Electronic Colloquium on Computational Complexity, Report No. 182 (2020)

An Improved Derandomization of the Switching Lemma

Zander Kelley *

November 6, 2020

Abstract

We prove a new derandomization of Håstad's switching lemma, showing how to efficiently generate restrictions satisfying the switching lemma for DNF or CNF formulas of size m using only $\widetilde{O}(\log m)$ random bits. Derandomizations of the switching lemma have been useful in many works as a key building-block for constructing objects which are in some way provably-pseudorandom with respect to AC^0-circuits (e.g., [AW85, TX13, GW14, SS16, AS17, ST17, ST19, BDSG+18, DHH19, Tel19]).

Here, we use our new derandomization to give an improved analysis of the pseudorandom generator of Trevisan and Xue for AC^0-circuits (CCC'13): we show that the generator ε-fools size-m, depth-D circuits with n-bit inputs using only $\widetilde{O}(\log(m/\varepsilon)^D \cdot \log n)$ random bits. In particular, we obtain (modulo the $\log\log$-factors hidden in the $\widetilde{O}$-notation) a dependence on m/ε which is best-possible with respect to currently-known AC^0-circuit lower bounds.

1 Introduction

The switching lemma (originally proved by Håstad [Hås86]) is an important and well-known tool used to analyze low-depth boolean circuits. It says if $F : \{0,1\}^n \to \{0,1\}$ is a DNF (or CNF) formula, with terms (or clauses) of width at most w, then if we randomly fix, or "restrict", all but (roughly) a $\frac{1}{w}$-fraction of inputs to F, then with high probability the resulting function on the remaining $\frac{n}{w}$ bits can be represented as a low-depth decision tree. Since a decision tree of depth d can be expressed either as a width-d DNF or a width-d CNF, this shows that random restrictions be used to "switch" e.g. a bounded-width DNF into a bounded-width CNF. Using this fact, one can argue that iteratively applying D random restrictions to a depth-D AC^0-circuit will likely cause the entire circuit to collapse to a small-depth decision tree.

*Email: awk2@illinois.edu. Department of Computer Science, University of Illinois at Urbana-Champaign. Supported by NSF grants CCF-1755921 and CCF-1814788.

ISSN 1433-8092

The switching lemma lies at the heart of all strong unconditional hardness and pseudorandomness[1] results known for the circuit class AC^0.

- The lemma is used to derive strong correlation bounds for the parity function against AC^0-functions, and plugging this average-case hard function into the Nisan-Wigderson framework yields a pseudorandom generator with seedlength $O(\log^{2D+6} m)$ for depth-D AC^0 circuits of size m [Nis91].
- One can obtain tight bounds on certain Fourier-analytic properties of AC^0-functions using the switching lemma [LMN93, Tal17]. These Fourier-analytic bounds are crucial in the works of Bazzi [Baz09] and Braverman [Bra08] (which were subsequently improved by [Raz09, DETT10, Tal17, HS19]) that show that AC^0-functions are fooled by $\log(m)^{O(D)}$-wise independent distributions.
- The flexible polarizing-walk framework introduced in [CHHL19] can also utilize these Fourier-analytic bounds to give a generator with seedlength $O(\log^{2D} m)$.

The current best pseudorandom generators for AC^0, which obtain seedlength $\log^{D+O(1)} m$ ([TX13, ST19]), stem from the work of Trevisan and Xue, whose pseudorandom construction relies on the switching lemma in the most direct way of all.

Restrictions and Selections. Before we continue, we define some notation for restrictions. A restriction is a vector $\rho \in \{0,1,*\}^n$, which intuitively corresponds to a partial n-bit input, with stars in the locations which are left unspecified. Two restrictions ρ and τ can be composed to form a new restriction $\rho \circ \tau$, defined so that in each coordinate, $(\rho \circ \tau)_i = \tau_i$ if $\rho_i = *$ and $(\rho \circ \tau)_i = \rho_i$ otherwise. Given a boolean function $F : \{0,1\}^n \to \{0,1\}$, the restricted function F_ρ is defined via $F_\rho(x) := F(\rho \circ x)$. Often, we wish to imagine the process of choosing a restriction $\rho \in \{0,1,*\}^n$ as first choosing a set on which to place the stars, and then deciding how to set the bits in the non-star coordinates using some independent process. For this purpose, for $T \in \{0,1\}^n$ and $\rho, \tau \in \{0,1,*\}^n$, we introduce the *selection* notation $T[\rho, \tau]$, which is defined so that

$$T[\rho,\tau]_i := \left\{ \begin{array}{ll} \rho_i, & \text{if } T_i = 0 \\ \tau_i, & \text{if } T_i = 1 \end{array} \right\}.$$

A p-random restriction is defined by $\rho := T[U, \star]$, where $T \in \{0,1\}^n$ is a p-random string (that is, $T_i = 1$ with probability p, independently in every coordinate), and U is a uniformly random vector in $\{0,1\}^n$, and $\star := *^n$ is the vector of all stars.

[1] A *pseudorandom generator* (PRG) is an explicit, efficiently computable mapping $G : \{0,1\}^\ell \to \{0,1\}^n$ that stretches ℓ-bit truly-random seeds into n-bit inputs which are indistinguishable from random to some function class. The corresponding *pseudorandom distribution* D is the uniform distribution over the (multi-) set $G(\{0,1\}^\ell)$. We say that D ε-fools a function F if $|\mathbb{E}[F(D)] - \mathbb{E}[F(U)]| \leq \varepsilon$, where U is the uniform distribution over $\{0,1\}^n$. Sometimes we simply say that D fools F without specifying an error parameter – by this we mean that D ε-fools F for some $\varepsilon \leq 1/3$.

The Trevisan-Xue Construction. In [TX13], Trevisan and Xue employ the generic "iterated pseudorandom restriction" approach to construct their pseudorandom generator for AC^0. This approach, which as first introduced by Ajtai and Wigderson [AW85], begins by constructing a *pseudorandom restriction* τ, which is drawn randomly from some small, efficiently sampleable set of restrictions. Then, one proceeds to compose multiple independent copies of τ, and then the output of the generator is $G = \tau^{(1)} \circ \tau^{(2)} \circ \cdots \circ \tau^{(r)}$, where r is chosen large enough so that it is highly likely that all the bits are set. The advantage of this approach is that, in order to show that $\mathbb{E}[F(G)] \approx \mathbb{E}[F(U)] \pm \varepsilon$ for functions F in some function-class $\mathcal{C}$ that is closed under restriction, it suffices (by a simple hybrid argument) to show, for all $F \in \mathcal{C}$, that $\mathbb{E}[F(\tau \circ U)] \approx \mathbb{E}[F(U)] \pm \varepsilon/r$ for a single pseudorandom restriction τ. This task can be considerably easier than trying to "set all the bits pseudorandomly in one shot". The seedlength of G then is r times the number of random bits needed to generate each independent copy of τ.

For size-m depth-D AC^0-circuits, Trevisan and Xue construct their pseudorandom restriction τ by composing roughly D independent copies of a more basic pseudorandom restriction. Specifically (in terms of the selection notation introduced above), they define

$$\tau := T^1 \wedge T^2 \wedge \cdots \wedge T^{D-1}[\star, X].$$

Here, $X \in \{0,1\}^n$ is chosen to be some basic pseudorandom distribution that fools small-depth decision trees, and each $T^j \in \{0,1\}^n$ is a pseudorandomly-chosen string such that $\mathbb{P}(T_i = 1) \approx 1/\log(m)$ in every coordinate, and the "$\wedge$" operation is a bitwise AND.

After applying a simple hybrid argument, the key to successfully analyzing this restriction is to choose the selection vectors T in such a way that applying a single restriction $\rho := T[U, \star]$ to a DNF or CNF[2] formula F is highly likely to cause the restricted function $F(\rho \circ x)$ to collapse to a low-depth decision tree. That is, if one can show that $\rho = T[U, \star]$, where the star-selection vector T is pseudorandom and the non-star inputs U are uniformly random, satisfies the switching lemma, then it can be argued that an application of ρ to a circuit of depth D will cause it to collapse to depth $D-1$ by switching all of the depth-2 circuits at the input layer. For this purpose, Trevisan and Xue prove the following lemma which is the main technical contribution of their work[3]. Below, the notation $\mathrm{DT}(F)$ stands for the depth of the smallest-depth decision tree that represents the boolean function F, and we use $\mathbb{1}(E)$ to denote the indicator-random-variable of an event E.

[2] Since a CNF formula is functionally equivalent to the negation of a DNF, and since the negation of a decision tree of depth d is also a decision tree of depth d, we can without loss of generality restrict our attention to DNFs in statements and proofs of the switching lemma, and the corresponding corollaries for CNFs follows easily.

[3] Actually, Trevisan and Xue prove a more general statement that allows the construction of restrictions $\rho = T[Y, \star]$ where both T and Y are pseudorandom, rather than just the selection vector T. This extension is important for some applications of their derandomized switching lemma given by later works, but (due to the use of a hybrid argument which we have discussed above) this extension is not needed in their original setting of derandomizing AC^0 via the iterated-restriction construction.

