})\right] && \text{(union bound)} \\
&\leq \sum_{P\in\mathcal{P}_{\mathcal{F},\ell,t}} \left(\mathop{\mathbf{E}}_{(\boldsymbol{\varrho},\boldsymbol{y})\leftarrow\mathcal{U}}[\mathcal{C}_P(\boldsymbol{\varrho},\boldsymbol{y})] + \frac{\delta}{M^{\lceil t/\ell \rceil} n^{O(t)}} \right) && (\mathcal{D}\ (\delta/(M^{\lceil t/\ell \rceil} n^{O(t)}))\text{-fools } \mathcal{C}_P) \\
&\leq \delta + \mathop{\mathbf{E}}_{(\boldsymbol{\varrho},\boldsymbol{y})\leftarrow\mathcal{U}}\left[\sum_{P\in\mathcal{P}_{\mathcal{F},\ell,t}} \mathcal{C}_P(\boldsymbol{\varrho},\boldsymbol{y})\right] && \text{(Proposition 16)} \\
&\leq \delta + 16^{t+\ell} \mathop{\mathbf{E}}_{(\boldsymbol{\varrho},\boldsymbol{y})\leftarrow\mathcal{U}}[\mathcal{C}_{\mathcal{F},\ell,t}(\boldsymbol{\varrho},\boldsymbol{y})] && \text{(Observation 19)} \\
&= \delta + 16^{t+\ell} \Pr_{(\boldsymbol{\varrho},\boldsymbol{y})\leftarrow\mathcal{U}}\left[\mathrm{depth}(\mathrm{CCDT}_\ell(\mathcal{F} \restriction \rho(\boldsymbol{\varrho},\boldsymbol{y}))) \geq t\right] && \text{(Claim 18)} \\
&= \delta + 16^{t+\ell} \Pr_{\boldsymbol{\rho}\leftarrow\mathcal{R}_p}\left[\mathrm{depth}(\mathrm{CCDT}_\ell(\mathcal{F} \restriction \boldsymbol{\rho})) \geq t\right] && \text{(Observation 13)} \\
&\leq \delta + 16^{t+\ell} M^{\lceil t/\ell \rceil} (32pk)^t. && \text{(Theorem 11)}
\end{aligned}$$

◀