Electronic Colloquium on Computational Complexity, Revision 2 of Report No. 174 (2014)

Tight Bounds on The Fourier Spectrum of $\mathbf{AC^0}$

Avishay Tal*

January 9, 2017

Abstract

We show that $\mathbf{AC^0}$ circuits on n variables with depth d and size m have at most $2^{-\Omega(k/\log^{d-1} m)}$ of their Fourier mass at level k or above. Our proof builds on a previous result by Håstad (SICOMP, 2014) who proved this bound for the special case $k = n$. Our result improves the seminal result of Linial, Mansour and Nisan (JACM, 1993) and is tight up to the constants hidden in the Ω notation.

As an application, we improve Braverman's celebrated result (JACM, 2010). Braverman showed that any $r(m, d, \varepsilon)$-wise independent distribution ε-fools $\mathbf{AC^0}$ circuits of size m and depth d, for

$$r(m, d, \varepsilon) = O(\log(m/\varepsilon))^{2d^2+7d+3} .$$

Our improved bounds on the Fourier tails of $\mathbf{AC^0}$ circuits allows us to improve this estimate to

$$r(m, d, \varepsilon) = O(\log(m/\varepsilon))^{3d+3} .$$

In contrast, an example by Mansour (appearing in Luby and Velickovic's paper - Algorithmica, 1996) shows that there is a $\log^{d-1}(m) \cdot \log(1/\varepsilon)$-wise independent distribution that does not ε-fool $\mathbf{AC^0}$ circuits of size m and depth d. Hence, our result is tight up to the factor 3 in the exponent.

*Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel. avishay.tal@weizmann.ac.il. Supported by an Adams Fellowship of the Israel Academy of Sciences and Humanities, by an ISF grant and by the I-CORE Program of the Planning and Budgeting Committee.

ISSN 1433-8092

Contents

1 Introduction

In this paper we discuss Boolean circuits in which every gate computes an unbounded fan-in OR or AND function of its inputs, and every leaf is marked with a literal from $x_1, \ldots, x_n, \neg x_1, \ldots, \neg x_n$. The number of gates in the circuit is called the circuit size and is denoted by m. The longest path in the circuit is called the circuit depth and is denoted by d. $\mathbf{AC^0}$ is the class of functions that can be realized by Boolean circuits of constant depth and polynomial size. (We also call Boolean circuits of polynomial size and constant depth $\mathbf{AC^0}$ circuits).

The study of bounded depth circuits flourished in the 1980s, culminating in the tight $\exp(\Omega(n^{1/(d-1)}))$ size lower bound for Boolean circuits of depth d computing the parity function [Ajt83, FSS84, Yao85, Hås86].[1] The main idea behind this lower bound was the following - Boolean circuits with size m and depth d become constant with high probability under random restrictions keeping each variable alive with probability $p = 1/O(\log m)^{d-1}$. In contrast, the parity function does not become a constant with probability at least 0.5 as long as $pn \geq 1$. Since the restricted circuit should compute the restricted function, we reach a contradiction if $m = \exp(o(n^{1/(d-1)}))$. The main idea is carried through a sequence of $d-1$ steps, where in each step the circuit depth is decreased by one with high probability, by applying Håstad's switching lemma [Hås86].

In their seminal paper, Linial, Mansour, and Nisan [LMN93] showed that $\mathbf{AC^0}$ circuits can be learned in quasipolynomial time, $n^{O(\log^d n)}$, using random samples, under the uniform distribution. They combined Håstad's switching lemma with Fourier analysis, to show that $\mathbf{AC^0}$ circuits may be well approximated (in L_2 norm) by low degree polynomials, namely polynomials of degree $O(\log^d n)$. Boppana [Bop97] improved their bound on the degree to $O(\log^{d-1} n)$, which is optimal for constant error. The existence of an approximating low degree polynomial implies a learning algorithm for $\mathbf{AC^0}$ circuits, using random examples. For polynomial size **DNFs** (depth 2 circuits), Mansour [Man95] showed that only $n^{O(\log\log n)}$ out of the $\binom{n}{\leq O(\log n)}$ monomials are needed to approximate the **DNF**, and achieved a $n^{O(\log\log n)}$ time learning algorithm for **DNFs**, using membership queries, via the Goldreich-Levin [GL89], Kushilevitz-Mansour [KM93] method.

The main technical result in [LMN93] was a bound on the Fourier tails of Boolean circuits. Namely, for any circuit f of size m and depth d,

$$\sum_{S\subseteq[n]:|S|\geq k} \hat{f}(S)^2 \leq m \cdot 2^{-\Omega(k^{1/d})} ,$$

where the LHS is called the Fourier tail of f at level k. This was later improved by Håstad [Hås01] to

$$\sum_{S\subseteq[n]:|S|\geq k} \hat{f}(S)^2 \leq \max\{2^{-\Omega((k/\log m)^{1/(d-1)})}, 2^{-\Omega(k/\log^{d-1}(m))}\} ,$$

which is tight for $k \leq O(\log^d(m))$, however not for larger values of k. Recently, Håstad [Hås14] and Impagliazzo, Matthews, and Paturi [IMP12] showed that any Boolean circuit f agrees with parity on at most a $1/2 + 2^{-n/O(\log m)^{d-1}}$ fraction of the inputs. In other words, they showed that $|\hat{f}([n])| \leq 2^{-n/O(\log m)^{d-1}}$.

[1] Lower bounds for the **DNF**-size of the parity function were known long before [Lup61].

1.1 Our Results

Based on the main lemma of [Hås14], we extend the results of [Hås14, IMP12] for all $k \in [0, n]$ and show the following.

Theorem 1.1 (Main Theorem)**.** *Let f be an Boolean circuit with depth d and size m. Then,*

$$\sum_{S:|S|\geq k} \hat{f}(S)^2 \leq 2 \cdot 2^{-k/O(\log m)^{d-1}} .$$

A few things to note first. Increasing k from 0 to n, the first time that Theorem 1.1 is meaningful is at $k = \Theta(\log^{d-1}(m))$, which is only marginally better than in [LMN93] and exactly the same as in [Bop97, Hås01]. Nonetheless, for larger values, our bound decreases much faster, and in particular for $m = \text{poly}(n)$ we get a $2^{-n/\text{poly}\log(n)}$ tail at level $k = \Omega(n)$ as opposed to a $2^{-\Omega((n/\log n)^{1/(d-1)})}$ tail by [Hås01]. In addition, while [Hås14] and [IMP12] give bounds on an individual coefficient, $|\hat{f}(S)|$, we give bounds on the sum of $\exp(\Omega(n))$ many squares of coefficients (e.g., for $k = n/2$).

We point out that the results of [Hås14], [IMP12], and ours are quite surprising, considering the fact that most proofs for Boolean circuits follow by induction on the depth d; performing $d-1$ consecutive steps of Håstad's switching lemma. Our main theorem is equivalent to saying that degree $O(\log^{d-1}(m) \cdot \log(1/\varepsilon))$ polynomials ε-approximates Boolean circuits of size m and depth d, as opposed to degree $O(\log^d(m/\varepsilon))$ polynomials by [LMN93]. It seems at first glance that one *must* pay a factor of $\log(m/\varepsilon)$ for each step in the induction to ensure error at most ε, thus resulting in degree at least $\log^{d-1}(m/\varepsilon)$. However, Håstad and Impagliazzo et al. managed to avoid that. Håstad performs random restrictions keeping each variable alive with probability $p = 1/O(\log m)$ that does not depend on ε. This only guarantee that the switching succeeds with probability $1 - 1/\text{poly}(m)$, as opposed to probability of $1 - \varepsilon/m$ in the original proof of [LMN93]. However, in the cases where the switching "fails", Håstad fixes D additional variables using a decision tree of depth D. Under these additional fixings, the probability that the switching fails reduces to $m \cdot 2^{-D}$. We show that the parameters p and D translate into a **multiplicative** term of $1/p$ and an **additive** term of D in the degree, correspondingly. Choosing D to be roughly $\log(m/\varepsilon)$ and applying induction gives the desired dependency on m and ε.

Theorem 1.1 shows that the Fourier tail above level k decreases exponentially fast in k. In Section 5, we show that such behavior is related to three other properties of concentration. We establish many connections between these four properties, and show that three of them are essentially equivalent. We think that these connections are of independent interest.[2] As a result of these connections we establish the following theorem.

Theorem 1.2. *Let f be an Boolean circuit with depth d and size m. Then,*

1. *For all k, p, if ρ is a p-random restriction, then $\mathbf{Pr}_\rho[\deg(f|_\rho) \geq k] \leq O(p \cdot \log^{d-1}(m))^k$.*
2. *For all k,*

$$\sum_{S:|S|=k} |\hat{f}(S)| \leq O(\log^{d-1}(m))^k . \tag{1}$$

[2] In fact, some of these connections have been already used in the context of de Morgan formulae [Tal14].

3. *f is ε-concentrated on at most $2^{O(\log\log(m)\cdot\log^{d-1}(m)\cdot\log(1/\varepsilon))}$ Fourier coefficients.*

In Section 6, we show that Equation (1) gives new proofs for the following known results:

- Correlation bounds for the Majority function. If f is a size m depth d circuit, then $\mathbf{Pr}[f(x) = \mathrm{MAJ}(x)] \leq \frac{1}{2} + \frac{O(\log^{d-1}(m))}{\sqrt{n}}$. Our result holds for $\log^{d-1}(m) = O((n/\log n)^{1/3})$, which is an artifact of the proof. This result was originally proved by Smolensky [Smo93] (see also [Fil10]) and by O'Donnell and Wimmer [OW07], for the entire range of parameters.

- Boolean circuits cannot distinguish between fair coins and coins with bias at most $\frac{1}{O(\log^{d-1}(m))}$. This result was previously proved by Cohen, Ganor and Raz [CGR14], improving the results of Aaronson [Aar10], and Shaltiel and Viola [SV10].

1.2 Applications to Pseudorandomness and Learning

Since the result of [LMN93] had many applications, our main theorem improves some of them as well.

k-wise independence fools bounded-depth circuits. The most significant improvement is to the work of Braverman [Bra10] who proved a longstanding conjecture, showing that poly-logarithmic independent distributions fool $\mathbf{AC^0}$ circuits. To be more precise, Braverman showed that any k-wise independent distribution, where $k = O(\log(m/\varepsilon))^{2d^2+7d+3}$, ε-fools circuits of size m and depth d. In addition, it was long known [LV96] that k must be larger than $\Omega(\log^{d-1}(m) \cdot \log(1/\varepsilon))$; otherwise, there is a k-wise independent distribution that is ε-distinguishable from the uniform distribution by a depth d, size m circuit. Our theorem improves Braverman's bounds to $k = O\left(\log(m/\varepsilon)\right)^{3d+3}$, answering an open question posed by Braverman on the affirmative. In particular, our result is non-trivial for polynomial size circuits of depth $d \leq 0.3\log(n)/\log\log(n)$. Since $\mathbf{NC^1}$ circuits can be computed by Boolean circuits of depth $O(\log(n)/\log\log(n))$ and polynomial size, constructing a non trivial PRG for all $d = O(\log(n)/\log\log(n))$ is a major open challenge. While the dependence of k on m and d is close to optimal, we conjecture that the dependence on ε could be much better.[3]

Conjecture 1. *Any k-wise independence ε-fools circuits of size m and depth d, for*

$$k = (\log m)^{O(d)} \cdot \log(1/\varepsilon) \ .$$

k-wise independence fools DNFs. We improve in Section 4.2 the earlier result of Bazzi [Baz09], who showed that $O(\log^2(m/\varepsilon))$-wise independence ε-fools **DNFs** of size m. We improve the dependence on ε and get that $O(\log(m) \cdot \log(m/\varepsilon))$-wise independence suffices. Note that by [LV96] this is optimal for $\varepsilon \leq 1/m^{\Omega(1)}$. The range $\varepsilon \geq 1/m^{o(1)}$ is still not tightly understood.

[3] We have learned that subsequent to this work, Harsha and Srinivasan [HS16] proved this conjecture.

Task	Ref.	Bound
k-wise ind. fooling **DNFs**	[Baz09]	$k = O(\log^2(m/\varepsilon))$
	This Work	$k = O(\log(m/\varepsilon) \cdot \log(m))$
	Lower Bound	$k \geq \log(m) \cdot \log(1/\varepsilon)$
k-wise ind. fooling $\mathbf{AC^0}$	[Bra10]	$k = O((\log(m/\varepsilon))^{d^2+3d} \cdot (\log m)^{d^2+4d+3})$
	This Work	$k = O\left((\log(m/\varepsilon))^{d} \cdot (\log m)^{2d+3}\right)$
	Lower Bound	$k \geq \log^{d-1}(m) \cdot \log(1/\varepsilon)$
sparse polynomial	[Man95]	sparsity $= (m/\varepsilon)^{O(\log\log(m/\varepsilon)\cdot\log(1/\varepsilon))}$
approximating **DNFs** in L_2	This Work	sparsity $= m^{O(\log\log(m)\cdot\log(1/\varepsilon))}$
sparse polynomial	[LMN93]	sparsity $= 2^{O(\log(n)\cdot\log^d(m/\varepsilon))}$
approximating $\mathbf{AC^0}$ in L_2	[Hås01]	sparsity $= 2^{O(\log(n)\cdot\log^{d-2}(m/\varepsilon)\cdot\log(m)\cdot\log(1/\varepsilon))}$
	This Work	sparsity $= 2^{O(\log\log(m)\cdot\log^{d-1}(m)\cdot\log(1/\varepsilon))}$
	Lower Bound	sparsity $\geq 2^{\Omega(\log^{d-1}(m))}$
PRGs for **DNFs**	[DETT10]	seed $= O(\log n + \log^2(m/\varepsilon) \cdot \log\log(m/\varepsilon))$
	This Work	seed $= O(\log n + \log(m/\varepsilon) \cdot \log(m) \cdot \log\log m)$
PRGs for $\mathbf{AC^0}$	[TX13]	seed $= \tilde{O}(\log^{d+4}(m/\varepsilon))$
	This Work	seed $= \tilde{O}(\log^{d+1}(m/\varepsilon) \cdot \log n)$

Figure 1: Summary of Applications

PRGs for $\mathbf{AC^0}$ and DNFs. We improve the results of De et al. [DETT10] (see Appendix C) and of Trevisan and Xue [TX13] (see Appendix D) that give the best known PRGs for **DNFs** and $\mathbf{AC^0}$ circuits respectively. In the PRG of De et al., we improve the dependency of the seed-length in ε, as seen in Figure 1. Since Trevisan and Xue used De et al.'s generator as a black-box in their construction, we also improve the seed length of their PRG for $\mathbf{AC^0}$ circuits. We observe two more improvements in the Trevisan-Xue generator to reduce the seed-length to $\tilde{O}(\log^{d+1}(m/\varepsilon) \cdot \log(n))$. This seed-length comes closer to the barrier $O(\log^d(m/\varepsilon))$ noted by [TX13].

Sparse polynomial approximations of Boolean circuits. Theorem 1.2 shows that any Boolean circuit f of size m and depth d can be ε-approximated in L_2 by a polynomial $p(x)$ of sparsity $(\log m)^{O(\log^{d-1}(m)\cdot\log(1/\varepsilon))}$, improving the results of [LMN93] and [Man95]. As the inner product on $k = \log^{d-1} m$ variables can be realized by a size $\mathrm{poly}(m)$ depth d circuit, and requires at least $\Omega(2^k)$ coefficients in order to $\Omega(1)$ approximate in L_2, one cannot achieve sparsity $2^{o(\log^{d-1} m)}$.

A table summarizing all of the improvements mentioned above is presented in Figure 1.