Lemma 1.1 ([TX13], Implicit in Lemma 7 and its proof)**.** *Fix a* DNF $F(x) = \bigvee_{i=1}^{m} A_i(x)$ *of width* w*, and let* $\rho = T[y, \star]$ *with* $T, y \in \{0,1\}^n$*. Then there is a function* $B^y(T)$*, depending on* F *and* y*, such that*

$$\mathbb{1}(\mathrm{DT}(F_\rho) \geq d) \leq B^y(T) := \sum_{i=1}^{K} f_i^y(T),$$

where

- $K \leq (4^w m)^d$*,*
- *each* f_i^y *is a* CNF *with at most* $2^w m$ *clauses,*

and furthermore, if y *is chosen uniformly randomly from* $\{0,1\}^n$ *and* T *is a truly* p*-random string, then*

- $\mathbb{E}_y \mathbb{E}_T B_y(T) \leq 2^{d+w}(5pw)^d$*.*

By plugging in state-of-the-art pseudorandom generators for CNFs (see [Tal17]), it is therefore possible to generate pseudorandom selection vectors T satisfying the following derandomized switching lemma.

Corollary 1.2. *For any* $p \in [2^{-n}, 1]$ *that is a power of a half, there is an efficiently-computable pseudorandom distribution over vectors* $T \in \{0,1\}^n$*, which can be sampled using only*

$$O\left(\log n + (d \log(m) + \log(1/\varepsilon)) \cdot \log(m) \cdot \log\log m\right)$$

random bits, with the following property. If $\rho := T[U, \star]$ *is a random restriction defined by pseudorandom selection* T *and uniformly random assignment* $U \in \{0,1\}^n$*, and* F *is any* DNF *with* m *terms and width* $w \leq O(\log m)$*, then*

$$\mathbb{P}(\mathrm{DT}(F_\rho) \geq d) \leq m^{O(1)} \cdot (10pw)^d + \varepsilon.$$

Furthermore, the probability that $T_i = 1$ *is at least* $p - \varepsilon$ *in every coordinate.*

Our Contribution. Since d must be chosen to be at least $\Omega(\log m)$ for the this to be useful (in the standard setting where $p = \Theta(1/w)$), we can summarize the above as achieving a pseudorandom restriction with seedlength $\widetilde{O}(\log^3 m)$ that satisfies the switching lemma on DNFs of size m. The main result of this paper is an improved derandomization of the switching lemma – we show how to generate restrictions satisfying the switching lemma using only $\widetilde{O}(\log m)$ random bits.

Theorem 1.3. *For any $p \in [2^{-n}, 1]$ which is a power of a half, there is an efficiently computable pseudorandom distribution over vectors $T \in \{0,1\}^n$, which can be sampled using only*

$$O(\log n + (w+d) \cdot \log w + \log(1/\varepsilon))$$

random bits, with the following property. If $\rho := T[U, \star]$ is a random restriction defined by pseudorandom selection T and uniformly random assignment $U \in \{0,1\}^n$, and F is any DNF *of width at most w, then*

$$\mathbb{P}(\mathrm{DT}(F_\rho) \geq d) \leq O(pw)^d + \varepsilon.$$

Furthermore, the probability that $T_i = 1$ is at least $p - \varepsilon$ in every coordinate.

By combining this with the well-known (and easily derandomizable) observation that randomly restricting a constant-fraction of inputs to a size-m DNF will likely cause it collapse to a DNF of width at most $O(\log m)$, we recover the corresponding statement for size-m DNFs with unbounded width.

Corollary 1.4. *There is an efficiently computable pseudorandom distribution over vectors $T \in \{0,1\}^n$, with seedlength*

$$O(\log n + (d + \log(m/\varepsilon)) \cdot \log\log(m/\varepsilon)),$$

such that for any DNF *F with m terms, the restriction $\rho := T[U, \star]$ satisfies*

$$\mathbb{P}(\mathrm{DT}(F_\rho) \geq d) \leq O(p \log(m/\varepsilon))^d + \varepsilon.$$

Furthermore, the probability that $T_i = 1$ is at least $p - \varepsilon$ in every coordinate.

In fact, we show that in order to satisfy the switching lemma, the pseudorandom selection vector T must merely possess the following weak pseudorandomness property we call p-boundedness; this simple, "one-sided" property is much coarser than the requirement that T fool CNFs, or even certain more basic pseudorandom properties such as k-wise independence or δ-bias (see Section 2), which still require fine, "two-sided" control on the behavior on small sets of coordinates.

Definition 1.5. *Say that a distribution over vectors $T \in \{0,1\}^n$ is k-wise p-bounded if, for every set $S \subseteq [n]$ of size $s \leq k$, we have*

$$\mathop{\mathbb{E}}_{T}\left[\prod_{i \in S} T_i\right] \leq p^s.$$

Theorem 1.6. *Suppose that T is a $(w+d)$-wise p-bounded distribution over $\{0,1\}^n$, and U is uniform over $\{0,1\}^n$. If $p \leq \frac{1}{16w}$, then for any* DNF *F of width at most w, the restriction $\rho := T[U, \star]$ satisfies*

$$\mathbb{P}(\mathrm{DT}(F_\rho) \geq d) \leq 2 \cdot (8pw)^d.$$

1.1 Proof Technique

When we imagine the task of derandomization with respect to a particular application, the setting is typically as follows. We have some bad event B, depending on some random choices $x \in \{0,1\}^n$, and we must show (say, in order to show that some randomized algorithm is likely to succeed) that the probability of B occurring is small. Suppose we have a proof which does indeed establish such a bound. Now, identify the event B with its own indicator function $B : \{0,1\}^n \to \{0,1\}$. To derandomize this statement, we can try to peer into our proof and see how $B(x)$ depends on the choices x. If the dependence is simple enough – for instance maybe the proof is just a union-bound over some local events involving at most k variables at a time – we are in luck and we can instead draw x from some merely k-wise independent distribution and inherit the same probability-of-success guarantee. However, if the proof is not simple enough, then it seems that we would need to look for some other, simpler way to bound the probability of B, which is unfortunate since we potentially miss out on the power of more sophisticated proof techniques.

The crucial observation of Trevisan and Xue is that this need not be the case. Indeed, if one can show that the event B can be expressed as a simple function (e.g. a sum of CNFs as in Lemma 1.1) that can fooled by some pseudorandom distribution X, then we can bound $\mathbb{E}[B(X)]$ in two distinct steps: first, show that $\mathbb{E}[B(X)] \approx \mathbb{E}[B(U)]$ by some "simple" argument, and only then show that $\mathbb{E}[B(U)]$ is small via some separate "complicated" argument. This is especially important for derandomizing the switching lemma because both of the well-known proofs of the switching lemma (i.e. Håstad's original conditioning-based proof [Hås86] as well as Razborov's alternative encoding-based proof[4] [Raz95]) seem hopelessly sophisticated and extremely fragile from a direct-derandomization point-of-view.

Of course, the drawback of this abstract approach is that we can not hope to obtain from it pseudorandom restrictions with seedlength any better than our best PRGs for CNFs. As we discuss further in Section 1.2, obtaining a PRG with seedlength $\widetilde{O}(\log m/\varepsilon)$ for size-m CNFs (which is what would be required to obtain an "ideal" derandomization of the switching lemma) would require a major breakthrough in circuit-complexity. Here, we sidestep this barrier by analyzing (a suitable modification of) a recent new proof of the switching lemma (which we would describe as "coupling-based") due to Rossman [Ros19]. We show by a careful analysis that it is (in our opinion, just barely) amenable to direct-derandomization.

Originally, the purpose of Rossman's alternative approach was to prove the switching lemma directly for size-m DNFs with unbounded width. This is in contrast to the more standard two-step argument, where one first shows that randomly restricting a constant-fraction of the inputs causes the DNF to collapse to width $w \leq O(\log m)$, and then argues (via the proof of Håstad or Razborov) that further restricting this width-w DNF with a $\Theta(1/w)$-random restriction will cause it to collapse to a small-depth decision tree. Rossman's argument gives a better bound (for a certain range of parameters) than this two-step argument. Rossman describes his own proof as "entropy-based", because the calculations which are required in

[4] See also the expositions by [Juk12, Tha09].

order to handle DNFs of unbounded width resemble the calculations one would make to prove the bound $\sum_{i=1}^{m} \pi_i \cdot \log(1/\pi_i) \leq \log(m)$ for arbitrary probability distributions $\pi \in \mathbb{R}^m$.

Here, we use the approach of Rossman for a completely different purpose, in a completely different setting. Specifically, we apply the approach in the setting of width-w DNFs that have unbounded size. We do not use any of the calculations which Rossman describes as "entropy-based", and so we describe the core of the remaining argument, a key re-randomization and coupling step, as "coupling-based". It is our understanding that, prior to our work, it was not known that Rossman's proof offered any advantage over the earlier proofs of Håstad or Razborov in the setting of bounded-width DNFs. An important message of our work is that the coupling-based approach indeed has a substantial advantage in the context of derandomization (and, as we discuss further in Section 1.3, we believe this advantage could be relevant to applications beyond the switching lemma).

The coupling-based approach leads to a proof of the switching lemma that is in many ways more "explicit" in how the bad event depends on the restriction ρ than earlier proofs. Unfortunately, this explicitness comes at the price of some fairly elaborate notation. So, for the benefit of the reader, we include a section explaining how the coupling-based approach can be used to prove (and derandomize) the fact that a p-random restriction applied to a width-w DNF will cause the DNF to become identically equal to a constant, except with probability $O(pw)$ – this fact is sometimes referred to as the "baby" switching lemma. This section (Section 3) can be freely skipped as it is not critical to any of our results. However we advise against this, as understanding the derandomization in this simpler setting is enough to grasp the key aspects of the technique; in particular, is sufficient to understand why direct-derandomization can succeed here while it has failed before.

1.2 Applications

Various works that construct objects which are in some way pseudorandom with respect to AC^0-circuits often rely on some kind of derandomization of the switching lemma. Examples of such constructions include the pseudorandom generators of [AW85, TX13, ST19, AS17, DHH19], the quantified derandomizations of [GW14, Tel19], the stochastic list-decodable codes of [SS16], and the non-malleable codes of [BDSG+18].

However, the type of guarantee given by the derandomized switching lemma proved in this work does not universally suffice for all of these applications; in particular, some applications require restrictions $\rho = T[Y, \star]$, where both T and Y are generated pseudorandomly, while we construct pseudorandom-selection distributions T such that the restriction $\rho = T[U, \star]$ satisfies the switching lemma when U is uniformly random. We discuss two applications where this type of derandomization is sufficient, and explain how our improved derandomization leads to more efficient solutions than were previously known.

Pseudorandom Generators for AC^0-Circuits. In their paper, Trevisan and Xue showed that the construction outlined in Section 1 gives a PRG that ε-fools size-m, depth-D AC^0-circuits and has seedlength $\widetilde{O}(\log(m/\varepsilon)^{D+3} \cdot \log(n/\varepsilon))$. In [Tal17], Tal gives an improved analysis, showing that (a minor alteration of) the Trevisan-Xue construction achieves seedlength $\widetilde{O}(\log(m/\varepsilon)^{D+1} \cdot \log n)$. Plugging our improved derandomization of the switching lemma into the construction yields the following.

Theorem 1.7. *There is an explicit pseudorandom generator that ε-fools size-m, depth-D AC^0-circuits[5], and has seedlength $\widetilde{O}(\log(m/\varepsilon)^D \cdot \log n)$. More specifically, the seedlength is*

$$O(\log(m/\varepsilon))^D \cdot \log(n) \cdot (\log\log(m/\varepsilon))^3.$$

Obtaining this specific seedlength is somewhat of a landmark, as it can be shown (by an easy argument sometimes referred to as the "discriminator lemma") that achieving a seedlength of, say,

$$O(\log(m/\varepsilon))^{D-0.01} \cdot \log(n)^{O(1)},$$

would imply a stronger worst-case lower bound against depth-$(D+1)$ circuits than is currently known for any explicit hard function[6]. Thus, modulo the log log-factors hidden in the $\widetilde{O}$-notation, the seedlength we obtain is best-possible without improving upon AC^0-circuit lower-bounds which have remained best-known for over 30 years.

Deterministic Search for CNF Satisfying Assignments. Suppose you have a CNF formula[7] for which it is known that at least a fraction $\varepsilon = 0.01$ of all possible inputs are satisfying, and you are tasked with finding some specific satisfying assignment. It is easy to give a randomized solution: just try random strings $x \in \{0,1\}^n$ until you find a satisfying assignment. However, it is nontrivial to give a deterministic solution to this problem.

Perhaps the most natural approach is to use a pseudorandom generator that (say) $\varepsilon/2$-fools polynomially-sized CNFs – then, one of the possible outputs of the generator is guaranteed to be satisfying. Since the best-known generators for poly(n)-sized CNFs have seedlength $\widetilde{O}(\log^2(n))$, this approach yields a deterministic search algorithm running in time $n^{\widetilde{O}(\log n)}$.

In [ST17], Servedio and Tan improve upon this by combining together two ingredients into a clever "decision-to-search reduction"-type solution to this problem. The first ingredient is a deterministic approximate-counting algorithm due to [GMR13] that, given a poly(n)-sized CNF, reports the fraction of satisfying assignments to the CNF (up to an approximation error $\pm\gamma$), and runs in time

$$\left(\frac{n}{\gamma}\right)^{\widetilde{O}(\log\log n + \log(1/\gamma))}.$$

[5] Here we assume we are in the standard setting where $m \geq n$, where the circuit is large enough to at least read all of the input bits.

[6] See [TX13] for further discussion of this barrier.

[7] For simplicity, we will in this section restrict our attention to CNFs of size at most $n^{O(1)}$.

The second ingredient, which is also due to [GMR13], is a particular derandomization of the switching lemma that uses $O(\log(n) \cdot \log\log n)$ random bits to restrict roughly a p-fraction of the inputs to a poly(n)-sized CNF in a way that, on average, approximately preserves the fraction of satisfying assignments, where

$$p \approx \frac{1}{\log(n)^{\log\log n}}.$$

We observe that by using our new derandomization of the switching lemma, we can do the same, but with the improved parameter $p \approx 1/\log(n)$.

Theorem 1.8. *There is an efficiently computable pseudorandom distribution over vectors $T, X \in \{0,1\}^n$, with total seedlength*

$$O(\log n + \log(m/\gamma) \cdot \log\log(m/\gamma)),$$

such that for any DNF *or* CNF *F of size m, the restriction $\rho := T[\star, X]$ satisfies*

$$\left| \mathop{\mathbb{E}}_{\rho} \mathop{\mathbb{E}}_{U} F(\rho \circ U) - \mathop{\mathbb{E}}_{U} F(U) \right| \leq \gamma,$$

where U is uniformly distributed over $\{0,1\}^n$. Furthermore, the probability that $T_i = 1$ is at least $\Omega(1/\log(m))$ in every coordinate $i \in [n]$.

Servedio and Tan proceed to iterate the following until all input-bits are fixed (and thus a satisfying assignment has been found):

Given a CNF $F : \{0,1\}^n \to \{0,1\}$ of size $m \leq n^{O(1)}$:

- Generate all $n^{O(\log\log n)}$ restrictions $\rho \in \{0,1,*\}^n$ from the pseudorandom distribution described by [GMR13].
- For each ρ, use the approximate-counting algorithm of [GMR13] to estimate the number of satisfying assignments to $F_\rho(x)$.
- Pick the ρ which resulted in the largest estimate, set all input bits that are not yet set according the the restriction ρ, and continue on the restricted CNF F_ρ.

In order to fix all the input-bits, this process must be iterated $r \approx \log(n)/p$ times, and since the approximation-error accumulates from every iteration, the approximate-counting algorithm must be run with parameter $\gamma := \varepsilon/r \approx \varepsilon p/\log(n)$. Thus, using the pseudorandom restriction distribution from [GMR13], Servedio and Tan set $\gamma \approx \varepsilon \cdot 2^{-(\log\log n)^2}$ and obtain a deterministic search algorithm that runs in time

$$\left(\frac{n}{\varepsilon}\right)^{\widetilde{O}(\log\log n + \log(1/\varepsilon))^2}.$$

If we instead plug in the pseudorandom restriction distribution given by Theorem 1.8, we can afford to set $\gamma \approx \varepsilon/\log(n)^2$, and we obtain an improved deterministic search algorithm running in time

$$\left(\frac{n}{\varepsilon}\right)^{\widetilde{O}(\log\log n+\log(1/\varepsilon))},$$

thus bringing the time required to solve this task in line with the time required by the best-known algorithms for approximate-counting.

Theorem 1.9. *There is a deterministic algorithm that, given any* CNF $F : \{0,1\}^n \to \{0,1\}$ *of size* $m \leq n^{O(1)}$ *for which at least an* ε*-fraction of all inputs* $x \in \{0,1\}^n$ *are satisfying, finds such an input in time*

$$\left(\frac{n}{\varepsilon}\right)^{\widetilde{O}(\log\log n+\log(1/\varepsilon))}.$$

1.3 Open Problems

For future work, we ask whether the approach in this paper can be used to give high-quality direct-derandomizations in other cases where this previously seemed impossible. In particular, we highlight the multi-switching lemma of [Hås14], the robust-sunflower lemma of [ALWZ20], and the work on DNF compression due to [LWZ20] as potential candidates.

The multi-switching lemma. The multi-switching lemma, which is also due to Håstad [Hås14], is a more refined statement concerning the "common-decision-tree complexity" of a sequence of DNFs that are all hit by the same random restriction. In typical style, he originally gave a Håstad-type conditioning-based proof of this result. Alternative Razborov-type encoding-based proofs were given by [Tal17] and [ST19] .

In [ST19], Servedio and Tan prove a Trevisan-Xue-style derandomization of the multi-switching lemma, showing how to generate restrictions satisfying the lemma with $\log(m)^{O(1)}$ random bits. They use this derandomization to give a pseudorandom generator for AC^0 with seedlength

$$\log(m)^{D+O(1)}\log(1/\varepsilon),$$

which is incomparable to the seedlength obtained in this work due to its superior dependence on ε. We leave it as an open question whether it is possible to use the approach of this work to obtain a better derandomization of the multi-switching lemma, and whether such a derandomization can lead to a pseudorandom generator for AC^0 with the best qualities of both works.

Robust-sunflowers, DNF compression, and the power of p-boundedness. The celebrated robust-sunflower lemma due to Alweiss, Lovett, Wu, and Zhang [ALWZ20] is the statement that DNFs with a certain structural property known as "spreadness" are highly likely to be satisfied by a random input. Besides its important combinatorial applications, the robust-sunflower lemma has recently been applied to obtain improved lifting theorems in

communication complexity [LMZ20, MP20]. Lovett has suggested [personal communication] that in order to push these lifting applications further, what is needed is an appropriate derandomization of the robust-sunflower lemma. More specifically, what is desired is a proof that the robust-sunflower lemma is true even for input-distributions which merely possess some natural, "one-sided" weak pseudorandomness property similar in spirit to e.g. spreadness or p-boundedness.