1.3 Organization

In Section 2, we lay out some preliminary definitions and results that will be used in the rest of the paper. In Section 3, we prove our main theorem, i.e. Theorem 1.1. In Section 4, we improve Braverman's and Bazzi's results in the field of pseudorandomness. In Section 5, we prove Theorem 1.2, by relating different notions of Fourier concentration. Then, in section 6, we use Theorem 1.2 to deduce simpler proofs for two known results: the inapproximability of the Majority function by bounded-depth circuits, and the indistinguishability of biased-coins from uniform coins by bounded-depth circuits. In Section 7, we give a self-contained new proof of the main lemma in the work of Håstad [Hås14], that plays a crucial role in the proof of Theorem 1.1. This serves two purposes. First, it makes the main result in our paper self-contained. Second, in our opinion, it gives a simpler proof of Håstad's main lemma ([Hås14]).

In the appendices, we revisit the works of Braverman [Bra10] (Appendix B), De et al. [DETT10] (Appendix C), and Trevisan and Xue [TX13] (Appendix D) in the field of pseudorandomness. We show how our main results (Theorem 1.1 and Theorem 1.2) improve these results. Furthermore, we reduce the seed-length of the PRG of [TX13] even further using several other observations.

2 Preliminaries

We denote by $[n] = \{1, \ldots, n\}$. We denote by log and ln the logarithms in bases 2 and e, respectively. For $f : \{-1,1\} \to \mathbb{R}$ we denote by $\|f\|_p = \left(\mathbf{E}_{x \in \{-1,1\}^n}[|f(x)|^p]\right)^{1/p}$.

2.1 Restrictions

Definition 2.1 (Restriction)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. A* restriction *ρ is a vector of length n of elements from $\{0,1,*\}$. We denote by $f|_\rho : \{0,1\}^n \to \{0,1\}$ the function f restricted according to ρ, defined by*

$$f|_\rho(x) = f(y), \quad \text{where} \quad y_i = \begin{cases} x_i, & \rho_i = * \\ \rho_i, & \text{otherwise} \end{cases}.$$

We say that the variable x_i is fixed *if $\rho_i \in \{0,1\}$, and that x_i is* unassigned *(or* alive*) if $\rho_i = *$.*

Note that the function $f|_\rho$ is defined as a function with n variables, although it depends only on the non-fixed variables. When fixing only one bit to a constant, we may denote the restricted function by $f|_{x_i=b}$.

Definition 2.2 (p-Random Restriction)**.** *A p-*random restriction *is a restriction as in Definition 2.1 that is sampled in the following way. For every $i \in [n]$, independently, with probability p set $\rho_i = *$ and with probability $\frac{1-p}{2}$ set ρ_i to be -1 and 1, respectively. We denote this distribution of restrictions by $\mathcal{R}_p$.*

2.2 Fourier Analysis of Boolean Functions

Any function $f : \{-1,1\}^n \to \mathbb{R}$ has a unique Fourier representation:

$$f(x) = \sum_{S \subseteq [n]} \hat{f}(S) \cdot \prod_{i \in S} x_i \ ,$$

where the coefficients $\hat{f}(S) \in \mathbb{R}$ are given by $\hat{f}(S) = \mathbf{E}_x[f(x) \cdot \prod_{i \in S} x_i]$. Parseval's identity states that $\sum_S \hat{f}(S)^2 = \mathbf{E}_x[f(x)^2] = \|f\|_2^2$, and in the case that f is Boolean (i.e., $f : \{-1,1\}^n \to \{-1,1\}$), all are equal to 1. The Fourier representation is the unique multilinear polynomial which agrees with f on $\{-1,1\}^n$. We denoted by $\deg(f)$ the degree of this polynomial, which also equals $\max\{|S| : \hat{f}(S) \neq 0\}$. We denote by

$$\mathbf{W}^k[f] \triangleq \sum_{S \subseteq [n], |S|=k} \hat{f}(S)^2$$

the Fourier weight at level k of f. Similarly, we denote $\mathbf{W}^{\geq k}[f] \triangleq \sum_{S \subseteq [n], |S| \geq k} \hat{f}(S)^2$. The truncated Fourier expansion of degree k of f is simply $f^{\leq k}(x) = \sum_{|S| \leq k} \hat{f}(S) \prod_{i \in S} x_i$. By Parseval, $\|f - f^{\leq k}\|_2^2 = \mathbf{W}^{\geq k+1}[f]$. The following fact relates the Fourier coefficients of f and $f|_\rho$, where ρ is a p-random restriction.[4]

Fact 2.3 (Proposition 4.17, [O'D14])**.** *Let $f : \{-1,1\}^n \to \mathbb{R}$, $S \subseteq [n]$, and $p > 0$. Then,*

$$\mathop{\mathbf{E}}_{\rho \sim \mathcal{R}_p} \left[\widehat{f|_\rho}(S) \right] = \hat{f}(S) p^{|S|}$$

and

$$\mathop{\mathbf{E}}_{\rho \sim \mathcal{R}_p} \left[\widehat{f|_\rho}(S)^2 \right] = \sum_{U \subseteq [n]} \hat{f}(U)^2 \cdot \mathop{\mathbf{Pr}}_{\rho \sim \mathcal{R}_p} [\{i \in U : \rho(i) = *\} = S] \ .$$

Summing the last equation over all sets S of size d gives the following corollary.

Fact 2.4. *Denote by $\mathrm{Bin}(k,p)$ a binomial random variable with parameters k and p. Then,*

$$\mathop{\mathbf{E}}_{\rho \sim \mathcal{R}_p} \left[\mathbf{W}^d[f|_\rho] \right] = \sum_{k=d}^{n} \mathbf{W}^k[f] \cdot \mathbf{Pr}[\mathrm{Bin}(k,p) = d]$$

Definition 2.5 (Fourier Sparsity, Spectral Norm)**.** *We define the* sparsity *of $f : \{-1,1\}^n \to \mathbb{R}$ as $\mathrm{sparsity}(f) \triangleq |\{S : \hat{f}(S) \neq 0\}|$; the* spectral norm of f *as $L_1(f) \triangleq \sum_S |\hat{f}(S)|$; and the* spectral norm of the k-th level of f *as $L_{1,k}(f) \triangleq \sum_{S:|S|=k} |\hat{f}(S)|$.*

We state the following known fact regarding the Fourier sparsity, spectral norm and granularity of low degree Boolean functions.

Fact 2.6 (Ex. 1.11, [O'D14])**.** *Let $f : \{-1,1\}^n \to \{-1,1\}$ with $\deg(f) = d$. Then,*

1. *$\forall S : |\hat{f}(S)| = k_S \cdot 2^{-d}$ where $k_S \in \mathbb{Z}$.*
2. *$\mathrm{sparsity}(f) \leq 2^{2d}$*
3. *$L_1(f) \leq 2^d$.*

[4] Note that $\widehat{f|_\rho}(S) = 0$ if ρ fixes one of the variables in S.

3 Exponentially Small Fourier Tails for Bounded Depth Circuits

We generalize the proof of Håstad ([Hås14]), who showed that the correlation between the parity function and any Boolean circuit of depth d and size m is at most $2^{-\Omega(n/\log^{d-1}(m))}$. This bound is tight up to the constants in the exponent, as shown by an example in [Hås14], and improves upon previous bounds from [LMN93, Hås01].

We will use two simple lemmata which explain the behavior of Fourier tails with respect to random restrictions, and arbitrary restrictions.

Lemma 3.1 ([LMN93])**.** *For any $f : \{-1,1\}^n \to \mathbb{R}$, $k \in \mathbb{N} \cup \{0\}$ and $p \in [0,1]$,*

$$\mathbf{W}^{\geq k}[f] \leq 2 \cdot \mathop{\mathbf{E}}_{\rho \sim \mathcal{R}_p} \left[\mathbf{W}^{\geq \lfloor kp \rfloor}[f|_\rho]\right] .$$

Proof. Let $k \in \mathbb{N} \cup \{0\}$ and $p \in [0,1]$. We have

$$\begin{aligned}
\mathop{\mathbf{E}}_{\rho \sim \mathcal{R}_p} \left[\mathbf{W}^{\geq \lfloor kp \rfloor}[f|_\rho]\right] &= \sum_{\ell \geq \lfloor kp \rfloor} \mathbf{W}^{\ell}[f] \cdot \mathbf{Pr}[\mathrm{Bin}(\ell, p) \geq \lfloor kp \rfloor] && \text{(Fact 2.4)} \\
&\geq \sum_{\ell \geq k} \mathbf{W}^{\ell}[f] \cdot \mathbf{Pr}[\mathrm{Bin}(\ell, p) \geq \lfloor kp \rfloor] \\
&\geq \sum_{\ell \geq k} \mathbf{W}^{\ell}[f] \cdot 1/2 && (\mathrm{median}(\mathrm{Bin}(\ell,p)) \geq \lfloor \ell p \rfloor \geq \lfloor kp \rfloor, \text{ [KB80]}) \\
&= 1/2 \cdot \mathbf{W}^{\geq k}[f]. && \square
\end{aligned}$$

The second lemma, taken from [IK14], states that if, for some bit, we have Fourier tail bounds for both restrictions fixing that bit to either $+1$ or -1, then we have Fourier tail bounds for the unrestricted function.

Lemma 3.2 ([IK14])**.** *Let $f : \{-1,1\}^n \to \mathbb{R}$ and $i \in [n]$. Then,*

$$\mathbf{W}^{\geq k}[f] \leq \frac{1}{2} \cdot \mathbf{W}^{\geq k-1}[f|_{x_i=-1}] + \frac{1}{2} \cdot \mathbf{W}^{\geq k-1}[f|_{x_i=1}].$$

In order to generalize the last lemma, we introduce the following definition, which is very similar to the definition of a decision tree, except we are not making any decision!

Definition 3.3 (Restriction Tree)**.** *A* restriction tree *is a rooted directed binary tree such that each internal node is labeled by a variable from $x_1, \ldots, x_n$ and has two outgoing edges: one marked with 1 and one marked with -1. The leaves of the tree are not labeled. Each leaf in the tree, ℓ, corresponds to a restriction τ_ℓ on the variables $x_1, \ldots, x_n$ in the most natural way: we fix the variables along the path from the root to the leaf ℓ according to the values on the path edges.*

Using induction, Lemma 3.2 implies (informally) that if, for some restriction tree, we have Fourier tail bounds for restrictions corresponding to all root-leaf paths in the tree, then we have Fourier tail bounds for the unrestricted function as well. The exact statement follows.

Lemma 3.4. *Let $f : \{-1,1\}^n \to \mathbb{R}$ be a function, and let T be a restriction tree of depth $\leq D$ such that for any leaf ℓ, under the corresponding restriction $\mathbf{W}^{\geq k}[f|_{\tau_\ell}] \leq \varepsilon$. Then, $\mathbf{W}^{\geq k+D}[f] \leq \varepsilon$.*

Proof. Apply induction on the depth of the restriction tree. For depth 0 this obviously holds. For depth D, consider both subtrees that are rooted by the children of the original root. If the root queries x_i, these are restriction trees for $\{x : x_i = 1\}$ and $\{x : x_i = -1\}$, and we may apply the induction hypothesis on each subtree to get $\mathbf{W}^{\geq k+(D-1)}[f|_{x_i=1}] \leq \varepsilon$ and $\mathbf{W}^{\geq k+(D-1)}[f|_{x_i=-1}] \leq \varepsilon$. Finally, applying Lemma 3.2 gives $\mathbf{W}^{\geq k+D}[f] \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$. □

Our proof relies on the main lemma in Håstad's work [Hås14]. We begin with a definition from [Hås14] and the statement of his main lemma.

Definition 3.5 (Common Partial Decision Tree)**.** *A set of functions $(g_i)_{i=1}^m$ has a* common s-partial decision tree of depth D*, if there is a restriction tree of depth D such that at each leaf ℓ of this restriction tree, each function g_i, restricted by τ_ℓ, is computable by an ordinary decision tree of depth s.*

Lemma 3.6 ([Hås14], Lemma 3.8)**.** *Let $(f_i)_{i=1}^m$ be a collection of depth-2 circuits, each of bottom fan-in t. Let ρ be a random restriction from $\mathcal{R}_p$. Then the probability that $(f_i|_\rho)_{i=1}^m$ is not computable by a common $\log(2m)$-partial decision tree of depth D is at most $m \cdot (24pt)^D$.*

In Appendix 7 we give a new proof for Lemma 3.6 (with constant 49 instead of 24) following the proof approach of [Raz95], [Bea94] and [Tha09] for the original switching lemma.

We are ready to prove the Fourier tail bounds for Boolean circuits. We define the effective size of a Boolean circuit as the number of gates in the circuit at distance 2 or more from the inputs.

Theorem 3.7. *Let f be a Boolean circuit of depth d, effective size m, and bottom fan-in t. Then, $\mathbf{W}^{\geq k}[f] \leq 8^{d-1} \cdot 2^{-k/\left(20t(96\log(2m))^{d-2}\right)}$.*

Proof. We prove by induction on d. The base case $d = 2$ was proved by Mansour [Man95], who showed that **DNFs** with bottom fan-in t have

$$\mathbf{W}^{\geq k}[f] \leq 4 \cdot 2^{-k/20t} .$$

For the induction step, we apply a p-random restriction with $p = 1/48t$. Consider the gates at distance 2 from the inputs: $f_1, \ldots, f_{m'}$, for $m' \leq m$. These gates compute functions given by depth-2 circuits with bottom fan-in $\leq t$. Setting $D = \lfloor kp/2 \rfloor$ and using Lemma 3.6 gives that with probability at least $1 - m \cdot 2^{-D} \geq 1 - 2^{\log(m)-D}$ over the random restrictions, $(f_i|_\rho)_{i=1}^{m'}$ can be computed by a common $\log(2m)$-partial decision tree of depth D. In this case, we say that the restriction ρ is *good*. Using Lemma 3.1 we have $\mathbf{W}^{\geq k}[f] \leq 2 \cdot \mathbf{E}_\rho[\mathbf{W}^{\geq \lfloor kp \rfloor}[f|_\rho]]$.

Since $\mathbf{W}^{\geq \lfloor kp \rfloor}[f|_\rho]$ is a random variable bounded in $[0,1]$ we have

$$\begin{aligned}\mathbf{W}^{\geq k}[f] &\leq 2 \cdot \mathop{\mathbf{E}}_{\rho \sim \mathcal{R}_p}\left[\mathbf{W}^{\geq \lfloor kp \rfloor}[f|_\rho]\right] \\ &= 2 \cdot \mathop{\mathbf{E}}_{\rho \sim \mathcal{R}_p}\left[\mathbf{W}^{\geq \lfloor kp \rfloor}[f|_\rho] \mid \rho \text{ is good}\right] \cdot \mathop{\mathbf{Pr}}_{\rho \sim \mathcal{R}_p}\left[\rho \text{ is good}\right] \\ &\quad + 2 \cdot \mathop{\mathbf{E}}_{\rho \sim \mathcal{R}_p}\left[\mathbf{W}^{\geq \lfloor kp \rfloor}[f|_\rho] \mid \rho \text{ is bad}\right] \cdot \mathop{\mathbf{Pr}}_{\rho \sim \mathcal{R}_p}\left[\rho \text{ is bad}\right] \\ &\leq 2 \cdot \mathop{\mathbf{E}}_{\rho \sim \mathcal{R}_p}\left[\mathbf{W}^{\geq \lfloor kp \rfloor}[f|_\rho] \mid \rho \text{ is good}\right] + 2 \cdot \mathop{\mathbf{Pr}}_{\rho \sim \mathcal{R}_p}\left[\rho \text{ is bad}\right],\end{aligned}$$

where $\mathbf{Pr}_\rho[\rho \text{ is bad}] \leq 2^{\log(m) - \lfloor k/96t \rfloor} \leq 2^{\log(2m) - k/96t}$. Using the following simple claim, we get $\mathbf{Pr}_\rho[\rho \text{ is bad}] \leq 2 \cdot 2^{-k/(96t \log(2m))}$.