In [ALWZ20], the core of the proof of the robust-sunflower lemma is a key width-reduction step which is proved using a Razborov-type encoding argument. We propose that a sensible approach to obtaining an appropriately derandomized robust-sunflower lemma is the following:

1. Give a Rossman-type coupling-based proof for this width-reduction step.
2. Derandomize this proof using the approach of this work.

However, we suggest to first start by derandomizing the following (simpler) related statement, which is a key lemma due to Lovett, Wu, and Zhang in their work on decision-list compression (for simplicity we state it here only for DNFs). This lemma is also proved by a Razborov-style encoding argument.

Lemma 1.10. *Let $F(x) = \bigvee_{i=1}^{m} A_i(x)$ be a* DNF *of width w, and let $\rho \in \{0,1,*\}^n$ be a p-random restriction. Say that a term $A_i(x)$ is "useful" in $F(x)$ if there is any input x such that $A_i(x) = 1$ and $A_j(x) = 0$ for all $j < i$. We have the bound*

$$\mathop{\mathbb{E}}_{\rho} \sum_{i=1}^{m} \mathbb{1}(A_i(\rho \circ x) \text{ is useful in } F(\rho \circ x)) \leq \left(\frac{4}{1-p}\right)^w.$$

We observe that we can carry out the first step of our suggested plan; namely, in Section 5 we include an alternative Rossman-style coupling-based proof of this lemma (in fact, with the improved constant 2 instead of 4). However, unlike the situation with the switching lemma, we do not see how to derandomize this proof; although we believe it should be possible, it will require new ideas. Concretely, we ask for a proof or refutation of the following conjecture.

Conjecture 1.11. *Lemma 1.10 is true even for $\rho = T[U, \star]$, where U is distributed uniformly over $\{0,1\}^n$, and T is any w-wise p-bounded distribution over $\{0,1\}^n$.*

2 Preliminaries

Decision Lists. A decision list is, for the purpose of this paper, a mathematical operator that takes two boolean vectors $a, b \in \{0,1\}^m$ and produces a single boolean output defined by the following process: find the smallest index $i \in [m]$ such that $a_i = 1$, and output b_i. If there is no such index i, the decision list returns a default value of 0. Thus, the value of a decision list on (a, b) is given by the summation

$$\sum_i a_i \cdot \phi_i \cdot b_i,$$

where

$$\phi_i := \prod_{j<i} (1 - a_i).$$

In order to clean up some expressions within this paper, we introduce the notational shorthand

$$\mathop{\mathbb{L}}_i (a_i \to b_i) := \sum_i a_i \cdot \phi_i \cdot b_i.$$

Restrictions. We use the notation for restrictions introduced near the beginning of Section 1.

Small-Bias Distributions. We make use of δ-biased distributions, which are a basic pseudorandomness primitive with efficient constructions due to [NN93] and [AGHP92]. A distribution X over $\{0,1\}^n$ is said to be δ-biased if, for every nonzero $\alpha \in \{0,1\}^n$, X δ-fools the parity function specified by the bits in α. That is,

$$|\mathop{\mathbb{E}}_X (-1)^{\langle \alpha, X \rangle}| \leq \delta.$$

Standard constructions of δ-biased distributions have seedlength $O(\log n + \log 1/\delta)$. We make use of the following simple properties of δ-biased distributions X.

Proposition 2.1. *If $f : \{0,1\}^n \to \{0,1\}$ is an AND of any k (logically-consistent) literals, then*

$$|\mathop{\mathbb{E}}_X f(X) - \left(\tfrac{1}{2}\right)^k| \leq \delta.$$

Proof. The function $f(x)$ can be expressed in the form

$$f(x) = \left(\tfrac{1}{2} \pm \tfrac{1}{2}(-1)^{x_{i_1}}\right)\left(\tfrac{1}{2} \pm \tfrac{1}{2}(-1)^{x_{i_2}}\right) \cdots \left(\tfrac{1}{2} \pm \tfrac{1}{2}(-1)^{x_{i_k}}\right).$$

Expanding this product gives a convex-combination of parity functions. □

Proposition 2.2. *If $f : \{0,1\}^n \to \{0,1\}$ depends on at most k variables, then*[8]

$$|\mathop{\mathbb{E}}_X f(X) - \mathop{\mathbb{E}}_U f(U)| \leq \delta \cdot 2^k.$$

Proof. By considering its truth table, the function $f(x)$ can be expressed as a sum of at most 2^k ANDs of literals. □

[8] This error bound can be improved to $\delta \cdot 2^{k/2}$ using a Fourier-analytic argument, but the simple argument given here suffices for most applications.

3 Proof of the Baby Switching Lemma

We begin this section by fixing a DNF $F(x) = \bigvee_{i=1}^{m} A_i(x)$, where each term A_i is an AND of at most w literals. For $i \in [m]$, we let $V_i \subseteq [n]$ be the set of variables contained in term A_i. Also for each i, fix some vector $\sigma_i \in \{0,1\}^n$ which satisfies $A_i(\sigma_i) = 1$, and define the boolean functions

$$\phi_i(x) := \prod_{j<i} (1 - A_j(x)).$$

In this section we prove the following pseudorandom "baby" switching lemma.

Theorem 3.1. *Let $\rho := T[y, \star]$, where T is a w-wise p-bounded distribution over $\{0,1\}^n$ and y is uniformly random in $\{0,1\}^n$. If F is a DNF of width w, then*

$$\mathbb{P}(F_\rho \text{ is non-constant}) \leq 4pw.$$

Proof. If $F_\rho(x) := F(\rho \circ x)$ is not identically a constant, then there is some term A_i such that

- A_i is the first term that is not falsified by ρ, and
- A_i is not satisfied by ρ.

Equivalently, we might say (more explicitly) that there is some index i and some integer $s \in [1, w]$ such that

- $A_i(\rho \circ \sigma_i) = 1$,
- $\phi_i(\rho \circ x) \equiv 1$ as a function of x, and
- $|\text{Stars}(\rho) \cap V_i| = s$.

This proves the inequality of random variables variables

$$\mathbb{1}(F_\rho \text{ is non-constant}) \leq \sum_{s=1}^{w} \sum_{i=1}^{m} A_i(\rho \circ \sigma_i) \cdot \mathbb{1}(\phi_i(\rho) \equiv 1) \cdot \mathbb{1}(|\text{Stars}(\rho) \cap V_i| = s)\,.$$

Now, let's fix an index i and a restriction ρ such that $|\text{Stars}(\rho) \cap V_i| = s \geq 1$. At this point we want to consider what happens to the value $A_i(\rho \circ \sigma_i)$ when we replace σ_i by a uniformly random input $x \in \{0,1\}^n$. If ρ falsifies A_i, then the value remains unchanged. If instead ρ is consistent with A_i, then $A_i(\rho \circ x)$ is functionally equivalent to an AND of s literals of x, so

$$\mathop{\mathbb{E}}_{x} A_i(\rho \circ x) = \left(\tfrac{1}{2}\right)^s.$$

Thus, in any case, we have the inequality $A_i(\rho \circ \sigma_i) \le 2^s \, \mathbb{E}_x \, A_i(\rho \circ x)$. We conclude that

$$\begin{aligned}
\mathbb{1}(F_\rho \text{ is non-constant}) &\le \sum_{s=1}^{w} \sum_{i=1}^{m} A_i(\rho \circ \sigma_i) \cdot \mathbb{1}(\phi_i(\rho) \equiv 1) \cdot \mathbb{1}(|\mathrm{Stars}(\rho) \cap V_i| = s) \\
&\le \sum_{s=1}^{w} 2^s \mathop{\mathbb{E}}_{x} \sum_{i} A_i(\rho \circ x) \cdot \mathbb{1}(\phi_i(\rho) \equiv 1) \cdot \mathbb{1}(|\mathrm{Stars}(\rho) \cap V_i| = s) \\
&\le \sum_{s=1}^{w} 2^s \mathop{\mathbb{E}}_{x} \sum_{i} A_i(\rho \circ x) \cdot \phi_i(x) \cdot \mathbb{1}(|\mathrm{Stars}(\rho) \cap V_i| = s) \\
&= \sum_{s=1}^{w} 2^s \mathop{\mathbb{E}}_{x} \mathop{\mathbb{L}}_{i} A_i(\rho \circ x) \to \mathbb{1}(|\mathrm{Stars}(\rho) \cap V_i| = s) \\
&= \sum_{s=1}^{w} 2^s \mathop{\mathbb{E}}_{x} \mathop{\mathbb{L}}_{i} A_i(T[y, x]) \to \mathbb{1}(|T_{V_i}| = s) \,,
\end{aligned}$$

where x is a uniformly random vector in $\{0,1\}^n$ which we introduce purely for the sake of analysis. Averaging over $\rho = T[y, \star]$ gives

$$\mathbb{P}(F_\rho \text{ is non-constant}) \le \sum_{s=1}^{w} 2^s \mathop{\mathbb{E}}_{T} \mathop{\mathbb{E}}_{y} \mathop{\mathbb{E}}_{x} \mathop{\mathbb{L}}_{i} A_i(T[y, x]) \to \mathbb{1}(|T_{V_i}| = s) \,.$$

Here, we make the key observation that, for any fixed vector T, the distribution $T[y, x]$ is simply the uniform distribution over $\{0,1\}^n$, and in particular it does not depend on T. So, the above expression if equivalent to

$$\begin{aligned}
\sum_{s=1}^{w} 2^s \mathop{\mathbb{E}}_{T} \mathop{\mathbb{E}}_{U} \mathop{\mathbb{L}}_{i} A_i(U) \to \mathbb{1}(|T_{V_i}| = s) &= \sum_{s=1}^{w} 2^s \mathop{\mathbb{E}}_{T} \sum_{i} \pi_i \cdot \mathbb{1}(|T_{V_i}| = s) \\
&= \sum_{i} \pi_i \cdot \sum_{r=1}^{w} 2^s \cdot \mathbb{P}(|T_{V_i}| = s) \,,
\end{aligned}$$

where π_i is the probability that, upon uniformly random input $U \in \{0,1\}^n$, A_i is the first term with $A_i(U) = 1$. Note of course that $\sum_i \pi_i \le 1$.