Claim 3.8. *If $0 \leq X \leq 1$ and $X \leq 2^{a-b}$, where $a \geq 1$, then $X \leq 2^{1-b/a}$.*

Proof. Since $0 \leq X \leq 1$ and $a \geq 1$, we have $X \leq X^{1/a}$, and $X^{1/a}$ is at most $2^{1-b/a}$. □

We are left to analyze $\mathbf{E}[\mathbf{W}^{\geq \lfloor kp \rfloor}[f|_\rho] \mid \rho \text{ is good}]$. Fixing ρ to be some specific good restriction, we will bound $\mathbf{W}^{\geq \lfloor kp \rfloor}[f|_\rho]$ for this specific ρ. By the definition of good restrictions, we have a common $\log(2m)$-partial decision tree of depth $D = \lfloor kp/2 \rfloor$ computing $(f_i|_\rho)_{i=1}^{m'}$. For each leaf ℓ of the common partial decision tree, let τ_ℓ be the restriction defined by the path leading to this leaf. We have that $f_i|_\rho|_{\tau_\ell}$ for $i = 1, \ldots, m'$ can be expressed as a decision tree of depth $\leq \log(2m)$, hence as a **CNF**/**DNF** formula of bottom fan-in at most $\log(2m)$. This means that applying the restriction $\rho \circ \tau_\ell$, the circuit f collapses to a depth $d-1$ Boolean circuit with bottom fan-in $t' \leq \log(2m)$ and effective size at most m.[5] By the induction hypothesis, for any k' we have $\mathbf{W}^{\geq k'}[f|_\rho|_{\tau_\ell}] \leq 8^{d-2} \cdot 2^{-\Omega(k'/(t' \log^{d-3}(2m)))}$. Setting $k' = \lfloor kp \rfloor - D \geq \lfloor kp/2 \rfloor \geq \frac{k}{96t} - 1$ and applying Lemma 3.4 we have

$$\mathbf{W}^{\geq \lfloor kp \rfloor}[f|_\rho] \leq \max_\ell \mathbf{W}^{\geq k'}[f|_\rho|_{\tau_\ell}] \leq 8^{d-2} \cdot 2^{-k'/(20t' \cdot (96 \log(2m))^{d-3})} \leq 8^{d-2} \cdot 2^{1 - k/\left(20t(96\log(2m))^{d-2}\right)},$$

and

$$\mathbf{W}^{\geq k}[f] \leq 4 \cdot 8^{d-2} \cdot 2^{-k/\left(20t(96\log(2m))^{d-2}\right)} + 4 \cdot 2^{-k/(96t\log(2m))} \leq 8^{d-1} \cdot 2^{-k/\left(20t(96\log(2m))^{d-2}\right)}. \quad \square$$

Theorem 3.9 (Theorem 1.1, restated)**.** *Let f be an Boolean circuit of depth d and size $m > 1$. Then, $\mathbf{W}^{\geq k}[f] \leq 2 \cdot 2^{-k/(c_d \log^{d-1}(m))}$ where $c_d = 60d \cdot 192^{d-1} \leq 216^d$. Equivalently, $\mathbf{W}^{\geq k}[f] \leq 2 \cdot e^{-k/(c'_d \log^{d-1}(m))}$ where $c'_d = \log_2(e) \cdot 60d \cdot 192^{d-1} \leq 2 \cdot 216^d$.*

Proof. Let f be a function computed by a Boolean circuit of depth d and m gates. We add a dummy layer of fan-in 1 gates in between the inputs and the layer next to them. Thus, f is realized by an Boolean circuit of depth $d+1$, effective size m and bottom fan-in 1. Plugging this into Theorem 3.7 gives $\mathbf{W}^{\geq k}[f] \leq 2^{3d - k/(20 \cdot 96^{d-1} \cdot \log^{d-1}(2m))}$. Hence, by Claim 3.8, we get

$$\mathbf{W}^{\geq k}[f] \leq 2 \cdot 2^{-k/(3d \cdot 20 \cdot 96^{d-1} \cdot \log^{d-1}(2m))} \leq 2 \cdot 2^{-k/(60d \cdot 96^{d-1} \cdot 2^{d-1} \cdot \log^{d-1}(m))},$$

where we used $\log(2m) \leq 2\log(m)$ for $m > 1$. □

[5] We only introduce new gates with distance 1 from the inputs - which does not increase the effective size.

4 Applications to Pseudorandomness

4.1 Improving Braverman's Analysis

Definition 4.1. *Denote by* $\mathrm{tail}(m,d,k)$ *the maximal* $\mathbf{W}^{\geq k}[F]$ *over all Boolean circuits* F *of size* $\leq m$ *and depth* $\leq d$.

By Theorem 3.9, $\mathrm{tail}(m,d,k) \leq 2 \cdot 2^{-k/(c_d \log^{d-1}(m))}$. Braverman's Theorem can be rephrased as follows (we show that this is indeed the case in Appendix B).

Theorem 4.2 ([Bra10])**.** *Let* $s_1, s_2 \geq \log m$ *be any parameters. Let* F *be a Boolean function computed by a circuit of depth* d *and size* m*. Let* μ *be an* r*-independent distribution where*

$$r = r(s_1, s_2, d) = 2((s_1 \cdot \log m)^d + s_2)$$

then

$$|\mathop{\mathbf{E}}_{\mu}[F] - E[F]| < \varepsilon(s_1, s_2, d),$$

where $\varepsilon(s_1, s_2, d) = 0.82^{s_1} \cdot (6m) + 2^{4(s_1 \cdot \log m)^d \log m} \cdot \mathrm{tail}(m^3, d+3, s_2)$

Picking $s_1 := 5\log(12m/\varepsilon)$ and $s_2 := \left(c_{d+3}\log(m^3)^{d+2}\right) \cdot 8 \cdot (s_1 \cdot \log m)^d \cdot \log(m)$ we get the following corollary.

Theorem 4.3. $r(m,d,\varepsilon)$*-independence* ε*-fools Boolean circuits of depth* d *and size* m*, where*

$$\begin{aligned} r(m,d,\varepsilon) &= 2((s_1 \cdot \log m)^d + s_2) \leq 4s_2 \\ &= 32 \cdot c_{d+3} \cdot (5\log(12m/\varepsilon))^d \cdot 3^{d+2} \cdot (\log m)^{2d+3} \\ &\leq O(\log(m/\varepsilon))^d \cdot (\log m)^{2d+3} \ . \end{aligned}$$

4.2 Improving Bazzi's Analysis

Bazzi [Baz09] showed that $O(\log^2(m/\varepsilon))$ independence ε-fools **DNFs** of size m. We show that $O(\log(m/\varepsilon) \cdot \log(m))$ independence suffices. For $\varepsilon \leq 1/m^{\Omega(1)}$ this bound is tight, due to the example of Mansour from [LV96].

Theorem 4.4 ([Baz09], [Raz09])**.** *Let* F *be a* **DNF** *with* m *terms, and* t *be some parameter. Then,* F *is* $m^3 \cdot \mathrm{tail}(m, 2, (k-3t)/2) + m2^{-t}$ *fooled by any* k*-wise independence.*

Picking $t := \log(2m/\varepsilon)$ and $k := 3t + 2c_2 \log(m)\log(4m^3/\varepsilon) = O(\log(m)\log(m/\varepsilon))$, we get that k-wise independence ε-fools **DNFs** with m terms since

$$m^3 \cdot \mathrm{tail}(2, m, (k-3t)/2) + m2^{-t} \leq m^3 \cdot 2 \cdot 2^{\frac{-c_2 \log(m)\log(4m^3/\varepsilon)}{c_2 \log(m)}} + \frac{\varepsilon}{2} \leq \varepsilon \ .$$

5 On Fourier Concentration, Switching Lemmas and Influence Moments

In this section, we connect different notions of Fourier concentration of Boolean functions. We begin by introducing some new definitions, and then move to state and prove the connections between the different notions. We end this Section, with the proof of Theorem 1.2, which is a result of Theorem 1.1 and the connections established in this section.

5.1 Influence Moments

In this section we introduce derivatives and influences of sets of variables. A different definition to the influence of a set was made in [KKL88]. There, the influence of a set J was defined to be the probability that under a uniform restriction of J^c to constants, the function's value is still undetermined. We choose a different variant, which has a much nicer Fourier expression.

We start with the standard definition of discrete derivatives and influences of Boolean functions.

Definition 5.1 (Discrete Derivative, Influence)**.** *Let $f\colon \{-1,1\}^n \to \mathbb{R}$ and $i \in [n]$. The* i-th discrete derivative operator *D_i maps the function f to the function $D_i f : \{-1,1\}^n \to \mathbb{R}$ defined by*

$$D_i f(x) = \frac{f(x^{(i\mapsto 1)}) - f(x^{(i\mapsto -1)})}{2} .$$

where $x^{(i\mapsto b)} = (x_1, \ldots, x_{i-1}, b, x_{i+1}, \ldots, x_n)$. The influence of coordinate i on f *is defined as*

$$\mathrm{Inf}_i(f) = \mathop{\mathbf{E}}_x[(D_i f(x))^2] .$$

The generalization to sets of more than one variable is the following.

Definition 5.2 (Discrete Derivative and Influence of a Set)**.** *Let $f\colon \{-1,1\}^n \to \mathbb{R}$ and $T \subseteq [n]$, and write $T = \{j_1, \ldots, j_k\}$. The T-*th (discrete) derivative operator*, D_T, maps the function f to the function $D_T f : \{-1,1\}^n \to \mathbb{R}$ defined by*

$$D_T f(x) = D_{j_1} D_{j_2} \ldots D_{j_k} f(x) .$$

The influence of subset T on f *is defined as*

$$\mathrm{Inf}_T(f) = \mathop{\mathbf{E}}_x\left[(D_T f(x))^2\right] .$$

The following claim gives equivalent formulations for the function $D_T f$ (and also implies that D_T is well defined, i.e., that $D_T f$ does not depend on the order of indices in T).

Claim 5.3.

$$D_T f(x) = \frac{1}{2^{|T|}} \sum_{z \in \{-1,1\}^T} f(x^{(T\mapsto z)}) \cdot \prod_{i \in T} z_i = \sum_{S \supseteq T} \hat{f}(S) \cdot \prod_{i \in S \setminus T} x_i$$

where $x^{(T\mapsto z)}$ is the vector in $\{-1,1\}^n$ whose i-th coordinate equals z_i whenever $i \in T$, and equals x_i otherwise.

The proof uses a straightforward inductive argument, and is given for completeness in Appendix A. Note that if $f : \{-1,1\}^n \to \{-1,1\}$, then the T-th derivative of f is $2^{-|T|}$ granular (i.e., $D_T f(x)$ is an integer times $2^{-|T|}$), since $D_T f(x)$ is a sum of integers divided by $2^{|T|}$. The following claim follows from Parseval's identity and the previous claim.

Claim 5.4. $\mathrm{Inf}_T(f) = \sum_{S \supseteq T} \hat{f}(S)^2$

Definition 5.5 (Total Degree-k Influence)**.** *The* total degree-k influence *is defined as*

$$\mathrm{Inf}^k(f) \triangleq \sum_{T:|T|=k} \mathrm{Inf}_T(f) \ .$$

Claim 5.4 gives the following Fourier expression for the total degree-k influence:

$$\mathrm{Inf}^k(f) = \sum_{S:|S|\geq k} \hat{f}(S)^2 \cdot \binom{|S|}{k} = \sum_{d\geq k} \mathbf{W}^d[f] \cdot \binom{d}{k} \ . \tag{2}$$

We state the following simple lemma expressing $\mathrm{Inf}^k(f)$ in terms of $\mathbf{W}^{\geq d}[f]$ instead of $\mathbf{W}^d[f]$.

Lemma 5.6. $\mathrm{Inf}^k(f) = \sum_{d\geq k} \mathbf{W}^{\geq d}[f] \cdot \binom{d-1}{k-1}$ *for all* $k \in \mathbb{N}$.

Proof. We perform some algebraic manipulations on Equation (2):

$$\begin{aligned}
\mathrm{Inf}^k(f) = \sum_{d\geq k} \mathbf{W}^d[f] \cdot \binom{d}{k} &= \sum_{d\geq k} \left(\mathbf{W}^{\geq d}[f] - \mathbf{W}^{\geq d+1}[f]\right) \cdot \binom{d}{k} \\
&= \mathbf{W}^{\geq k}[f] + \sum_{d\geq k+1} \mathbf{W}^{\geq d}[f] \cdot \left(\binom{d}{k} - \binom{d-1}{k}\right) \\
&= \mathbf{W}^{\geq k}[f] + \sum_{d\geq k+1} \mathbf{W}^{\geq d}[f] \cdot \binom{d-1}{k-1} \\
&= \sum_{d\geq k} \mathbf{W}^{\geq d}[f] \cdot \binom{d-1}{k-1}
\end{aligned}$$

□

5.2 Connections between Four Fourier Concentration Properties

In this section we show connections between four attributes of Boolean functions, and establish equivalence between three of them. The properties, each relative to a parameter t, are the following:

- **ESFT**: Exponentially small Fourier tails.

$$\forall k : \mathbf{W}^{\geq k}[f] \leq e^{-\Omega(k/t)}$$

- **SLTP**: Switching lemma type property / degree shrinkage

$$\forall d, p : \Pr_{\rho\sim\mathcal{R}_p}[\deg(f|_\rho) = d] \leq O(pt)^d$$

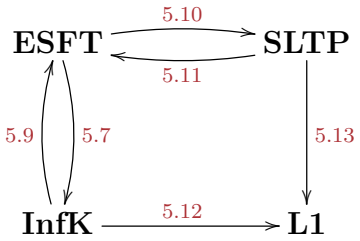
- **L1**: Bounded spectral norm of the k-th level.

$$\forall k : \sum_{|S|=k} |\hat{f}(S)| \leq O(t)^k$$

- **InfK**: Bounded total degree-k influence.

$$\forall k : \mathrm{Inf}^k[f] \leq O(t)^k \ .$$

In Lemmata 5.7, 5.9, 5.10, 5.11, 5.12, 5.13, we show the following connections:



We remark that Lemma 5.13 is due to Mansour [Man95], and Lemma 5.11 is due to Linial et al. [LMN93]. Note that **L1** does not imply any other property, because one can take for example the parity function, which has the **L1** property with $t = 1$. However, this function has very large Fourier tails, very high degree under random restriction, and $\binom{n}{k}$ total degree-k influence. Anything that implies **SLTP** and **L1** needs f to be Boolean. Other relations generalize to bounded real-valued functions.

In the remainder of this section we state Lemmata 5.7, 5.9, 5.10, 5.11, 5.12, 5.13 more accurately and prove them.

Lemma 5.7. *Let $t > 0, C > 0$. If $\mathbf{W}^{\geq d}[f] \leq C \cdot e^{-d/t}$ for all d, then $\mathrm{Inf}^k[f] \leq C \cdot t^k$ for all k.*

In the proof of Lemma 5.7, we use the following simple fact that follows from Newton's generalized binomial theorem.

Fact 5.8. *Let $|x| < 1$, and $k \in \mathbb{N}$. Then, $\sum_{d=k}^{\infty} \binom{d-1}{k-1} \cdot x^d = \frac{x^k}{(1-x)^k}$.*

Proof of Lemma 5.7. We shall prove for $C = 1$, the proof generalizes for all C. Denote $a := e^{-1/t}$. Using Lemma 5.6 we bound the total degree-k influence:

$$\mathrm{Inf}^k(f) = \sum_{d \geq k} \mathbf{W}^{\geq d}[f] \cdot \binom{d-1}{k-1} \leq \sum_{d \geq k} e^{-d/t} \cdot \binom{d-1}{k-1} = \sum_{d \geq k} a^d \cdot \binom{d-1}{k-1}$$

Using Fact 5.8 with $x := a$ gives

$$\mathrm{Inf}^k(f) \leq \frac{a^k}{(1-a)^k} = \frac{1}{(1/a-1)^k} = \frac{1}{\left(e^{1/t}-1\right)^k} \leq \frac{1}{(1/t)^k} = t^k$$

where in the last inequality we used the fact that $e^x - 1 \geq x$ for all $x \in \mathbb{R}$. □

The reverse relation holds too, i.e. **InfK** implies **ESFT**.

Lemma 5.9. *Let $t > 0, C > 0$. If $\mathrm{Inf}^k[f] \leq C \cdot t^k$ for all k, then $\mathbf{W}^{\geq d}[f] \leq C \cdot e \cdot t \cdot e^{-(d-1)/et}$ for all d.*

Proof. We shall prove for $C = 1$, the proof generalizes for all C. By Lemma 5.6, $\mathbf{W}^{\geq d}[f] \cdot \binom{d-1}{k-1} \leq \mathrm{Inf}^k[f] \leq t^k$. Hence $\mathbf{W}^{\geq d}[f] \leq t^k / \binom{d-1}{k-1}$. We can pick any k to optimize this bound. Picking $k = \lfloor (d-1)/et \rfloor + 1$ we get

$$\mathbf{W}^{\geq d}[f] \leq t^k / \left(\frac{d-1}{k-1}\right)^{k-1} \leq t \cdot e^{-(k-1)} \leq e \cdot t \cdot e^{-(d-1)/et} . \qquad \square$$

In our previous work [Tal14], the following relation (**ESFT** implies **SLTP**) was established.

Lemma 5.10 ([Tal14])**.** *Let $t, C > 0$, and $f : \{-1,1\}^n \to \{-1,1\}$. If $\mathbf{W}^{\geq k}[f] \leq C \cdot e^{-k/t}$ for all k, then $\mathbf{Pr}_{\rho \sim \mathcal{R}_p}[\deg(f|_\rho) = d] \leq C \cdot (4pt)^d$ for all p, d.*

We give a slightly shorter proof, using the total degree-d influence.

Proof. We shall prove for $C = 1$, the proof generalizes for all C. The proof goes by showing that

$$\mathop{\mathbf{E}}_{\rho}[\mathbf{W}^d[f|_\rho]] \leq (pt)^d \tag{3}$$

and

$$\mathop{\mathbf{E}}_{\rho}[\mathbf{W}^d[f|_\rho]] \geq 4^{-d} \cdot \mathop{\mathbf{Pr}}_{\rho}[\deg(f|_\rho) = d] . \tag{4}$$

Equation (4) is true since

$$\mathop{\mathbf{E}}_{\rho}[\mathbf{W}^d[f|_\rho]] \geq \mathop{\mathbf{E}}_{\rho}[\mathbf{W}^d[f|_\rho] | \deg(f|_\rho) = d] \cdot \mathop{\mathbf{Pr}}_{\rho}[\deg(f|_\rho) = d] .$$

and the (random) Boolean function $f|_\rho$ has Fourier mass at least 4^{-d} if $\deg(f|_\rho) = d$, by the granularity of low degree functions - Fact 2.6.

We are left to prove Equation (3). Using Fact 2.4, we have

$$\mathop{\mathbf{E}}_{\rho}[\mathbf{W}^d[f|_\rho]] = \sum_{k=d}^{n} \mathbf{W}^k[f] \binom{k}{d} p^d (1-p)^{k-d} \leq p^d \sum_{k=d}^{n} \mathbf{W}^k[f] \binom{k}{d} = p^d \cdot \mathrm{Inf}^d[f] \leq (pt)^d ,$$

where in the last inequality we used Lemma 5.7. $\square$

Linial, Mansour and Nisan [LMN93] proved that **SLTP** implies **ESFT**.

Lemma 5.11 ([LMN93], restated slightly)**.** *Let $t > 0, C > 0$, and $f : \{-1,1\}^n \to [-1,1]$. If for all $d \in \mathbb{N}, p \in (0,1)$, $\mathbf{Pr}_{\rho \sim \mathcal{R}_p}[\deg(f|_\rho) \geq d] \leq C\,(tp)^d$, then for any k, $\mathbf{W}^{\geq k}[f] \leq 2e \cdot C \cdot e^{-k/te}$.*

The proof is given in [LMN93]; we give it here for completeness.

Proof. Pick $p = 1/et$, then by Lemma 3.1, and the fact that $\mathbf{W}^{\geq \lfloor kp \rfloor}[f|_\rho]$ is always at most 1 and equals 0 whenever $\deg(f|_\rho) < \lfloor kp \rfloor$, we get

$$\mathbf{W}^{\geq k}[f] \leq 2 \mathop{\mathbf{E}}_{\rho}\left[\mathbf{W}^{\geq \lfloor kp \rfloor}[f|_\rho]\right] \leq 2 \mathop{\mathbf{E}}_{\rho}\left[\mathbf{Pr}[\deg(f|_\rho) \geq \lfloor kp \rfloor]\right] \leq 2C(1/e)^{\lfloor k/et \rfloor} . \qquad \square$$

The next lemma proves that **InfK** implies **L1**.

Lemma 5.12. *If f is Boolean, then $L_{1,k}[f] \le 2^k \cdot \mathrm{Inf}^k[f]$.*

Proof. It is easy to see from Claim 5.3 that for any subset $T \subseteq [n]$,

$$\mathop{\mathbf{E}}_{x}[D_T f(x)] = \mathop{\mathbf{E}}_{x}\left[\sum_{S \supseteq T} \hat{f}(S) \prod_{i \in S \setminus T} x_i\right] = \hat{f}(T) \ .$$

Recall that if f is Boolean, then $D_T f(x)$ is $2^{-|T|}$ granular, which implies that $\forall x : |D_T f(x)| \le 2^{|T|}(D_T f(x))^2$. Hence,

$$|\hat{f}(T)| = |\mathop{\mathbf{E}}_{x}[D_T f(x)]| \le \mathop{\mathbf{E}}_{x}[|D_T f(x)|] \le 2^{|T|} \mathop{\mathbf{E}}_{x}[(D_T f(x))^2] = 2^{|T|}\mathrm{Inf}_T(f) \ .$$

Summing over all sets T of size k completes the proof. □

Remark: It is necessary that f is Boolean in Lemma 5.12, since otherwise we can have the function

$$f_{t,k}(x) = \sum_{S \subseteq [n], |S|=k} \frac{1}{\sqrt{\binom{n}{k}} e^{k/2t}} \prod_{i \in S} x_i$$

which maps $\{-1,1\}^n$ to $\mathbb{R}$, has $\mathbf{W}^{\ge k}[f_{t,k}] = \mathbf{W}^k[f_{t,k}] = e^{-k/t}$, and $\mathrm{Inf}^k[f_t] \le t^k$, but

$$L_{1,k}[f_t] = \sqrt{\binom{n}{k}} e^{-k/2t} \ge \left(\frac{n}{k e^{1/t}}\right)^{k/2}$$

is much larger than $O(t)^k$ for $n = \omega(kt^2 e^{1/t})$.

Next, Mansour [Man95] proved that **SLTP** implies **L1**.

Lemma 5.13 ([Man95])**.** *Let $t > 0$, and $f : \{-1,1\}^n \to \{-1,1\}$. If for all d, p, $\mathbf{Pr}_{\rho \sim \mathcal{R}_p}[\deg(f|_\rho) = d] \le C(pt)^d$, then $\forall k : L_{1,k}[f] \le 2C(4t)^k$.*

Proof. We shall prove for $C = 1$, the proof generalizes for all C. We first prove that for any function $f : \{-1,1\}^n \to \mathbb{R}$, $p \in [0,1], k \in \mathbb{N}$ we have $L_{1,k}(f) \le \frac{1}{p^k} \mathbf{E}_{\rho \sim \mathcal{R}_p}[L_{1,k}[f|_\rho]]$.

$$\begin{aligned} L_{1,k}[f] &= \sum_{S:|S|=k} |\hat{f}(S)| = \sum_{S:|S|=k} \left| \frac{1}{p^k} \mathop{\mathbf{E}}_{\rho \sim \mathcal{R}_p}\left[\widehat{f|_\rho}(S)\right]\right| && \text{(Fact 2.3)} \\ &\le \sum_{S:|S|=k} \frac{1}{p^k} \mathop{\mathbf{E}}_{\rho \sim \mathcal{R}_p}\left[|\widehat{f|_\rho}(S)|\right] = \frac{1}{p^k} \mathop{\mathbf{E}}_{\rho \sim \mathcal{R}_p}\left[\sum_{S:|S|=k} |\widehat{f|_\rho}(S)|\right] \\ &= \frac{1}{p^k} \mathop{\mathbf{E}}_{\rho \sim \mathcal{R}_p}[L_{1,k}[f|_\rho]] \ . && (5) \end{aligned}$$

Next, we show that for $f : \{-1,1\}^n \to \{-1,1\}$, if there exists $t > 0$ such that for all d, p, $\mathbf{Pr}[\deg(f|_\rho) = d] \le (pt)^d$, then $\mathbf{E}_{\rho \sim \mathcal{R}_p}[L_1[f|_\rho]] \le 2$ for $p = 1/4t$. Conditioning on $\deg(f|_\rho) = d$ and using Fact 2.6, we have $L_1[f|_\rho] \le 2^d$. Hence,

$$\mathop{\mathbf{E}}_{\rho \sim \mathcal{R}_p}[L_1[f|_\rho]] = \sum_{d=0}^{n} \mathop{\mathbf{E}}_{\rho \sim \mathcal{R}_p}[L_1[f|_\rho] | \deg(f|_\rho) = d] \cdot \mathbf{Pr}[\deg(f|_\rho) = d] \le \sum_{d=0}^{n} 2^d \cdot \left(\tfrac{1}{4}\right)^d \le 2 \ . \quad (6)$$

Plugging Equation (6) in Equation (5) with $p = 1/4t$ we get

$$L_{1,k}[f] \leq \frac{1}{p^k} \mathop{\mathbf{E}}_{\rho \sim \mathcal{R}_p} [L_{1,k}[f|_\rho]] \leq \frac{1}{p^k} \mathop{\mathbf{E}}_{\rho \sim \mathcal{R}_p} [L_1[f|_\rho]] \leq (4t)^k \cdot 2 \,. \qquad \square$$

The next lemma is relevant to the learnability results given in [Man95] and [LMN93].

Lemma 5.14. *Let f be a Boolean function, let $t \geq 1$ and C be some positive constant. If $\mathbf{W}^{\geq k}[f] \leq C \cdot e^{-k/t}$ for all k, then f is ε-concentrated on at most $t^{O(t \log(1/\varepsilon))}$ Fourier coefficients.*

Here, by ε-concentrated on r coefficients we mean that there exist r subsets of $[n]$, $\{S_1, \ldots, S_r\}$, which captures $1 - \varepsilon$ of the Fourier mass of f, i.e. $\sum_{i=1}^{r} \hat{f}(S_i)^2 \geq 1 - \varepsilon$.

Proof. We shall prove for $C = 1$, the proof generalizes for all constant C. Let $w := t \cdot \ln(2/\varepsilon)$. First it is enough to consider Fourier coefficients of sets of size $\leq w$, since the sum of squares of Fourier coefficients of larger sets is at most $\varepsilon/2$. Now $\sum_{S:|S|\leq w} |\hat{f}(S)| = \sum_{i=0}^{w} L_{1,i}[f]$. Using Lemmata 5.7 and 5.12 we get

$$\sum_{i=0}^{w} L_{1,i}[f] \leq \sum_{i=0}^{w} 2^i t^i \underset{(t \geq 1)}{\leq} t^w 2^{w+1} \,.$$

Letting $\mathcal{F} = \{S : |S| \leq w, |\hat{f}(S)| \geq \frac{\varepsilon/2}{t^w 2^{w+1}}\}$ we get by Parseval's identity that

$$\sum_{S \in \mathcal{F}} \hat{f}(S)^2 = 1 - \sum_{|S|>w} \hat{f}(S)^2 - \sum_{|S|\leq w, S \notin \mathcal{F}} \hat{f}(S)^2 \,,$$

where we already noted that $\sum_{|S|>w} \hat{f}(S)^2 \leq \varepsilon/2$. To bound the last term

$$\sum_{|S|\leq w, S \notin \mathcal{F}} \hat{f}(S)^2 \leq \max\{|\hat{f}(S)| : |S| \leq w, S \notin \mathcal{F}\} \cdot \sum_{|S|\leq w} |\hat{f}(S)| \leq \varepsilon/2 \,.$$

Hence, $\sum_{S \in \mathcal{F}} \hat{f}(S)^2 \geq 1 - \varepsilon$. It remain to figure out the size of $\mathcal{F}$. Since every coefficient in $\mathcal{F}$ contributes at least $\frac{\varepsilon/2}{t^w 2^{w+1}}$ to the sum $\sum_{i=0}^{w} L_{1,i}[k]$, and this sum is at most $t^w 2^{w+1}$ we get that the size of $\mathcal{F}$ is at most $2(t^w 2^{w+1})^2/\varepsilon = O(t)^{2t \ln(1/\varepsilon)}$, which completes the proof. $\square$

5.3 Theorem 1.2

Immediate from Theorem 3.9, Lemmata 5.7, 5.10, 5.12, and 5.14 we get the following corollary.

Theorem 5.15 (Thm. 1.2, restated)**.** *Let f be a Boolean circuit of depth d and size $m > 1$. Then,*

1. *For all k, p, $\mathbf{Pr}_{\rho \sim \mathcal{R}_p}[\deg(f|_\rho) = k] \leq 2 \cdot (4p \cdot c'_d \log^{d-1}(m))^k$.*
2. *For all k, $\mathrm{Inf}^k[f] \leq 2 \cdot (c'_d \log^{d-1}(m))^k$.*
3. *For all k, $L_{1,k}[f] = \sum_{S:|S|=k} |\hat{f}(S)| \leq 2 \cdot (2c'_d \log^{d-1}(m))^k$.*
4. *f is ε-concentrated on at most $O(\log^{d-1} m)^{O(\log^{d-1}(m) \log(1/\varepsilon))} = 2^{O(\log\log(m) \log^{d-1}(m) \log(1/\varepsilon))}$ Fourier coefficients.*

6 Short Proofs for Known Results

In this section, we give simple proofs for two known results based on Theorem 5.15.

6.1 Bounded-Depth Circuits Cannot Approximate Majority

The next result states that nearly balanced symmetric functions, and in particular the Majority function, cannot be well approximated by a small and shallow circuit.