To conclude the calculation, we use the p-boundedness assumption on T to say that

$$\mathbb{P}(|T_{V_i}| = s) \le \mathop{\mathbb{E}}_{T} \sum_{S \in \binom{V_i}{s}} \prod_{j \in S} T_j \le \binom{w}{s} \cdot p^s.$$

Finally, summing over all s gives

$$\mathbb{P}(F_\rho \text{ is non-constant}) \le \sum_{s=1}^{w} \binom{w}{s} (2p)^s = (1 + 2p)^w - 1 \le e^{2pw} - 1.$$

We assume that $p \le \frac{1}{2w}$, since otherwise the desired bound is trivial, and so we finish by applying the estimate $e^t \le 1 + 2t$, which is valid for $t \in [0, 1]$. $\square$

4 Proof of the Full Switching Lemma

We begin this section by fixing a DNF $F(x) = \bigvee_{i=1}^{m} A_i(x)$, where each term A_i is an AND of at most w literals. For $i \in [m]$, we let $V_i \subseteq [n]$ be the set of variables contained in term A_i. We will insist that these sets are presented in increasing order so that we can refer to the "j-th entry of V_i", which we denote $V_i[j]$.

We recall the notion of the canonical decision tree for a DNF. We use the notation $\mathrm{CDT}(F, \rho)$ to refer to the canonical decision tree of the restricted function $F(\rho \circ x)$. The canonical decision tree is defined by the simple, greedy, recursive construction described below. For $Q \subseteq [n]$ and $\alpha \in \{0,1\}^{|Q|}$, we let $Q \leftarrow \alpha$ denote the restriction which sets the variables in Q according to α in the natural way, and has stars elsewhere.

$\mathrm{CDT}(F, \rho)$:

- If F is empty, return 0.
- If A_1 is satisfied by ρ, return 1.
- If A_1 is falsified by ρ, return $\mathrm{CDT}(\bigvee_{i=2}^{m} A_i(x), \rho)$
- Otherwise, let $Q = \mathrm{Stars}(\rho) \cap V_1$ be the set of free variables in A_1, and query all of them. That is, we construct a complete binary tree of depth $|Q|$, and to each path $\alpha \in \{0,1\}^{|Q|}$, we assign the value $\mathrm{CDT}(\bigvee_{i=2}^{m} A_i, \rho \circ (Q \leftarrow \alpha))$.

We remark that in the context of the final bullet point above, the restrictions $\rho \circ (Q \leftarrow \alpha)$ and $(Q \leftarrow \alpha) \circ \rho$ are in fact the same since $Q \subseteq \mathrm{Stars}(\rho)$. So we can equivalently say that we recurse on "$\mathrm{CDT}(\bigvee_{i=2}^{m} A_i, (Q \leftarrow \alpha) \circ \rho)$" – it will be preferable for us to instead imagine composing the restrictions in this way.

We wish to unpack this recursive definition of the canonical decision tree so that we can express the event that $\mathrm{CDT}(F, \rho)$ has depth $\geq d$ in terms of some more explicit conditions depending on ρ. Unfortunately, this will require us to introduce quite a bit of additional notation; to get started, for a set $Q \in \binom{[w]}{t}$, a vector $\alpha \in \{0,1\}^t$, and an index $\ell \in [m]$, we define the restriction

$$Q \leftarrow_\ell \alpha$$

so that for all $j \in Q, j \leq |V_\ell|$,

$$(Q \leftarrow_\ell \alpha)_{V_\ell[j]} := \alpha_j,$$

and elsewhere we have $(Q \leftarrow_\ell \alpha)_i := *$. Thus, $Q \leftarrow_\ell \alpha$ corresponds to the restriction which fixes a subset of the variables in V_ℓ, where the subset is specified by $Q \subseteq [w]$, according to α.

Given some sets $Q_i \in \binom{[w]}{s_i}$, vectors $\alpha_i \in \{0,1\}^{s_i}$, and indices $\ell_i \in [m]$, we denote the corresponding restrictions by

$$\widetilde{\alpha}_i := Q_i \leftarrow_{\ell_i} \alpha_i.$$

Lastly, given some sets $Q_1, Q_2, \ldots, Q_r \subseteq [w]$ and a tuple of indices $\ell = (\ell_1, \ell_2, \ldots, \ell_r)$, define

$$\widetilde{Q}(\ell) := \{V_i[j] \ : \ i \in [r],\ j \in Q_i,\ j \leq |V_i|\},$$

that is, $\widetilde{Q}(\ell) \subseteq \bigcup_{i=1}^{r} V_i$ is a set of variables which is selected based on the subsets $Q_i \subseteq [w]$.

Lemma 4.1. *Suppose* $\mathrm{CDT}(F, \rho)$ *has a path* $\alpha \in \{0,1\}^{d-1}$ *that fails to reach a leaf of the decision tree. Then there exist*

- *integers* $r \in [d]$, $s \in [d, d+w-1]$, *and* $s_1, \ldots s_r \geq 1$ *with* $s_1 + \cdots + s_r = s$,
- *indices* $1 \leq \ell_1 < \ldots < \ell_r \leq m$,
- *sets* $Q_i \in \binom{[w]}{s_i}$ *for all* $i \in [r]$, *and vectors* $\alpha_i \in \{0,1\}^{s_i}$ *for all* $i \in [r-1]$,

such that

1. *For each* $i \in [r]$, A_{ℓ_i} *is the first term in* F *that is not falsified by* $\widetilde{\alpha}_1 \circ \cdots \circ \widetilde{\alpha}_{i-1} \circ \rho$,
2. $\mathrm{Stars}(\rho) \cap (\bigcup_{i=1}^{r} V_{\ell_i}) = \widetilde{Q}(\ell)$.

Proof. We simply unpack the recursive definition of $\mathrm{CDT}(F, \rho)$, following along the path in the decision tree defined by the bits in α. For each i,

- We record the index ℓ_i of the first term in F that is not falsified by $\widetilde{\alpha}_1 \circ \cdots \circ \widetilde{\alpha}_{i-1} \circ \rho$.
- We record the variables queried while processing A_{ℓ_i} (encoded as $Q_i \subseteq [w]$), and set $s_i := |Q_i|$,
- If $\sum_{j \leq i} s_i \geq d$, we set $r := i$ and $s := \sum_{j \leq r} s_i$ and terminate.
- Otherwise, we use the next s_i bits of α to determine α_i and continue.

In the first item above, it must be the case that such a term exists, and is not satisfied by $\widetilde{\alpha}_1 \circ \cdots \circ \widetilde{\alpha}_{i-1} \circ \rho$, or else we have reached the end of the path in the decision tree defined by α before reaching depth d. Thus, $s_i \geq 1$ for all i. Since we terminate as soon as possible after reaching depth d, and the terms all have width at most w, we are guaranteed $r \leq d$ and $s \leq d + w - 1$. □

Lemma 4.1 suggests the following approach for bounding the probability that F, randomly restricted by ρ, has a canonical decision tree of depth at least d: first, fix some data $r, s, \ell_i, Q_i, \alpha_i$; then, bound the probability that items (*1.*) and (*2.*) occur for a particular fixing of the data, and finally sum over all possibilities for that data to get an overall bound.

Towards this end, given data $\ell = (\ell_1, \ldots, \ell_r)$, $\alpha = (\alpha_1, \ldots, \alpha_{r-1})$, and $Q = (Q_1, \ldots, Q_r)$, define the functions

$$f_\ell^{Q,\alpha}(x) := \prod_{i=1}^{r} A_i(\widetilde{\alpha}_1 \circ \cdots \circ \widetilde{\alpha}_{i-1} \circ x).$$

We now come to a key trick needed for our proof: we can apply the function $f_\ell^{Q,\alpha}$ to a random completion of ρ in order to detect whether it satisfies the conditions described in (*1.*) of Lemma 4.1. Indeed, fix a restriction ρ and suppose that $|\mathrm{Stars}(\rho) \cap (\bigcup_{i=1}^{r} V_{\ell_i})| = s$. Then, as a function of x, $f_\ell^{\alpha,Q}(\rho \circ x)$ is either (a) identically zero (in the case that $\widetilde{\alpha}_1 \circ \cdots \circ \widetilde{\alpha}_{i-1} \circ \rho$ falsifies A_i for some i – in this case say that ρ is "miss"), or (b) functionally equivalent to an AND of s literals (in the complement case that ρ is "hit"). Thus, in any case we have

$$\mathbb{1}(\rho \text{ is a "hit" w.r.t. } \ell, Q, \alpha) \leq 2^s \mathop{\mathbb{E}}_x f_\ell^{Q,\alpha}(\rho \circ x),$$

where x is a uniformly random vector in $\{0,1\}^n$ which we introduce purely for the sake of our analysis.

Our next step will be to simplify the event (*1.*) in Lemma 4.1 by refactoring quantifiers.