Theorem 6.1. *Let $g : \{-1,1\}^n \to \{-1,1\}$ be a symmetric function on n variables. Let $f : \{-1,1\}^n \to \{-1,1\}$ be depth d size m circuit, and assume that*

$$c'_d \log^{d-1}(m) \leq (n/100\ln(n))^{1/3} \ .$$

Then,

$$\mathrm{Cor}(f,g) \triangleq \left|\mathop{\mathbf{E}}_x[f(x)g(x)]\right| \leq |\hat{g}(\emptyset)| + \frac{\sqrt{2}+8c'_d\log^{d-1}(m)}{\sqrt{n}}$$

Proof. Since g is a symmetric Boolean function, for all $S \subseteq [n]$, $\hat{g}(S)^2 \cdot \binom{n}{|S|} = \sum_{T:|T|=|S|} \hat{g}(T)^2 \leq 1$. Hence, $|\hat{g}(S)| \leq \frac{1}{\sqrt{\binom{n}{|S|}}}$. Let ℓ be some parameter we shall set later. Then,

$$\left|\mathop{\mathbf{E}}_x[f(x)g(x)]\right| \leq \sum_S |\hat{f}(S)\hat{g}(S)| = |\hat{f}(\emptyset)\hat{g}(\emptyset)| + \sum_{k=1}^{\ell}\sum_{S:|S|=k}|\hat{f}(S)\hat{g}(S)| + \sum_{S:|S|>\ell}|\hat{f}(S)\hat{g}(S)|. \tag{7}$$

We bound each of the three terms in the RHS of Equation (7). The first term is at most $|\hat{g}(\emptyset)|$. For the third term we use Cauchy-Schwartz, Theorem 3.9, and Parseval's identity ($\sum_{S:|S|>\ell}\hat{g}(S)^2 \leq 1$), to get

$$\sum_{S:|S|>\ell}|\hat{f}(S)\hat{g}(S)| \leq \sqrt{\sum_{S:|S|>\ell}\hat{f}(S)^2 \sum_{S:|S|>\ell}\hat{g}(S)^2} \leq \sqrt{2\cdot e^{-\ell/\left(c'_d\log^{d-1}(m)\right)}} \ .$$

Picking $\ell := \ln(n)\cdot c'_d\log^{d-1}(m)$ this is smaller than $\sqrt{2/n}$. For the second term in the RHS of Equation (7), we use the estimates on $L_{1,k}(f)$ and $|\hat{g}(S)|$, to get

$$\sum_{S:|S|=k}|\hat{g}(S)\hat{f}(S)| \leq \frac{1}{\sqrt{\binom{n}{k}}}\cdot\sum_{S:|S|=k}|\hat{f}(S)| \leq \frac{2\cdot(2c'_d\log^{d-1}(m))^k}{\sqrt{\binom{n}{k}}} \leq 2\cdot\left(\frac{2c'_d\log^{d-1}(m)}{\sqrt{n/k}}\right)^k \ .$$

We denote by $D_k := 2\cdot\left(\frac{2c'_d\log^{d-1}(m)}{\sqrt{n/k}}\right)^k$. The ratio between two consecutive terms D_{k+1}/D_k for $k+1 \leq \ell$ is at most

$$\frac{2c'_d\log^{d-1}(m)}{\sqrt{n}}\sqrt{\frac{(k+1)^{k+1}}{k^k}} \leq \frac{2c'_d\log^{d-1}(m)}{\sqrt{n}}\sqrt{e\cdot(k+1)} \leq \frac{2c'_d\log^{d-1}(m)}{\sqrt{n}}\sqrt{e\cdot\ell} \leq \frac{1}{2} \ ,$$

where we used the choice of ℓ and the assumption $c'_d \log^{d-1}(m) \leq \left(\frac{n}{100 \ln n}\right)^{1/3}$ for the last inequality to hold. We get that the sum $\sum_{1 \leq |S| \leq \ell} |\hat{f}(S)\hat{g}(S)|$ is at most $D_1 + D_2 + \ldots + D_\ell \leq 2D_1$. Overall, we get

$$\mathop{\mathbf{E}}_{x}[f(x)g(x)] \leq |\hat{g}(\emptyset)| + \frac{\sqrt{2} + 8c'_d \log^{d-1}(m)}{\sqrt{n}} \,. \qquad \square$$

We remark that although our proof is Fourier analytical, it differs from the standard argument that is used to bound the correlation of bounded depth circuits with parity for example. The standard argument shows that two functions are $o(1)$ correlated by proving that one is $1 - o(1)$ concentrated on the low levels of the Fourier spectrum while the other is $1 - o(1)$ concentrated on the high levels. Here, however, if we take g to be the Majority function, and f to be an $\mathbf{AC^0}$ circuit, then both f and g are 0.99-concentrated on the first $O(\text{poly}\log(n))$ levels of their Fourier spectrum. We deduce the small correlation by showing that f must be very imbalanced on those levels, which is captured by having small $L_{1,k}$ norm. In contrast, the Majority function is symmetric - its Fourier mass on level k is equally spread on the different coefficients. Combining these two properties guarantees small correlation.

6.2 The Coin-Problem

Theorem 6.2. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be a depth d size m circuit, and let $p \in [0,1]$. Then, f distinguishes between unbiased coins and coins with bias p with advantage at most $6c'_d p \log^{d-1}(m)$.*

Proof. We can assume $pc'_d \log^{d-1}(m) \leq 1/6$, since otherwise the result is trivial. For $-1 \leq p \leq 1$, a p-biased coin is a random variable which gets 1 with probability $(1+p)/2$ and -1 with probability $(1-p)/2$, i.e., this is a biased coin whose expectation is p. Let U_n be the distribution of n independent 0-biased coins, and $B(n,p)$ be the distribution of n independent p-biased coins. We have

$$\begin{aligned}
\text{Distinguishability}(f) &\triangleq \left| \mathop{\mathbf{E}}_{x \sim U_n}[f(x)] - \mathop{\mathbf{E}}_{x \sim B(p,n)}[f(x)] \right| = \left| \hat{f}(\emptyset) - \sum_{S \subseteq [n]} \hat{f}(S) p^{|S|} \right| \\
&= \left| \sum_{S \neq \emptyset} \hat{f}(S) p^{|S|} \right| \leq \sum_{k=1}^{n} p^k \cdot 2 \cdot \left(2c'_d \log^{d-1}(m)\right)^k \\
&\leq 2p \cdot \left(2c'_d \log^{d-1}(m)\right) \cdot \sum_{k=1}^{\infty} (1/3)^{k-1} = 3p \cdot \left(2c'_d \log^{d-1}(m)\right) \,. \qquad \square
\end{aligned}$$

7 A New Proof for Håstad's Switch-Many Lemma

In this section, we give a new proof for Håstad's [Hås14] Switch-Many Lemma, i.e., Lemma 3.6. The new proof follows Razborov's [Raz95] approach, and its recent simplification by Thapen [Tha09] for Håstad's original switching lemma [Hås86].

Notation. We denote by $\mathcal{R}$ the set of all restrictions on n variables. For a sequence of indices $S \in [n]^k$ with no repetitions, and a string $\sigma \in \{0,1\}^k$ we denote by $(S \to \sigma)$ the restriction which fixes S_i to σ_i for $i \in [k]$ and leaves all other variables free. For two restrictions ρ, σ we denote by $\rho\sigma$ their composition. For a sequence $S \in \Sigma^k$ over some alphabet Σ, and two indices i and j such that $1 \leq i \leq j \leq k$, we denote by $S[i:j]$ the subsequence $(S_i, \ldots, S_j)$, and by $S[i]$ the element S_i.

7.1 The Canonical Decision Tree

Let F be an r-**DNF**, i.e., an OR of ANDs where each AND has at most r input literals from $x_1, \ldots, x_n, \neg x_1, \ldots, \neg x_n$. Let ρ be a restriction. The canonical tree $T(F, \rho)$ is defined by the following decision procedure: Look through F for the first term C_1, such that $C_1|_\rho \not\equiv 0$. If no such term exists, then halt and output 0. Otherwise, let A be the set of free variables in C_1 under ρ. Query the variables in A and let $\pi_1, \ldots, \pi_{|A|}$ be their assignment. If the term C_1 is satisfied under the assignment (in particular if $A = \emptyset$), then halt and output 1 . Otherwise, repeat the process with $\rho(A \to (\pi_1, \pi_2, \ldots, \pi_{|A|}))$ instead of ρ. We keep iterating until one of the aforementioned halting conditions hold.

7.2 Restriction Tree for Multiple DNFs

Let $F_1, \ldots, F_m$ be r-**DNFs**. We define the d-restriction-tree complexity of $F_1, \ldots, F_m$ to be the minimal depth of a restriction tree such that under the restriction defined by each leaf, each **DNF** F_i is of canonical decision-tree-complexity at most d. We denote this complexity by $\mathrm{RT}_d(\{F_1, \ldots, F_m\})$.

Theorem 7.1.

$$\Pr_{\rho \sim \mathcal{R}_p}\left[\mathrm{RT}_d(\{F_1|_\rho, \ldots, F_m|_\rho\}) \geq k\right] \leq m^{\lceil k/(d+1)\rceil} \cdot \left(\frac{24pr}{1-p}\right)^k$$

The following is a corollary of Theorem 7.1.

Corollary 7.2. *Let $F_1, \ldots, F_m$ be r-**DNFs**. Let k, d be positive integers, $0 \leq p \leq 1$, and assume $2^{d+1} \geq m$. Then,*

$$\Pr_{\rho \sim \mathcal{R}_p}\left[\mathrm{RT}_d(\{F_1|_\rho, \ldots, F_m|_\rho\}) \geq k\right] \leq m \cdot (49pr)^k \,. \tag{8}$$

Proof. We can assume without loss of generality that $p < 1/49$ since otherwise the RHS of Eq. (8) is at least 1 and the LHS is always at most 1. We get

$$\begin{aligned}
\Pr_{\rho \sim \mathcal{R}_p}\left[\mathrm{RT}_d(\{F_1|_\rho, \ldots, F_m|_\rho\}) \geq k\right] &\leq m^{\lceil k/(d+1)\rceil} \cdot (24pr/(1-p))^k && \text{(Theorem 7.1)} \\
&\leq m^{1+k/(d+1)} \cdot (24pr/(1-p))^k \\
&= m \cdot \left(\frac{m^{1/(d+1)} \cdot 24pr}{1-p}\right)^k \\
&\leq m \cdot \left(\frac{2 \cdot 24pr}{1-p}\right)^k && (m \leq 2^{d+1}) \\
&\leq m \cdot (49pr)^k \,. && (1-p > 48/49)
\end{aligned}$$

□

We will prove Theorem 7.1 based on the approach of Thapen [Tha09] which simplified Razborov's [Raz95] and Beame's [Bea94] proofs for the (original) switching lemma. The idea of the proof is that in order to show that some event A happens with low probability, it is sufficient to show that there exists some other event B (not necessarily disjoint of A) that happens with probability much larger than A. For example, if $\mathbf{Pr}[B] \geq M \cdot \mathbf{Pr}[A]$ (think of M as some large factor) then since $\mathbf{Pr}[B] \leq 1$ it means that $\mathbf{Pr}[A] \leq 1/M$.

The following is the main lemma in this section, from which we deduce Theorem 7.1 quite easily.

Lemma 7.3. *Let S be the set of restrictions under which $\mathrm{RT}_d(\{F_1|_\rho, \ldots, F_m|_\rho\}) \geq k$. Then, there is a 1:1 mapping*

$$\theta : S \to \mathcal{R} \times [3r]^k \times \{0,1\}^k \times \{0,1\}^k \times [m]^{\lceil k/d+1 \rceil}$$

given by $\theta : \rho \mapsto (\rho\sigma, \beta, \pi, \tau, \mathcal{I})$ where σ fixes exactly k additional variables that weren't fixed by ρ.

Proof of Theorem 7.1, assuming Lemma 7.3. For a (fixed) restriction $\rho \in \mathcal{R}$ we denote by $\mathbf{Pr}[\rho]$ the probability to sample ρ when sampling a restriction from the distribution $\mathcal{R}_p$. For a (fixed) set of restrictions $A \subseteq \mathcal{R}$ we denote by $\mathbf{Pr}[A]$ the probability to sample a restriction in A when sampling a restriction from the distribution $\mathcal{R}_p$. Recall that by the definition of $\mathcal{R}_p$, we have $\mathbf{Pr}[\rho] = p^a \cdot \left(\frac{1-p}{2}\right)^{b+c}$ where a, b and c are the number of $*$'s, 0's and 1's in ρ respectively.

For a fixed value of β, π, τ, and $\mathcal{I}$, consider the set $S' = S_{\beta,\pi,\tau,\mathcal{I}} := \{\rho \in S \mid \exists \rho' : \theta(\rho) = (\rho', \beta, \pi, \tau, \mathcal{I})\}$. Since θ is 1:1 (Lemma 7.3), the first component $\theta_1 : \rho \mapsto \rho\sigma$ is also 1:1 on the set S'.[6] This implies that $\mathbf{Pr}[\theta_1(S')] = \sum_{\rho \in S'} \mathbf{Pr}[\theta_1(\rho)]$. By the definition of $\mathcal{R}_p$, for any $\rho \in \mathcal{R}$ and any σ that fixes k additional variables that were free in ρ, we have $\mathbf{Pr}[\rho\sigma] = \left(\frac{1-p}{2p}\right)^k \cdot \mathbf{Pr}[\rho]$. We get

$$1 \geq \mathbf{Pr}[\theta_1(S')] = \sum_{\rho \in S'} \mathbf{Pr}[\theta_1(\rho)] = \sum_{\rho \in S'} \mathbf{Pr}[\rho] \cdot \left(\frac{1-p}{2p}\right)^k = \mathbf{Pr}[S'] \cdot \left(\frac{1-p}{2p}\right)^k ,$$

hence, $\mathbf{Pr}[S'] \leq \left(\frac{2p}{1-p}\right)^k$. Taking a union bound over all possible $\beta, \pi, \tau, \mathcal{I}$ we get, as desired,

$$\mathbf{Pr}[S] \leq \sum_{\beta,\pi,\tau,\mathcal{I}} \mathbf{Pr}[S_{\beta,\pi,\tau\mathcal{I}}] \leq (3r)^k \cdot 2^k \cdot 2^k \cdot m^{\lceil k/(d+1) \rceil} \cdot \left(\frac{2p}{1-p}\right)^k . \qquad \square$$

Proof of Lemma 7.3. Let $\rho \in S$ be a restriction such that $\mathrm{RT}_d(\{F_1|_\rho, \ldots, F_m|_\rho\}) \geq k$. We describe in detail how to map ρ into $(\rho\sigma, \beta, \pi, \tau, \mathcal{I})$, where $\sigma \in \mathcal{R}, \beta \in [3r]^k, \pi \in \{0,1\}^k, \tau \in \{0,1\}^k$, and $\mathcal{I} \in [m]^{\lceil k/(d+1) \rceil}$. Then, we shall describe how to decode from $(\rho\sigma, \beta, \pi, \tau, \mathcal{I})$ the restriction ρ, showing that the mapping is 1:1.

[6] Since a collision in θ_1 on S' implies a collision in θ.

Encoding. We are going to choose a sequence of k variables that weren't fixed by ρ, and assign them values according to three adversarial strategies:

Global Strategy This strategy ensures that $\mathrm{RT}_d(\{F_1|_\rho, \ldots, F_m|_\rho\}) \geq k$. We will denote its answers by $\pi_1, \ldots, \pi_k \in \{0,1\}$.

Local Strategy This will be the local adversary strategy based on one **DNF** we are focusing on. We will denote its answers by $\tau_1, \ldots, \tau_k \in \{0,1\}$.

Bread-Crumbs Strategy The objective of this strategy is to leave the necessary traces, so that the mapping will be invertible. We will denote its answers by $\sigma_1, \ldots, \sigma_k \in \{0,1\}$.