Proposition 4.2. *The event (1.) in Lemma 4.1 is equivalent to the following event:*

1'. $\ell = (\ell_1, \ldots \ell_r)$ is the first (with respect to the lexicographic ordering on increasing tuples in $[m]^r$) tuple such that for all $i \in [r]$, A_{ℓ_i} is not falsified by $\widetilde{\alpha}_1 \circ \cdots \circ \widetilde{\alpha}_{i-1} \circ \rho$.

Proof. Fix a restriction ρ and some data r, Q, α. For $k \in [m]$, We introduce a notational shorthand $E_i(k)$ to refer the the event that A_k is not falsified by $\widetilde{\alpha}_1 \circ \cdots \circ \widetilde{\alpha}_{i-1} \circ \rho$.

We consider two different methods for generating an increasing tuple of indices. First, we define $\ell = (\ell_1, \ldots, \ell_r)$ by letting ℓ_1 be the first index such that $E_1(\ell_1)$, and, for $i > 1$, letting ℓ_i be the first index larger than ℓ_{i-1} such that $E_i(\ell_i)$. Second, we define $\ell' = (\ell'_1, \ldots, \ell'_r)$ as the lexicographically-first increasing tuple such that $E_i(\ell_i)$ for all i.

In the case that either of these tuples are well-defined (i.e. there is at least one increasing tuple satisfying the conditions), we show that they are the same. Seeking contradiction, suppose $\ell \neq \ell'$, and let j be the first coordinate in which they differ. If $\ell_j < \ell'_j$ then we could get a lexicographically-smaller increasing tuple ℓ' that still satisfies the conditions by replacing ℓ'_j with ℓ_j. On the other hand, $\ell_j > \ell'_j$ would clearly contradict our procedure for defining ℓ. □

In order to check that a tuple is indeed the first tuple "hit" by ρ, we introduce the function

$$\phi_\ell^{\alpha,Q}(x) := \prod_{\ell' < \ell} (1 - f_{\ell'}^{Q,\alpha}(x)),$$

where the product is taken over all increasing tuples $\ell' \in [m]^r$ which are lexicographically smaller than ℓ. Recalling our discussion from earlier, we have that, as a function of x, the restricted function $\phi_\ell^{Q,\alpha}(\rho \circ x)$ is identically 1 if ρ is a "miss" with respect to ℓ', Q, α for all $\ell' < \ell$.

Using Lemma 4.1, Proposition 4.2, and the recent discussion, we have that for fixed data r, s, Q, α,

$$\begin{aligned}
&\mathbb{1}(\rho \text{ satisfies events } (1.) \text{ and } (2.) \text{ w.r.t. } Q, \alpha) \\
&\leq \sum_\ell \mathbb{1}(\rho \text{ is a "hit" w.r.t. } \ell, Q, \alpha) \cdot \mathbb{1}(\rho \text{ is a "miss" for } \ell' < \ell) \cdot \mathbb{1}\left(\text{Stars}(\rho) \cap \left(\bigcup_{i=1}^r V_{\ell_i}\right) = \widetilde{Q}(\ell)\right) \\
&\leq \sum_\ell 2^s \mathop{\mathbb{E}}_x f_\ell^{Q,\alpha}(\rho \circ x) \cdot \phi_\ell^{Q,\alpha}(\rho \circ x) \cdot \mathbb{1}\left(\text{Stars}(\rho) \cap \left(\bigcup_{i=1}^r V_{\ell_i}\right) = \widetilde{Q}(\ell)\right) \\
&\leq \sum_\ell 2^s \mathop{\mathbb{E}}_x f_\ell^{Q,\alpha}(\rho \circ x) \cdot \phi_\ell^{Q,\alpha}(\rho \circ x) \cdot \mathbb{1}\Big(\text{Stars}(\rho) \supseteq \widetilde{Q}(\ell)\Big) \\
&= 2^s \mathop{\mathbb{E}}_x \mathop{\mathbb{L}}_\ell f_\ell^{Q,\alpha}(\rho \circ x) \to \mathbb{1}\Big(\text{Stars}(\rho) \supseteq \widetilde{Q}(\ell)\Big),
\end{aligned}$$

where the sum is over all increasing tuples $\ell \in [m]^r$. Thus, we have proved the main technical lemma of this section (which can be compared with Lemma 1.1):

Lemma 4.3. *Let $F(x) = \bigvee_{i=1}^m A_i(x)$ by a DNF of width w. Then for any restriction $\rho \in \{0,1,*\}^n$,*

$$\mathbb{1}(\text{CDT}(F,\rho) \text{ has depth } \geq d) \leq B_d(\rho) := \sum_{s=d}^{d+w-1} \sum_{r=1}^{d} \sum_{Q,\alpha} 2^s \mathop{\mathbb{E}}_x \mathop{\mathbb{L}}_\ell f_\ell^{Q,\alpha}(\rho \circ x) \to \mathbb{1}\Big(\text{Stars}(\rho) \supseteq \widetilde{Q}(\ell)\Big),$$

where the inner summation is over all $Q = (Q_1, \ldots Q_r) \subseteq \binom{[w]}{s_1} \times \cdots \times \binom{[w]}{s_r}$ and all $\alpha = (\alpha_1, \ldots, \alpha_{r-1}) \in \{0,1\}^{s_1} \times \ldots \{0,1\}^{s_{r-1}}$, for all choices of $s_1, \ldots s_r \geq 1$ such that $s_1 + \cdots + s_r = s$. The decision list is indexed over the set of all increasing tuples $\ell \in [m]^r$, which is ordered lexicographically.

The important features of the bounding expression B_d are summarized by the next two claims.

Proposition 4.4. *The expression $B_d(\rho)$ from Lemma 4.3 is the sum of at most $(8w)^{d+w}$ functions of the form*

$$\mathop{\mathbb{E}}_x \mathop{\mathbb{L}}_i f_i(\rho \circ x) \to \mathbb{1}(\text{Stars}(\rho) \supseteq S_i),$$

where x is a uniformly random vector in $\{0,1\}^n$, each S_i is a set of size at most $d + w$, and each f_i is an AND of at most wd literals.

Proof. For each s, we count the number of choices for the data r, Q, α, and also account for the scaling factor 2^s. There are at most 2^s choices for $r, s_1, \dots s_r$ such that $s_1 + \cdots + s_r = s$. Finally, There are at most 2^s choices for α, and at most $w^{s_1} \cdots w^{s_r} = w^s$ choices for Q. Summing over s, we get

$$\sum_{s=d}^{d+w-1} (8w)^s \leq (8w)^{d+w}.$$

The claims about the form of S_i and f_i correspond the the facts that by construction, $\widetilde{Q}(\ell)$ is a set of size $s \leq d + w$, and $f_\ell^{Q,\alpha}$ is a product of (restrictions of) $r \leq d$ terms from F. □

Proposition 4.5. *Suppose $\rho \in \{0, 1, *\}^n$ is a p-random restriction with $p \leq \frac{1}{16w}$. Then*

$$\mathop{\mathbb{E}}_{\rho} B_d(\rho) \leq 2 \cdot (8pw)^d$$

Proof. We first argue for a single function of the form $\mathbb{E}_x \mathbb{L}_i f_i(\rho \circ x) \to \mathbb{1}(\mathrm{Stars}(\rho) \supseteq S_i)$ such that $|S_i| = s$ for all i, and then sum.

We imagine sampling ρ by first making a random selection to determine the locations of the stars, and then randomly setting the non-star coordinates to 0 or 1 using a separate random process. That is, let $\rho = T[y, \star]$, where T is a p-random string in $\{0, 1\}^n$ and y is a uniformly random vector in $\{0, 1\}^n$. Averaging over ρ gives

$$\mathop{\mathbb{E}}_{\rho} \mathop{\mathbb{E}}_{x} \mathop{\mathbb{L}}_{i} f_i(\rho \circ x) \to \mathbb{1}(\mathrm{Stars}(\rho) \supseteq S_i) = \mathop{\mathbb{E}}_{T} \mathop{\mathbb{E}}_{y} \mathop{\mathbb{E}}_{x} \mathop{\mathbb{L}}_{i} f_i(T[y, x]) \to \prod_{j \in S_i} T_j.$$

At this point, we make the key observation that for any fixed vector T, the distribution $T[y, x]$ is simply the uniform distribution over $\{0, 1\}^n$. In particular, it does not depend on T. So, letting U be a uniformly random vector in $\{0, 1\}^n$, the above is equivalent to

$$\mathop{\mathbb{E}}_{T} \mathop{\mathbb{E}}_{U} \mathop{\mathbb{L}}_{i} f_i(U) \to \prod_{j \in S_i} T_j = \mathop{\mathbb{E}}_{T} \sum_i \pi_i \prod_{j \in S_i} T_j = \sum_i \pi_i \mathop{\mathbb{E}}_{T} \prod_{j \in S_i} T_j,$$

where π_i is the probability, that upon random input U, i is the first index such that $f_i(U) = 1$. We conclude this estimate by noting that $\mathbb{E}_T \prod_{j \in S_i} T_j = p^s$, and so $\sum_i \pi_i p^s \leq p^s$.

Summing this bound over all summands in the expression $B_d(\rho)$ gives

$$\mathop{\mathbb{E}}_{\rho} B_d(\rho) \leq \sum_{s=d}^{d+w-1} (8w)^s \cdot p^s \leq (8pw)^d \sum_{i=0}^{\infty} (8pw)^{-i} = \frac{(8pw)^d}{1 - 8pw}.$$ □

We observe that the argument above is robust to the use of less-than-perfect random selection procedures.

Theorem 4.6. *Let $F(x) = \bigvee_{i=1}^{m} A_i(x)$ by a DNF of width w. Suppose T is a $(d+w)$-wise p-bounded distribution over $\{0,1\}^n$, and y is a random vector in $\{0,1\}^n$. If $p \leq \frac{1}{16w}$, Then the random restriction $\rho = T[y, \star]$ satisfies*

$$\mathbb{P}(\mathrm{CDT}(F, \rho) \text{ has depth } \geq d) \leq \mathop{\mathbb{E}}_{\rho} B_d(\rho) \leq 2 \cdot (8pw)^d.$$

Proof. We follow the proof of Proposition 4.5 exactly, except that we use the p-boundedness assumption to say that

$$\mathop{\mathbb{E}}_{T} \prod_{j \in S} T_j \leq p^s$$

whenever S is a set of size $|S| = s \leq d + w$. □

5 Further Proofs

Proof of Theorem 1.3

Proof. Let $t := \log(1/p)$. We first generate $Y \in (\{0,1\}^t)^n \cong \{0,1\}^{tn}$ according to a δ-biased distribution. We define the selection vector T in every coordinate via $T_i := Y_{i,1} \wedge Y_{i,2} \wedge \cdots \wedge Y_{i,t}$.

Now, we follow the proof of Proposition 4.5 exactly, except that we use the δ-bias assumption to say that

$$\mathop{\mathbb{E}}_{T} \prod_{j \in S} T_j \leq p^s + \delta$$

whenever S is a set of size s. For the restriction $\rho := T[U, \star]$, this yields

$$\mathbb{P}(\mathrm{CDT}(F, \rho) \text{ has depth } \geq d) \leq 2 \cdot (8pw)^d + \delta \cdot (8w)^{d+w}.$$

Setting $\delta = \varepsilon/(8w)^{d+w}$ results in the desired seedlength

$$O(\log(nt) + \log(1/\delta)) = O(\log n + (w+d) \cdot \log w + \log(1/\varepsilon)).$$ □

Proof of Corollary 1.4

Proof. We first restrict according to a selection vector $T^1 \in \{0,1\}^n$ which we draw directly from a δ-biased distribution. Let $\rho^1 = T^1[U, \star]$. By a simple union bound over all terms, the probability that $F(\rho^1 \circ x)$ has a surviving term of width at least w is at most

$$((3/4)^w + \delta)m.$$

We then compose with a restriction ρ^2 that satisfies the switching lemma on DNFs of width w as given by Theorem 1.3. The overall restriction is $\rho := \rho^1 \circ \rho^2 = T^1 \wedge T^2[U, \star]$. and the overall selection vector is $T := T^1 \wedge T^2$. By picking parameters $\delta = \Theta(\varepsilon/m)$ and $w = \Theta(\log(m/\varepsilon))$, we can obtain overall error ε and obtain the overall desired seedlength. □

Proof of Theorem 1.7

Proof Sketch. The construction of the pseudorandom generator is essentially the construction we outline in Section 1. Specifically, the output of the generator is given by

$$G_r(Z) := \tau^{(1)} \circ \tau^{(2)} \circ \cdots \circ \tau^{(r)} \circ Z,$$

where Z is a δ-biased distribution and each $\tau^{(i)}$ is an independent copy of a pseudorandom restriction defined as follows. We set

$$\rho^0 := X \text{ and } \rho^j := T^j[\star, \rho^{j-1}],$$

and $\tau := \rho^D = T^1 \wedge \cdots \wedge T^D[\star, X]$, where T^1 and X are (independently) drawn from a δ-biased distribution, and for $j \geq 2$ we (independently) generate each T^j according to a δ-biased distribution $Y \in (\{0,1\}^t)^n$ in the manner described in the proof of Theorem 1.3.

Fix a size-m, depth-D AC^0-function $F : \{0,1\}^n \to \{0,1\}$, where we assume $m \geq n$. We also assume $D \geq 2$, as the $D = 1$ case simply corresponds to ANDs and ORs of literals, which is easy.

The probability that, as a function of X, the restricted function $F(\tau \circ U)$ cannot be expressed as a depth-w decision tree is at most

$$\gamma_1 := ((3/4)^w + \delta) \cdot m + (D-1) \cdot m \cdot (2 \cdot (8pw)^w + \delta \cdot (8w)^{2w}),$$

where $p = 1/2^t$ and w is a parameter we are free to choose. By choosing $w \geq \Omega(\log(m))$ and $t = \log(w) + 4$, we can say that

$$\gamma_1 \leq 2^{-\Omega(w)} + \delta \cdot w^{O(w)}.$$

Now, it is easy to argue that X fools depth-w decision trees with error $\delta \cdot 2^w$, since such decision trees can be expressed as a sum over at most 2^w ANDs of literals. So, overall, the distribution $\tau \circ U$ fools such functions F with error at most

$$\gamma_2 := \gamma_1 + \delta \cdot 2^w \leq 2^{-\Omega(w)} + \delta \cdot w^{O(w)}.$$

Now, we bound the overall error of the generator by a simple hybrid argument:

$$\begin{aligned} |\mathbb{E}\, F(G_r(Z)) - \mathbb{E}\, F(U)| &\leq |\mathbb{E}\, F(G_r(Z)) - \mathbb{E}\, F(G_r(U))| + \sum_{i=0}^{r-1} \left|\mathbb{E}\, F(G_i(\rho^{(i+1)} \circ U)) - \mathbb{E}\, F(G_i(U))\right| \\ &\leq |\mathbb{E}\, F(G_r(Z)) - \mathbb{E}\, F(G_r(U))| + r \cdot \gamma_2. \end{aligned}$$

It remains to bound the error of the base of the hybrid argument, $|\mathbb{E}\, F(G_r(Z)) - \mathbb{E}\, F(G_r(U))|$. To do this, we set r large enough so that, as a function of Z, it is highly likely that $F(G_r(Z))$ depends only on a few bits of Z, say w. In this case, Z fools $F(G_r(Z))$ with error at most $\delta \cdot 2^w$. We can set $r = \Theta(\log(n)\log(w)/p^{D-1})$ and obtain the following claim, which is proved using standard tail-bound arguments for δ-biased random variables.

Claim 5.1.

$$\mathbb{P}\big(|\mathrm{Stars}(\tau^{(1)} \circ \tau^{(2)} \circ \cdots \circ \tau^{(r)})| \geq w\big) \leq 2^{-\Omega(w)} + \delta^{\Omega(1/D \log w)} \cdot w^{O(Dw)}.$$

Given this claim, we can bound the overall error of the pseudorandom generator by

$$\gamma_3 := 2^{-\Omega(w)} + \delta^{\Omega(1/D \log w)} \cdot w^{O(Dw)} + r \cdot \gamma_2 = 2^{-\Omega(w)} + \delta^{\Omega(1/D \log w)} \cdot w^{O(Dw)}.$$

By choosing $w := \Theta(\log(m/\varepsilon))$ and $\log(1/\delta) := \Theta(D^2 \log(m/\varepsilon)(\log\log(m/\varepsilon))^2)$, we get error ε with an overall seedlength of

$$O(r \cdot D \cdot \log(1/\delta)) \leq O(\log(m/\varepsilon))^D \cdot \log(n) \cdot (\log\log(m/\varepsilon))^3. \qquad \square$$

Proof Sketch for Claim 5.1. We set up a matrix of boolean random variables W with r rows and m columns, corresponding to $W_{ij} := \mathbb{1}\big(j \in \mathrm{Stars}(\tau^{(i)})\big)$. We note that each W_{ij} is, as a function of the underlying δ-biased variables used to generate selection vectors, a negation of an AND of literals. As a result, the product of any subset of variables from W is in fact a read-once CNF, and it is shown in [DETT10] that read-once CNFs with c clauses are fooled by δ-biased distributions with error $\delta' := \delta^{\Omega(1/\log(c))}$. It can also be shown by an elementary argument (i.e. just expand the product into 2^c terms) that such products are also fooled with error $\delta \cdot 2^c$. We note that for each variable,

$$\big|\mathbb{E}[W_{ij}] - (1 - p^{D-1}/2)\big| \leq \delta.$$

Define $q := p^{D-1}/2$, and note for later that $1/q \leq O(w)^D$. We set $r := 4\log(n)\log(w)/q$.

Define $n_0 := n$ and

$$n_i := |\mathrm{Stars}(\tau^{(1)} \circ \tau^{(2)} \circ \cdots \circ \tau^{(i)})| = \sum_{j=1}^{n} \prod_{i' \leq i} W_{i'j}.$$

We argue in two stages that $n_{r/2} \leq w^{O(D)}$ with high probability, and then (conditioning on this likely event) that $n_r \leq w$ with high probability.