We consider the following iterative encoding process, which is divided into phases. Each phase, except for maybe the last phase, contains at least $d+1$ steps. In each phase, t, we will focus on one specific **DNF** out of $F_1, \ldots, F_m$, and identify a sequence of variables T_t of length $d_t \geq d+1$ to be queried. The strings $\pi^t, \tau^t, \sigma^t \in \{0,1\}^{d_t}$ will be the answers to the sequence of queries T_t according to the Global, Local or Bread-Crumbs strategies, respectively.

At phase $t = 1, 2, \ldots$, we consider the restriction $\rho_t = \rho(T_1 \to \pi^1) \ldots (T_{t-1} \to \pi^{t-1})$. We identify some **DNF**, F_{i_t}, whose canonical decision tree depth under ρ_t is $d_t \geq d+1$ (if no such **DNF** exists, then we stop). We add i_t to $\mathcal{I}$. Next, we run the canonical decision tree on F_{i_t} and ρ_t, answering according to the local adversarial strategy which keeps F_{i_t} undetermined after less than d_t queries.

We initialize $T_t := \emptyset$ and τ^t, σ^t to be empty-strings. In each step, we find the first term $\mathcal{T}$ in F_{i_t} which is not equivalent to 0 under $\rho_t(T_t \to \tau^t)$. By the assumption that F_{i_t} has canonical decision tree depth d_t, we get that $\mathcal{T}$ is not equivalent to 1 either. Let A be the non-empty set of variables whose literals appear in $\mathcal{T}$ and are unassigned by ρ_t. We order the set A in some canonical order. For each $x_j \in A$, let $j_{\text{ind}} \in [r]$ be the index of the literal containing x_j in term $\mathcal{T}$. We let $j_{\text{type}} = 1$ if x_j is the last variable to be queried in the DNF F_{i_t}, otherwise $j_{\text{type}} = 2$ if x_j is the last variable in A, and otherwise $j_{\text{type}} = 3$. For each $x_j \in A$, according to the A's order, we add $(j_{\text{ind}}, j_{\text{type}})$ to β. In addition, we query the local adversary according to the variables in A under the restriction $\rho_t \cdot (T_t \to \tau^t)$ and update τ^t to contain its new answers. We concatenate to σ^t the values to the variables in A that satisfy $\mathcal{T}$ (these are the "bread-crumbs"). We update $T_t := T_t \cup A$, and continue with $\rho_t \cdot (T_t \to \tau^t)$ until querying d_t variables.

After ending the phase, we ask the global adversary the sequence of queries in T_t (by the order they were asked) and consider its sequence of answers as π^t. We continue to the next phase with $\rho_{t+1} = \rho(T_1 \to \pi^1) \ldots (T_t \to \pi^t)$ (i.e., we "discard" the answers to T_t according to the local adversary and add the answers according to the global adversary). We stop the encoding process after querying k variables overall, even if we are in the middle of a phase.

We show by induction that $\mathrm{RT}_d(\{F_1|_{\rho_t}, \ldots, F_m|_{\rho_t}\}) \geq k - \sum_{i=1}^{t-1} d_i$. This is trivially true for $t = 1$ since this is equivalent to the assumption that $\rho \in S$. Assuming it is true for t, we show that it is true for $t+1$. Since $\mathrm{RT}_d(\{F_1|_{\rho_t}, \ldots, F_m|_{\rho_t}\}) \geq k - \sum_{i=1}^{t-1} d_i$ it means that there exists a set of answers for T_t, namely π^t, under which $\mathrm{RT}_d(\{F_1|_{\rho_{t+1}}, \ldots, F_m|_{\rho_{t+1}}\}) \geq k - \sum_{i=1}^{t-1} d_i - |T_t| = k - \sum_{i=1}^{t} d_i$, which completes the induction.

Let p be the number of phases in the encoding process. By the above process, we get that $\pi = \pi^1 \ldots \pi^p \in \{0,1\}^k$, $\tau = \tau^1 \ldots \tau^p \in \{0,1\}^k$, $\beta \in [3r]^k$, $\sigma := (T_1 \to \sigma^1) \ldots (T_p \to \sigma^p)$

fixes k additional variables to those fixed by ρ, and $\mathcal{I}$ is a sequence of p indices from $[m]$. In addition, $p \leq \lceil k/(d+1) \rceil$ since in each phase, except for maybe the last phase, we query at least $d+1$ variables and overall we query at most k variables. If less than $\lceil k/(d+1) \rceil$ phases exists, we may pad $\mathcal{I}$ with 1's.

Decoding. We wish to show that θ is 1:1. Let $(\rho\sigma, \beta, \pi, \tau, \mathcal{I})$ be an image of θ; we will show how to decode ρ from this image. It is enough to show by induction on $t = 1, \ldots, p$, that we can recover $T_1, \ldots, T_t$, since this allows to reconstruct ρ by simply setting the values of $\bigcup_{i=1}^{p} T_i$ to $*$ in $\rho\sigma$.

Assuming we already recovered $T_1, \ldots, T_{t-1}$ correctly, we show how to decode T_t as well. Knowing $T_1, \ldots, T_{t-1}$ allows the decoder to define $\rho'_t := \rho(T_1 \to \pi^1) \ldots (T_{t-1} \to \pi^{t-1})(T_t \to \sigma^t) \ldots (T_p \to \sigma^p)$ by replacing the assignment of $T_1, \ldots, T_{t-1}$ in $\rho\sigma$ according to $\pi^1, \ldots, \pi^t$. Using the set of indices $\mathcal{I}$, we know i_t, i.e. the index of the **DNF** out of $F_1, \ldots, F_m$ that was considered by the encoding process at phase t. We show that the first term in F_{i_t} under ρ'_t which is not equivalent to 0 is the same as the first such term under ρ_t (recall that $\rho_t := \rho(T_1 \to \pi^1) \ldots (T_{t-1} \to \pi^{t-1})$). Let i' be the index of the first nonzero term in F_{i_t} under ρ_t. Then, all terms prior to i' were fixed to 0 under ρ_t and this remains true when we refine ρ_t to ρ'_t. In addition, since σ^t satisfies all the literals in term i' which are unassigned by ρ_t (except if we finished the entire encoding process while in the middle of processing this term, in this case σ^t fixes some of the free variables to satisfy the literals and the rest remain free), we get that the term with index i' in F_{i_t} is not equivalent to 0 under ρ'_t. Thus, we identified the term i' correctly. We collect indices from β until reaching type 1 or 2, which yields the set of variables the encoder sets when processing term i'. We replace the assignment for these variables to be according to τ instead of according to σ.

We continue this way by identifying the next term the encoder examined, and decode the set of variables fixed in the encoding process, according to the information stored in β. This allows us to continue decoding the set T_t which completes the proof. □

Remark 7.4. *We remark that the information we are avoiding to store*[7] *is the index of the term on which a certain* **DNF** *is not fixed under a restriction ρ. We are using the Bread-Crumbs partial assignment σ to satisfy all the literals that are unassigned in this term, in order to allow the identification of the term in the decoding process. Once the term is known, we can encode/decode a variable using a number in $[r]$ rather than a number in $[n]$, which is much more "inexpensive" to encode. Storing the index to the DNF we are considering at each phase may seem "expensive". However, we are recording such an index at most once in every $d+1$ consecutive steps, making this reasonable.*

Acknowledgement I wish to thank my advisor Ran Raz for many helpful discussions, for his encouragement and his support. I thank Johan Håstad and Roei Tell for helpful discussions. I thank the anonymous referees for helpful comments.

[7] And we have good reasons to do so, since this will result in a non-effective switching lemma.

References

[Aar10] S. Aaronson. BQP and the polynomial hierarchy. In *STOC*, pages 141–150, 2010.

[AGHP92] N. Alon, O. Goldreich, J. Håstad, and R. Peralta. Simple construction of almost k-wise independent random variables. *Random Structures and Algorithms*, 3(3):289–304, 1992.

[Ajt83] M. Ajtai. Σ_1^1-formulae on finite structures. *Annals of Pure and Applied Logic*, 24:1–48, 1983.

[Baz09] L. M. J. Bazzi. Polylogarithmic independence can fool DNF formulas. *SIAM J. Comput.*, 38(6):2220–2272, 2009.

[Bea94] P. Beame. A switching lemma primer. 1994.

[Bop97] R. B. Boppana. The average sensitivity of bounded-depth circuits. *Inf. Process. Lett.*, 63(5):257–261, 1997.

[Bra10] M. Braverman. Polylogarithmic independence fools ac^0 circuits. *J. ACM*, 57(5):28:1–28:10, 2010.

[CGR14] G. Cohen, A. Ganor, and R. Raz. Two sides of the coin problem. In *APPROX-RANDOM*, pages 618–629, 2014.

[DETT10] A. De, O. Etesami, L. Trevisan, and M. Tulsiani. Improved pseudorandom generators for depth 2 circuits. In *APPROX-RANDOM*, pages 504–517, 2010.

[Fil10] Y. Filmus. Smolensky's lower bound. Unpublished Manuscript, 2010.

[FSS84] M. Furst, J. B. Saxe, and M. Sipser. Parity, circuits, and the polynomial-time hierarchy. *Mathematical Systems Theory*, 17(1):13–27, apr 1984.

[GL89] O. Goldreich and L. A. Levin. A hardcore predicate for all one-way functions. In *STOC*, pages 25–32, 1989.

[Hås86] J. Håstad. Almost optimal lower bounds for small depth circuits. In *STOC*, pages 6–20, 1986.

[Hås01] J. Håstad. A slight sharpening of LMN. *J. Comput. Syst. Sci.*, 63(3):498–508, 2001.

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

[HS16] P. Harsha and S. Srinivasan. On polynomial approximations to ac^0. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2016, September 7-9, 2016, Paris, France*, pages 32:1–32:14, 2016.

[IK14] R. Impagliazzo and V. Kabanets. Fourier concentration from shrinkage. In *CCC*, pages 321–332, 2014.

[IMP12] R. Impagliazzo, W. Matthews, and R. Paturi. A satisfiability algorithm for AC^0. In *SODA*, pages 961–972, 2012.

[KB80] R. Kaas and J. M. Buhrman. Mean, median and mode in binomial distributions. *Statistica Neerlandica*, 34(1):13–18, 1980.

[KKL88] J. Kahn, G. Kalai, and N. Linial. The influence of variables on Boolean functions. In *FOCS*, pages 68–80, 1988.

[KM93] E. Kushilevitz and Y. Mansour. Learning decision trees using the Fourier spectrum. *SIAM J. Comput.*, 22(6):1331–1348, 1993.

[LMN93] N. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, Fourier transform and learnability. *J. ACM*, 40(3):607–620, 1993.

[Lup61] O. Lupanov. Implementing the algebra of logic functions in terms of constant depth formulas in the basis $\{\&, \vee, \neg\}$. *Dokl. Akad. Nauk. SSSR*, 136:1041–1042, 1961. In Russian.

[LV96] M. Luby and B. Velickovic. On deterministic approximation of DNF. *Algorithmica*, 16(4/5):415–433, 1996.

[Man95] Y. Mansour. An $O(n^{\log\log n})$ learning algorithm for DNF under the uniform distribution. *J. Comput. Syst. Sci.*, 50(3):543–550, 1995.

[NN93] J. Naor and M. Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM J. on Computing*, 22(4):838–856, 1993.

[O'D14] R. O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

[OW07] R. O'Donnell and K. Wimmer. Approximation by DNF: examples and counterexamples. In *ICALP*, pages 195–206, 2007.

[Raz95] A. A. Razborov. Bounded arithmetic and lower bounds in boolean complexity. In *Feasible Mathematics II*, volume 13 of *Progress in Computer Science and Applied Logic*, pages 344–386. Birkhuser Boston, 1995.

[Raz09] A. A. Razborov. A simple proof of Bazzi's theorem. *TOCT*, 1(1), 2009.

[Smo93] R. Smolensky. On representations by low-degree polynomials. In *FOCS 1993*, pages 130–138, 1993.

[SV10] R. Shaltiel and E. Viola. Hardness amplification proofs require majority. *SIAM J. Comput.*, 39(7):3122–3154, 2010.

[Tal14] A. Tal. Shrinkage of de Morgan formulae from quantum query complexity. In *FOCS*, pages 551–560, 2014.

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

[TX13] L. Trevisan and T. Xue. A derandomized switching lemma and an improved derandomization of AC0. In *CCC*, pages 242–247, 2013.

[Yao85] A. C. Yao. Separating the polynomial hierarchy by oracles. In *FOCS*, pages 1–10, 1985.

A Equivalent Expressions for the T-th Discrete Derivatives

Claim (Claim 5.3, restated)**.**

$$D_T f(x) = \frac{1}{2^{|T|}} \sum_{z \in \{-1,1\}^T} f(x^{(T \mapsto z)}) \cdot \prod_{i \in T} z_i = \sum_{S \supseteq T} \hat{f}(S) \cdot \prod_{i \in S \setminus T} x_i$$

where $x^{(T \mapsto z)}$ is the vector in $\{-1,1\}^n$ whose i-th coordinate equals z_i whenever $i \in T$, and x_i otherwise.

Proof. We prove by induction on the size of T. For $T = \emptyset$ the claim trivially holds. For $T = \{j_1, \ldots, j_k\}$, let $T' = \{j_2, \ldots, j_k\}$ and $g = D_{T'} f$, then $D_T f = D_{j_1} D_{T'} f = D_{j_1} g$. By the definition of the j_1-th derivative, we have

$$D_T f(x) = \frac{g(x^{(j_1 \mapsto 1)}) - g(x^{(j_1 \mapsto -1)})}{2} \,.$$

By the induction hypothesis, this equals

$$\begin{aligned} D_T f(x) &= \frac{1}{2} \cdot \left(D_{T'} f(x^{(j_1 \mapsto 1)}) - D_{T'} f(x^{(j_1 \mapsto -1)}) \right) \\ &= \frac{1}{2} \frac{1}{2^{k-1}} \left(\sum_{z' \in \{-1,1\}^{T'}} f\left(\left(x^{(j_1 \mapsto 1)}\right)^{(T' \mapsto z')} \right) \prod_{i \in T'} z'_i - \sum_{z' \in \{-1,1\}^{T'}} f\left(\left(x^{(j_1 \mapsto -1)}\right)^{(T' \mapsto z')} \right) \prod_{i \in T'} z'_i \right) \\ &= \frac{1}{2^k} \sum_{z \in \{-1,1\}^T} f(x^{(T \mapsto z)}) \prod_{i \in T} z_i \,. \end{aligned}$$

As for the second item, by induction, $g(x) = \sum_{S \supseteq T'} \hat{f}(S) \cdot \prod_{i \in S \setminus T'} x_i$. Thus,

$$\begin{aligned} D_T f(x) = \frac{g(x^{(j_1 \mapsto 1)}) - g(x^{(j_1 \mapsto -1)})}{2} &= \frac{1}{2} \sum_{S \supseteq T'} \hat{f}(S) \cdot \prod_{i \in S \setminus T} x_i \cdot \begin{cases} 1 - (-1), & j_1 \in T' \\ 1 - 1, & \textit{otherwise} \end{cases} \\ &= \sum_{S \supseteq T} \hat{f}(S) \cdot \prod_{i \in S \setminus T} x_i \,. \end{aligned}$$

□

B Rephrasing Braverman's Result

Lemma B.1 ([Bra10, Lemma 8]). *Let ν be any probability distribution on $\{0,1\}^n$. For a circuit of depth d and size m computing a function F, for any s, there is a degree $r = (s\cdot\log(m))^d$ polynomial f and a Boolean function $\mathcal{E}_\nu$ computable by a circuit of depth $\leq d+3$ and size $O(m^2 r)$ such that*

1. $\mathbf{Pr}_\nu[\mathcal{E}_\nu(x)=1] < 0.82^s \cdot m$, *and*
2. *whenever $\mathcal{E}_\nu = 0$, $f(x) = F(x)$.*

Proposition B.2 ([Bra10, Prop. 9]). *In Lemma B.1, for $s \geq \log(m)$, $\|f\|_\infty < (2m)^{\deg(f)-2} = (2m)^{(s\log(m))^d-2}$*

Lemma B.3 ([Bra10, Rephrasing of Lemma 10]). *Let F be computed by a circuit of depth d and size m. Let s_1, s_2 be two parameters with $s_1 \geq \log(m)$. Let μ be any probability distribution on $\{0,1\}^n$, and $U_{\{0,1\}^n}$ be the uniform distribution on $\{0,1\}^n$. Set*

$$\nu := \frac{1}{2}\left(\mu + U_{\{0,1\}^n}\right) .$$

Let $\mathcal{E}_\nu$ be the function from Lemma 8 with $s = s_1$. Set $F' = F \vee \mathcal{E}_\nu$. Then, there is a polynomial f' of degree $r_f = (s_1 \cdot \log m)^d + s_2$, such that

1. $\mathbf{Pr}_\mu[F \neq F'] < 2 \cdot 0.82^{s_1} \cdot m$
2. $\mathbf{Pr}_U[F \neq F'] < 2 \cdot 0.82^{s_1} \cdot m$
3. $\|F' - f'\|_2^2 \leq 0.82^{s_1} \cdot (4m) + 2^{4(s_1\cdot\log m)^d \log m} \cdot \mathrm{tail}(m^3, d+3, s_2)$, *and*
4. $f'(x) = 0$ *whenever* $F'(x) = 0$.

Proof. The first two properties follow from Lemma B.1 directly, since

$$\mathop{\mathbf{Pr}}_{\mu}[\mathcal{E}_\nu = 1], \mathop{\mathbf{Pr}}_{U_n}[\mathcal{E}_\nu = 1] \leq 2 \cdot \mathop{\mathbf{Pr}}_{\nu}[\mathcal{E}_\nu = 1] \leq 2 \cdot 0.82^{s_1} m .$$

Let f be the degree $(s_1 \cdot \log m))^d$ approximation of F from Lemma B.1. By Proposition B.2,

$$\|f\|_\infty < (2m)^{(s_1\cdot\log m)^d - 2} < 2^{2(s_1 \log m)^d \log(m) - 2} .$$

Let $\tilde{\mathcal{E}}_\nu$ be the truncated Fourier expansion of $\mathcal{E}_\nu$ of degree s_2. We have

$$\|\mathcal{E}_\nu - \tilde{\mathcal{E}}_\nu\|_2^2 \leq \mathrm{tail}(m^3, d+3, s_2) .$$

Let

$$f' := f \cdot (1 - \tilde{\mathcal{E}}_\nu)$$

Then $f' = 0$ whenever $F' = 0$ (since $(F' = 0) \implies (\mathcal{E}_\nu = 0, F = 0) \implies (f = 0) \implies (f' = 0)$). It remains to estimate $\|F' - f'\|_2^2$:

$$\begin{aligned}\|F' - f'\|_2^2 &\le 2 \cdot \|F' - f \cdot (1 - \mathcal{E}_\nu)\|_2^2 + 2 \cdot \|f \cdot (1 - \mathcal{E}_\nu) - f'\|_2^2 \\ &= 2 \cdot \|\mathcal{E}_\nu\|_2^2 + 2 \cdot \|f \cdot (\mathcal{E}_\nu - \tilde{\mathcal{E}}_\nu)\|_2^2 \\ &\le 2 \cdot \mathbf{Pr}[\mathcal{E}_\nu = 1] + 2 \cdot \|f\|_\infty^2 \cdot \|\mathcal{E}_\nu - \tilde{\mathcal{E}}_\nu\|_2^2 \\ &\le 0.82^{s_1} \cdot (4m) + 2^{4(s_1 \cdot \log m)^d \log m} \cdot \mathrm{tail}(m^3, d+3, s_2),\end{aligned}$$

which completes the proof. □

Theorem B.4 ([Bra10, Rephrasing of Main Theorem])**.** *Let $s_1, s_2 \ge \log m$ be any parameters. Let F be a Boolean function computed by a circuit of depth d and size m. Let μ be an r-independent distribution where*

$$r = r(s_1, s_2, d) = 2((s_1 \cdot \log m)^d + s_2)$$

then

$$|\mathop{\mathbf{E}}_{\mu}[F] - E[F]| \le \varepsilon(s_1, s_2, d),$$

where $\varepsilon(s_1, s_2, d) = 0.82^{s_1} \cdot (6m) + 2^{4(s_1 \cdot \log m)^d \log m} \cdot \mathrm{tail}(m^3, d+3, s_2)$

Proof of Theorem B.4. Denote by $\varepsilon_1 := 0.82^{s_1} \cdot (2m)$ and

$$\varepsilon_2 := 0.82^{s_1} \cdot (4m) + 2^{4(s_1 \cdot \log m)^d \log m} \cdot \mathrm{tail}(m^3, d+3, s_2)\ .$$

Applying Lemma B.3 with parameters s_1 and s_2 gives

$$\|F' - f'\|_2^2 \le \varepsilon_2\ .$$

Now take $f'_\ell := 1 - (1 - f')^2$. Then $f'_\ell \le 1$ and $f'_\ell = 0$ whenever $F' = 0$, hence $f'_\ell \le F'$. To estimate $\mathbf{E}[F'(x) - f'_\ell(x)]$ we note that $F'(x) - f'_\ell(x)$ equals 0 whenever $F' = 0$, and is equal to

$$F'(x) - f'_\ell(x) = (1 - f'(x))^2 = (F'(x) - f'(x))^2$$

whenever $F' = 1$. We get

$$\mathbf{E}[F'(x) - f'_\ell(x)] \le \|F' - f'\|_2^2 \le \varepsilon_2\ .$$

In addition, $\deg(f'_\ell(x)) \le 2(s_2 + (s_1 \cdot \log m)^d)$.

To finish the proof, if μ is a $\left(2 \cdot (s_2 + (s_1 \cdot \log m)^d)\right)$-wise independent distribution then

$$\begin{aligned}\mathop{\mathbf{E}}_{\mu}[F(x)] &\ge \mathop{\mathbf{E}}_{\mu}[F'(x)] - \varepsilon_1 \ge \mathop{\mathbf{E}}_{\mu}[f'_\ell(x)] - \varepsilon_1 =^* \mathbf{E}[f'_\ell(x)] - \varepsilon_1 \\ &= \mathbf{E}[F'(x)] - \mathbf{E}[F'(x) - f'_\ell(x)] - \varepsilon_1 \ge \mathbf{E}[F'(x)] - \varepsilon_2 - \varepsilon_1 \ge \mathbf{E}[F(x)] - \varepsilon_2 - \varepsilon_1\end{aligned}$$

where we used in * the fact that $\deg(f'_\ell) \le 2(s_2 + (s_1 \cdot \log m)^d)$ and μ is $\deg(f'_\ell)$-wise independent. In a similar way, one can show $\mathbf{E}_\mu[F(x)] \le \mathbf{E}[F(x)] + \varepsilon_1 + \varepsilon_2$. Combining both cases we get

$$|\mathop{\mathbf{E}}_{\mu}[F] - E[F]| \le \varepsilon_1 + \varepsilon_2 = \varepsilon(s_1, s_2, d)\ .$$ □

C Improving the Analysis of De, Etesami, Trevisan and Tulsiani

De et al. [DETT10] proved that any ε-biased distribution δ-fools depth-2 circuits (**DNFs** or **CNFs**) of size m, for some $\varepsilon = \varepsilon(\delta, m)$. In fact, their work shows that generators of ε-biased distributions are the best known pseudorandom generators fooling depth-2 circuits. We are able to improve their analysis slightly, getting an optimal dependence between ε and δ.

Some notation is needed first. Throughout this section (and the next), we shall think of Boolean functions as functions $f : \{0,1\}^n \to \mathbb{R}$ (as opposed to $f : \{-1,1\}^n \to \mathbb{R}$). We can identify each function $f : \{0,1\}^n \to \mathbb{R}$ with a function $\tilde{f} : \{-1,1\}^n \to \mathbb{R}$ by $\tilde{f}(y_1, \ldots, y_n) = f(\frac{1-y_1}{2}, \ldots, \frac{1-y_n}{2})$ or equivalently $f(x_1, \ldots, x_n) = \tilde{f}((-1)^{x_1}, \ldots, (-1)^{x_n})$. When talking about the Fourier expansion of f, we mean the Fourier expansion of $\tilde{f}$ as defined in Section 2. In this notation, $f(x) = \sum_{S \subseteq [n]} \hat{f}(S) \cdot (-1)^{\sum_{i \in S} x_i}$.

Next, we discuss **DNFs** and **CNFs**. Disjunctive normal forms (**DNFs**) are expressions of the form $F(x) = \bigvee_{i=1}^{m} t_i(x)$ where each term $t_i(x)$ is an AND of some literals from $x_1, \ldots, x_n, \neg x_1, \ldots, \neg x_n$. If any term in F is an AND of at most w literals, then we say that F is of width w, and we call F a w-**DNF**. Similarly conjunctive normal forms (**CNFs**) are expressions of the form $F(x) = \bigwedge_{i=1}^{m} c_i(x)$ where each clause c_i is an OR of some literals. We define w-**CNFs** similarly to w-**DNFs**. The size of a **DNF** (**CNF**, resp.) is the number of terms (clauses, resp.) in it, i.e., m in the examples above.

Recall the definition of the spectral norm of a Boolean function $L_1(f) = \sum_S |\hat{f}(S)|$ and denote by $L_1^*(f) = \sum_{S \neq \emptyset} |\hat{f}(S)|$. We denote by U_n the uniform distribution over $\{0,1\}^n$.

We cite a proposition and two lemmata from the work of De et al. [DETT10].

Proposition C.1 ([DETT10, Prop. 2.6])**.** *Suppose $f, f_\ell, f_u : \{0,1\}^n \to \mathbb{R}$ are three functions such that for every $x \in \{0,1\}^n$ we have $f_\ell(x) \leq f(x) \leq f_u(x)$. Furthermore, assume $\mathbf{E}_{x \sim U_n}[f(x) - f_\ell(x)] \leq \delta$ and $\mathbf{E}_{x \sim U_n}[f_u(x) - f(x)] \leq \delta$. Let $l = \max(L_1^*(f_\ell), L_1^*(f_u))$. Then, any ε-biased probability distribution $(\delta + \varepsilon l)$-fools f.*

Lemma C.2 ([DETT10, Lemma 4.3])**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be a* **DNF** *with m terms and $g : \{0,1\}^n \to \mathbb{R}$ be such that: $L_1(g) \leq l_1$, $\|f - g\|_2^2 \leq \varepsilon_1$ and $g(x) = 0$ whenever $f(x) = 0$. Then, we can get $f_\ell, f_u : \{0,1\}^n \to \mathbb{R}$ such that*

- $\forall x, f_\ell(x) \leq f(x) \leq f_u(x)$
- $\mathbf{E}_{x \sim U_n}[f_u(x) - f(x)] \leq m \cdot \varepsilon_1$ *and* $\mathbf{E}_{x \sim U_n}[f(x) - f_\ell(x)] \leq m \cdot \varepsilon_1$.
- $L_1(f_\ell), L_1(f_u) \leq (m+1)(l_1+1)^2 + 1$.

Lemma C.3 ([DETT10, Lemma 4.4])**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be a* **DNF** *with m terms and width-w. Suppose for every* **DNF** *with at most m terms and width-w, f_1, there is a function $g_1 : \{0,1\}^n \to \mathbb{R}$ such that: $L_1(g_1) \leq l_2$ and $\|f_1 - g_1\|_2^2 \leq \varepsilon_2$. Then, we can get $g : \{0,1\}^n \to \mathbb{R}$ such that $L_1(g) \leq m \cdot (l_2 + 1)$, $\|f - g\|_2^2 \leq m^2 \cdot \varepsilon_2$ and $g(x) = 0$ whenever $f(x) = 0$.*

De et al. [DETT10] used Lemma C.3 with a bound on the width of the approximated **DNF**. We will use Lemma C.3 without any assumption on the width.

The following is a corollary of Thm. 5.15.

Corollary C.4. *Let $f : \{0,1\}^n \to \{0,1\}$ be a* **DNF** *of size m and $\varepsilon_2 > 0$. Then, there is a function $g_1 : \{0,1\}^n \to \mathbb{R}$ such that $\mathbf{E}[(f - g_1)^2] \leq \varepsilon_2$ and $L_1(g_1) = 2^{O(\log m \cdot \log\log m \cdot \log(1/\varepsilon_2))}$.*

Proof. According to Thm. 5.15, there is a set $\mathcal{F}$ of coefficients such that $|\mathcal{F}| \leq 2^{O(\log m \cdot \log\log m \cdot \log(1/\varepsilon_2))}$ and $\sum_{S \notin \mathcal{F}} \hat{f}(S)^2 \leq \varepsilon_2$. Hence, $g_1(x) = \sum_{S \in \mathcal{F}} \hat{f}(S) \cdot (-1)^{\sum_{i \in S} x_i}$ is an approximation of f with

$$\mathop{\mathbf{E}}_{x}[(f(x) - g_1(x))^2] = \sum_{S \notin \mathcal{F}} \hat{f}(S)^2 \leq \varepsilon_2 \,.$$

where we used Parseval's identity. Since each Fourier coefficient is at most 1 in absolute value $L_1(g_1) = \sum_{S \in \mathcal{F}} |\hat{f}(S)| \leq |\mathcal{F}|$, which completes the proof. □

The following theorem is our refinement of [DETT10, Thm. 4.1].

Theorem C.5. *Let f be a* **DNF** *formula with m terms. Then, f is δ-fooled by any ε-biased distribution where $\varepsilon = 2^{O(\log m \cdot \log(m/\delta) \cdot \log\log m)}$.*

Proof. Set $\varepsilon_2 = \delta/2m^3$ and $\varepsilon_1 = \delta/2m$. By applying Corollary C.4 for every **DNF** formula of size m, f_1, there exists a function $g_1 : \{0,1\}^n \to \mathbb{R}$ such that

- $\mathbf{E}[(f_1 - g_1)^2] \leq \varepsilon_2$
- $L_1(g_1) \leq 2^{O(\log m \cdot \log\log m \cdot \log(1/\varepsilon_2))} = 2^{O(\log m \cdot \log\log m \cdot \log(m/\delta))}$

We apply Lemma C.3 with width n (this is a trivial choice of width, since all **DNFs** on n variables are of width at most n without loss of generality), $\varepsilon_2 = \delta/2m^3$ and $l_2 = 2^{O(\log m \cdot \log\log m \cdot \log(m/\delta))}$. Then, we get the existence of a function $g : \{0,1\}^n \to \mathbb{R}$ such that $g(x) = 0$ whenever $f(x) = 0$ and $\mathbf{E}[(g - f)^2] \leq m^2\varepsilon_2 = \delta/2m$. and $L_1(g) \leq (l_2 + 1) \cdot m = 2^{O(\log m \cdot \log\log m \cdot \log(m/\delta))}$. Then, we apply Lemma C.2 with g, $\varepsilon_1 = \delta/2m$ and $l_1 = L_1(g) = 2^{O(\log m \cdot \log\log m \cdot \log(m/\delta))}$ to get a sandwiching approximation of f by f_ℓ and f_u such that

- $\forall x : f_\ell(x) \leq f(x) \leq f_u(x)$
- $\mathbf{E}_{x \sim U_n}[f_u(x) - f(x)] \leq m \cdot \varepsilon_1 = \delta/2$ and $\mathbf{E}_{x \sim U_n}[f(x) - f_\ell(x)] \leq m \cdot \varepsilon_1 = \delta/2$.
- $L_1(f_u), L_1(f_\ell) \leq (l_1 + 1)^2 \cdot (m + 1) + 1 = 2^{O(\log m \cdot \log\log m \cdot \log(m/\delta))}$.