For the first stage, we use a standard bound on the k-th moment of a sum $Z := \sum_{j=1}^{n} Z_j$ of independent, mean-zero random variables $Z_i \in [-1, +1]$, (see e.g. the proof of theorem 4 in [SSS95]): for even k we have $\mathbb{E}\, Z^k \leq (kn)^{k/2}$. Now, if we consider the i-th row in W, and set $Z_j := W_{ij} - (1 - q)$, then we obtain the k-th moment bound $\mathbb{E}\, Z^k \leq (kn)^{k/2} + \delta \cdot 2^k \cdot n^k$. Applying a Markov argument, we derive the probability bound

$$\mathbb{P}(n_i \geq (1 - q/2)n_{i-1}) \leq \left(\frac{k}{n_{i-1}} \frac{4}{q^2}\right)^{k/2} + \delta \cdot (4/q)^k.$$

Observe that $(1 - q/2)^{r/2} \cdot n \leq 1$, and set $k := \Theta(w)$. By a simple union-bound argument, we conclude that we must have $n_{r/2} \leq w^{O(D)}$, except with probability at most

$$(r/2) \cdot (2^{-\Omega(w)} + \delta \cdot w^{O(Dw)}) \leq 2^{-\Omega(w)} + \delta \cdot w^{O(Dw)}.$$

Now for the second stage, we fix the first $r/2$ rows of W, and condition on the event that $n_{r/2} \leq w^{O(D)}$. We consider the chance that there is a set of w columns, among the $n_{r/2}$ columns which are still live, such that W has all 1's down each of these w columns. This event can be expressed as a read-once CNF with $c = w \cdot r/2 \leq w^{O(D)}$ clauses. Thus the chance that this event occurs (for a specific set of w columns) is at most

$$(1-q)^{w \cdot r/2} + \delta^{\Omega(1/\log(c))} \leq e^{-w \log(n) \log(w)} + \delta^{\Omega(1/D \log w)}.$$

We finish the estimate by union-bounding over all $\binom{n_{r/2}}{w} \leq w^{O(Dw)}$ choices of w columns. $\square$.

Proof of Theorem 1.8

Proof. Use Theorem 1.4, with $d = \Theta(\log(1/\gamma))$ and $p = \Theta(1/\log(m))$, to select T. We generate $X \in \{0,1\}^n$ using a δ-biased distribution. We exchange the order of expectations and estimate

$$\mathop{\mathbb{E}}_{T} \mathop{\mathbb{E}}_{U} \left(\mathop{\mathbb{E}}_{X} F(T[U,X]) - \mathop{\mathbb{E}}_{U'} F(T[U,U']) \right).$$

Let $\sigma := T[U, \star]$. Now, whenever F_σ successfully collapses to a depth-d decision tree, we have

$$\left| \mathop{\mathbb{E}}_{X} F(\sigma \circ X) - \mathop{\mathbb{E}}_{U'} F(\sigma \circ U') \right| \leq \delta \cdot 2^s.$$

This is because any decision tree of depth d can be expressed as a sum of at most 2^d ANDs of literals (we get one AND for each path in the decision tree that reaches a leaf which outputs 1). Whenever F_σ fails to collapse, we instead bound this quantity trivially by 1. We can set $\log(1/\delta) := \Theta(\log(1/\gamma))$ to get an overall error bound of γ. $\square$

Proof of Lemma 1.10

Proof. Let $F(x) = \bigvee_{i=1}^m A_i(x)$ be a width-w DNF, and let ρ be a p-random restriction. For each i, define $\phi_i(x) := \prod_{j<i}(1 - A_i(x))$, and let $\sigma_i \in \{0,1,*\}^n$ be the (unique) restriction that sets all the variables in A_i so that A_i becomes satisfied, and does not set any other variables.

We observe that

$$\mathbb{1}(A_i(\rho) \text{ is useful in } F(\rho)) = A_i(\rho \circ \sigma_i) \cdot \mathbb{1}(\phi_i(\rho \circ \sigma_i) \not\equiv 0)$$

and, with some consideration, that

$$\mathop{\mathbb{E}}_{\rho} A_i(\rho \circ \sigma_i) \cdot \mathbb{1}(\phi_i(\rho \circ \sigma_i) \not\equiv 0) \leq \left(\frac{2}{1-p} \right)^w \mathop{\mathbb{E}}_{\rho} \mathbb{1}(A_i(\rho) \equiv 1) \cdot \mathbb{1}(\phi_i(\rho) \not\equiv 0).$$

Thus, the average number of useful terms in $F(\rho)$ is bounded by

$$\left(\frac{2}{1-p} \right)^w \sum_{i=1}^m \mathop{\mathbb{E}}_{\rho} \mathbb{1}(A_i(\rho) \equiv 1) \cdot \mathbb{1}(\phi_i(\rho) \not\equiv 0) \leq \left(\frac{2}{1-p} \right)^w,$$

since the events in the sum are disjoint. $\square$

References

[AGHP92] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures & Algorithms*, 3(3):289–304, 1992.

[ALWZ20] Ryan Alweiss, Shachar Lovett, Kewen Wu, and Jiapeng Zhang. Improved bounds for the sunflower lemma. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 624–630, 2020.

[AS17] Sergei Artemenko and Ronen Shaltiel. Pseudorandom generators with optimal seed length for non-boolean poly-size circuits. *ACM Transactions on Computation Theory (TOCT)*, 9(2):1–26, 2017.

[AW85] Miklos Ajtai and Avi Wigderson. Deterministic simulation of probabilistic constant depth circuits. In *26th Annual Symposium on Foundations of Computer Science (sfcs 1985)*, pages 11–19. IEEE, 1985.

[Baz09] Louay MJ Bazzi. Polylogarithmic independence can fool DNF formulas. *SIAM Journal on Computing*, 38(6):2220–2272, 2009.

[BDSG+18] Marshall Ball, Dana Dachman-Soled, Siyao Guo, Tal Malkin, and Li-Yang Tan. Non-malleable codes for small-depth circuits. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 826–837. IEEE, 2018.

[Bra08] Mark Braverman. Polylogarithmic independence fools AC0 circuits. *Journal of the ACM (JACM)*, 57(5):1–10, 2008.

[CHHL19] Eshan Chattopadhyay, Pooya Hatami, Kaave Hosseini, and Shachar Lovett. Pseudorandom generators from polarizing random walks. *Theory of Computing*, 15(1):1–26, 2019.

[DETT10] Anindya De, Omid Etesami, Luca Trevisan, and Madhur Tulsiani. Improved pseudorandom generators for depth 2 circuits. In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, pages 504–517. Springer, 2010.

[DHH19] Dean Doron, Pooya Hatami, and William M. Hoza. Near-optimal pseudorandom generators for constant-depth read-once formulas. In *34th Computational Complexity Conference (CCC 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.

[GMR13] Parikshit Gopalan, Raghu Meka, and Omer Reingold. DNF sparsification and a faster deterministic counting algorithm. *Computational Complexity*, 22(2):275–310, 2013.

[GW14] Oded Goldreich and Avi Wigderson. On derandomizing algorithms that err extremely rarely. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 109–118, 2014.

[Hås86] Johan Håstad. Almost optimal lower bounds for small depth circuits. In *Proceedings of the eighteenth annual ACM symposium on Theory of computing*, pages 6–20, 1986.

[Hås14] Johan Håstad. On the correlation of parity and small-depth circuits. *SIAM Journal on Computing*, 43(5):1699–1708, 2014.

[HS19] Prahladh Harsha and Srikanth Srinivasan. On polynomial approximations to AC0. *Random Structures & Algorithms*, 54(2):289–303, 2019.

[Juk12] Stasys Jukna. *Boolean function complexity: advances and frontiers*, volume 27. Springer Science & Business Media, 2012.

[LMN93] Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, fourier transform, and learnability. *Journal of the ACM (JACM)*, 40(3):607–620, 1993.

[LMZ20] Shachar Lovett, Raghu Meka, and Jiapeng Zhang. Improved lifting theorems via robust sunflowers. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 27, page 48, 2020.

[LWZ20] Shachar Lovett, Kewen Wu, and Jiapeng Zhang. Decision list compression by mild random restrictions. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 247–254, 2020.

[MP20] Ian Mertz and Toniann Pitassi. Lifting: As easy as 1,2,3. In *Electronic Colloquium on Computational Complexity (ECCC)*, 2020.

[Nis91] Noam Nisan. Pseudorandom bits for constant depth circuits. *Combinatorica*, 11(1):63–70, 1991.

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM journal on computing*, 22(4):838–856, 1993.

[Raz95] Alexander A Razborov. Bounded arithmetic and lower bounds in boolean complexity. In *Feasible Mathematics II*, pages 344–386. Springer, 1995.

[Raz09] Alexander Razborov. A simple proof of Bazzi's theorem. *ACM Transactions on Computation Theory (TOCT)*, 1(1):1–5, 2009.

[Ros19] Benjamin Rossman. Criticality of regular formulas. In *34th Computational Complexity Conference (CCC 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.

[SS16] Ronen Shaltiel and Jad Silbak. Explicit list-decodable codes with optimal rate for computationally bounded channels. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2016)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2016.

[SSS95] Jeanette P Schmidt, Alan Siegel, and Aravind Srinivasan. Chernoff–Hoeffding bounds for applications with limited independence. *SIAM Journal on Discrete Mathematics*, 8(2):223–250, 1995.

[ST17] Rocco A. Servedio and Li-Yang Tan. Deterministic search for CNF satisfying assignments in almost polynomial time. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 813–823. IEEE, 2017.

[ST19] Rocco A. Servedio and Li-Yang Tan. Improved pseudorandom generators from pseudorandom multi-switching lemmas. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.

[Tal17] Avishay Tal. Tight bounds on the fourier spectrum of AC0. In *32nd Computational Complexity Conference (CCC 2017)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017.

[Tel19] Roei Tell. Improved bounds for quantified derandomization of constant-depth circuits and polynomials. *Computational Complexity*, 28(2):259–343, 2019.

[Tha09] Neil Thapen. Notes on switching lemmas. Unpublished Manuscript, 2009.

[TX13] Luca Trevisan and Tongke Xue. A derandomized switching lemma and an improved derandomization of AC0. In *2013 IEEE Conference on Computational Complexity*, pages 242–247. IEEE, 2013.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il