Denote by $l = (l_1 + 1) \cdot (m + 1) + 1$. Applying Prop. C.1, we get that any $\varepsilon = \delta/(2l) = 2^{-O(\log m \cdot \log\log m \cdot \log(m/\delta))}$ biased distribution γ-fools f, where $\gamma = \delta/2 + \varepsilon \cdot l \leq \delta$. □

It is well-known from the works of [NN93, AGHP92] that ε-biased distributions on n bits may be sampled using a $O(\log n + \log(1/\varepsilon))$-seed length, which gives the following corollary.

Theorem C.6. *There exists a polynomial time pseudorandom generator $\mathcal{G}$ of seed length $O(\log n + \log m \cdot \log\log m \cdot \log(m/\delta))$ that δ-fools all* **DNFs** *of size m on n variables.*

Note that by de Morgan laws every **DNF** of size m is the negation of a **CNF** of size m, and vice versa. Hence, any pseudorandom generator that fools **DNFs** also fools **CNFs**.

D Improving the Generator of Trevisan and Xue

In this section, we revisit the pseudorandom generator of Trevisan and Xue [TX13] that ε-fools $\mathbf{AC^0}$ circuits of size M and depth d. We improve its seed-length from $O(\log^{d+3}(M/\varepsilon) \cdot \log(n/\varepsilon))$ to $O(\log^{d+1}(M/\varepsilon) \cdot \log n)$ by two observations in addition to the improved analysis of the generator of De et al. (see Thm. C.6).

We start by explaining the sampling process of Trevisan and Xue's generator at a high-level. The generator applies $O(\log^{d-1}(M/\varepsilon) \cdot \log(n/\varepsilon))$ pseudorandom restrictions iteratively, where each pseudorandom restriction fixes each variable (that wasn't already fixed) with probability $\Theta(1/\log^{d-1}(M/\varepsilon))$. The seed length required per step is $\tilde{O}(\log^4(M/\varepsilon))$. Each pseudo-random restriction consists of a pseudorandom process that selects which variables to fix, in addition to a pseudorandom process that selects the values for these variables. The heart of Trevisan and Xue's analysis is a proof that the selection of which variables to fix can be done by sampling recursively d times (one per depth) from any distribution that fools **CNFs** with appropriate parameters. This is done by proving that any distribution that fools **CNFs**, also fools Håstad's switching lemma [Hås86] (see Lemma D.1 below).

Our improvement from seed-length $\tilde{O}(\log^{d+3}(M/\varepsilon) \cdot \log(n/\varepsilon))$ to $\tilde{O}(\log^{d+1}(M/\varepsilon) \cdot \log n)$ is a combination of three improvements:

- We get a factor of $\log(M/\varepsilon)$ improvement via a better analysis of the pseudorandom generator of De et al. [DETT10] (see Section C). We get a better dependency on the error parameter ε_0 in Thm. C.6, compared to the corresponding theorem of [DETT10]. Since Trevisan and Xue use Thm. C.6 with error parameter $\varepsilon_0 = 1/2^{\Theta(\log^2(M/\varepsilon))}$ that is much smaller than any polynomial in ε/M, this improvement is effective.

- We get a factor of $\log(M/\varepsilon)$ improvement by applying the switching lemma for one less step. We show that with high probability the circuit collapses to a depth-2 circuit instead of collapsing to a bounded depth decision tree. Since we are able to fool depth-2 circuits by Thm. C.6, this is enough.[8]

- We replace a factor of $\log(n/\varepsilon)$ by a factor of $\log(n)$ by noting that one can continue restricting variables until less than $O(\log(1/\varepsilon))$ variables are alive, and then fix the remaining variables using a $O(\log(1/\varepsilon))$-wise independent distribution. In the original analysis, one waited until all variables were fixed.

In the rest of the section, we will use the following notation (as suggested by Trevisan and Xue [TX13]). A restriction may be defined (not uniquely) by two binary strings of length n: $\theta \in \{*, \square\}^n$ and $\beta \in \{0,1\}^n$, where for $i \in [n]$,

$$\rho(i) = \begin{cases} \beta(i), & \theta(i) = \square \\ *, & \theta(i) = * \end{cases}.$$

We shall identify a string $w \in \{0,1\}^{n(q+1)}$ with a restriction as follows. We partition w to (l, r) where l consists of the first qn bits of w, and r consists of the last n bits of w. We

[8] To get another $\tilde{O}(\log(M/\varepsilon))$ improvement it is enough to construct PRGs for depth-3 circuits with seed-length $\tilde{O}(\log^3(M/\varepsilon))$. Then, one can stop one step sooner (i.e. when reaching depth-3) and apply the PRG for depth-3 on the remaining circuit.

further partition $l \in \{0,1\}^{nq}$ to n blocks of q consecutive bits each. For block $i \in [n]$, we take $\theta(i) = *$ iff all the q bits in the block equal 1. We take $\beta = r$ and yield the restriction defined by (θ, β).

If D is a distribution over $\{0,1\}^{(q+1)n}$, then $(\theta, \beta) \sim D$ means that we sample $w \sim D$ as a string of length $(q+1)n$ and use the aforementioned identification to get $\theta \in \{*, \square\}^n$ and $\beta \in \{0,1\}^n$. Note that sampling $w \in \{0,1\}^{n\cdot(q+1)}$ uniformly at random yields a restriction $\rho = (\theta, \beta)$ distributed according to $\mathcal{R}_p$ for $p = 2^{-q}$.

Lemma D.1 ([TX13, Lemma 7])**.** *Let F be a* **CNF** *of size M and width t over n variables, $p_0 = 2^{-q_0}$ where $q_0 \in \mathbb{N}$, and D be a distribution over $\{0,1\}^{(q_0+1)n}$ that ε_0-fools all* **CNFs** *of size at most $M \cdot 2^{t\cdot(q_0+1)}$. Then,*

$$\Pr_{(\theta,\beta)\sim D}[\mathrm{DT}(F|_{\theta,\beta}) > s] \leq 2^{s+t+1} \cdot (5p_0 t)^s + \varepsilon_0 \cdot 2^{(s+1)\cdot(2t+\log M)},$$

where $\mathrm{DT}(f)$ denotes the depth of the smallest decision tree computing a function f.[9]

The following is a slight generalization of [TX13, Fact 9].

Fact D.2 ([TX13, Fact 9])**.** *Let D_1 be a distribution over $\{0,1\}^{n_1}$ that ε_1-fools* **CNFs** *of size m on n_1 variables. Let D_2 be a distribution over $\{0,1\}^{n_2}$ that ε_2-fools* **CNFs** *of size m on n_2 variables. Let $D_1 \otimes D_2$ be the distribution over $\{0,1\}^{n_1+n_2}$ sampled by concatenating independent samples from D_1 and D_2. Then, $D_1 \otimes D_2$ is a distribution that $(\varepsilon_1 + \varepsilon_2)$-fools* **CNFs** *of size m on $n_1 + n_2$ variables.*

Proof. Let $F(X, Y)$ be a **CNF** of size m on $n_1 + n_2$ variables, where X consists of the first n_1 variables and Y consists of the last n_2 variables. We have

$$\mathop{\mathbf{E}}_{x\sim U_{n_1}, y\sim U_{n_2}}[F(x,y)] = \mathop{\mathbf{E}}_{x\sim U_{n_1}}\left[\mathop{\mathbf{E}}_{y\sim U_{n_2}}[F_x(y)]\right]$$

where $F_x(\cdot)$ is the **CNF** F when the variables X are fixed to x. Note that $F_x(\cdot)$ is in itself a **CNF** of size at most m on n_2 variables. By assumption, for all values of x,

$$\mathop{\mathbf{E}}_{y\sim U_{n_2}}[F_x(y)] = \mathop{\mathbf{E}}_{y\sim D_2}[F_x(y)] \pm \varepsilon_2 \,. \tag{9}$$

Similarly for any fixed assignment $Y = y$, we have

$$\mathop{\mathbf{E}}_{x\sim U_{n_1}}[F(x,y)] = \mathop{\mathbf{E}}_{x\sim D_1}[F(x,y)] \pm \varepsilon_1 \,. \tag{10}$$

Combining Eqs. (9) and (10) gives

$$\begin{aligned}
\mathop{\mathbf{E}}_{x\sim U_{n_1}, y\sim U_{n_2}}[F(x,y)] &= \mathop{\mathbf{E}}_{x\sim U_{n_1}}\left[\mathop{\mathbf{E}}_{y\sim D_2}[F_x(y)] \pm \varepsilon_2\right] \\
&= \mathop{\mathbf{E}}_{y\sim D_2}\left[\mathop{\mathbf{E}}_{x\sim U_{n_1}}[F(x,y)] \pm \varepsilon_2\right] \\
&= \mathop{\mathbf{E}}_{y\sim D_2}\left[\mathop{\mathbf{E}}_{x\sim D_1}[F(x,y)] \pm \varepsilon_1 \pm \varepsilon_2\right] \\
&= \mathop{\mathbf{E}}_{x\sim D_1, y\sim D_2}[F(x,y)] \pm (\varepsilon_1 + \varepsilon_2) \,.
\end{aligned}$$

□

[9] Actually, Trevisan and Xue show the stronger result where $\mathrm{DT}(f)$ is replaced by the depth of the canonical decision tree for f (see Section 7 for its definition). However, we do not benefit from this strengthening.

By induction, Fact D.2 implies the following corollary.

Corollary D.3. *Let D be a distribution over $\{0,1\}^n$ that ε-fools* **CNFs** *of size m. Let $t \in N$, and $D^{\otimes t}$ be the distribution over $\{0,1\}^{n \cdot t}$ sampled by concatenating t independent samples from D. Then, $D^{\otimes t}$ is a distribution that $(\varepsilon \cdot t)$-fools* **CNFs** *of size m on $n \cdot t$ variables.*

In the following theorems we shall assume that the circuit size M is larger than the length of the input n.

Theorem D.4 ([TX13, Thm. 11, "Derandomized Switching Lemma for $\mathbf{AC^0}$", restated])**.** *Let C be circuit on n variables with size M, depth d and a top* OR*-gate. Let $p = 2^{-q}$, where $q \in \mathbb{N}$, and $s \in \mathbb{N}$ be some positive parameter. Assume that there exists a pseudorandom generator $\mathcal{G}$ with seed length r that ε_0-fools* **CNFs** *of size $M \cdot 2^s \cdot 2^{s \cdot (q+1)}$. Then, there exists a pseudorandom selection generator $\mathcal{G}_0$ of seed length $(d-1) \cdot r$ such that:*

- $\mathbf{Pr}_{\theta \sim \mathcal{G}_0, \beta \sim \mathcal{U}}\left[F|_{\theta,\beta} \text{ is not an } s\text{-}\mathbf{DNF} \text{ of size } \leq M \cdot 2^s\right]$
  $\leq M \cdot \left(2^{2s+1} \cdot \max\{(5ps)^s, (5/64)^s\} + \varepsilon_0 \cdot 2^{(s+1)\cdot(3s+\log M)}\right)$.
- *For each set of variables $T \subseteq [n]$, the probability that all variables in T are fixed is at most $(1 - p^{d-2}/64)^{|T|} + \varepsilon_0 \cdot (d-1)$.*

Proof. We shall start by adding a dummy layer next to the inputs that transforms the circuit C into a circuit C' of size at most $M \cdot n$, depth $d+1$ and bottom fan-in 1. We construct $\mathcal{G}_0$ by running $(d-1)$ iterative pseudorandom selections, using the generator of [DETT10] in each iteration. By Fact D.2, the pair (θ, β) obtained by sampling $\theta \sim \mathcal{G}$ and $\beta \in \{0,1\}^n$ uniformly at random, ε_0-fools **CNFs** of size $M \cdot 2^s \cdot 2^{s \cdot (q+1)}$. We denote by $M_1, \ldots, M_d$ the number of gates in the original circuit C at distance $1, \ldots, d$ from the inputs, respectively.

The first iteration. For the first iteration of Lemma D.1, we pick $p_0 = 1/64$ and $t = 1$. The probability that under the pseudorandom restriction, one of the gates at distance 2 from the inputs cannot be computed by a decision tree of depth s is at most

$$M_1 \cdot \left(2^{s+1+1} \cdot (5/64)^s + \varepsilon_0 \cdot 2^{(s+1)(2+\log M)}\right) .$$

In the complement event, we may express each gate at distance 2 from the inputs both as an s-**DNF** and as an s-**CNF**, so we can collapse this layer with the layer above it. This simplification yields a circuit of depth d, fan-in s and does not introduce new gates at distance 2 or more from the inputs. The number of gates at distance 1 from the inputs is at most $M \cdot 2^s$, since each depth-s decision tree is an s-**DNF** of size at most 2^s (and similarly an s-**CNF** of size at most 2^s).

The other $d-2$ iterations. At iteration $i = 2, \ldots, d-1$, we apply Lemma D.1 with $t = s$ and $p_0 = p$. We get that under the pseudorandom restriction, the probability that there exists a gate at distance 2 from the inputs that cannot be computed by a decision tree of depth s is at most

$$M_i \cdot \left(2^{s+s+1} \cdot (5ps)^s + \varepsilon_0 \cdot 2^{(s+1)(2s+\log(M \cdot 2^s))}\right) .$$

We are using the fact that each gate at distance 2 from the inputs computes a **CNF**/**DNF** of size at most $M \cdot 2^s$ and bottom fan-in s, an invariant that is preserved during the iterative process. Again, if a gate is computed by a decision tree of depth s then it is also computed by an s-**CNF** and by an s-**DNF** of size at most 2^s, and we may collapse the layers at distances 2 and 3 from the inputs.

Overall, after $(d-1)$-iterations, with probability at least

$$1 - M \cdot \left(2^{s+s+1} \cdot \max\{(5ps)^s, (5/64)^s\} + \varepsilon_0 \cdot 2^{(s+1)(3s+\log M)}\right),$$

all "switchings" were successful and we got a circuit of depth-2, size at most $M \cdot 2^s$, bottom fan-in s and a top OR gate, i.e. we got an s-**DNF**.

As for the second item of the theorem, observe that θ is selected by sampling $d-1$ binary strings from $\mathcal{G}$: one string w_1 of length $n \cdot \log_2(64)$ (consisting of n blocks of $\log_2(64)$ bits each) and $(d-2)$ strings, $w_2, \ldots, w_{d-1}$, of length $n \cdot q$ (consisting of n blocks of q bits each). We denote the concatenation of these $d-1$ strings by w. The i-th bit in θ is fixed (i.e. $\theta_i = \square$) iff the i-th block in one of the $d-1$ strings contains a zero. Thus, the event $\theta(i) = \square$ may be expressed as an OR of $6+(d-2)q$ literals over w. The event that a set of T variables is fixed may be expressed as a **CNF** of size $|T| \leq n$ in the bits of w. By Corollary D.3, the distribution of w is $\varepsilon_0 \cdot (d-1)$-pseudorandom for **CNFs** of size at most $M \cdot 2^s \cdot 2^{s(q+1)}$ and in particular to **CNFs** of size at most n. Thus,

$$\begin{aligned}\Pr_{w \text{ pseudo-random}}[T \text{ is fixed under } w] &\leq \Pr_{w \text{ random}}[T \text{ is fixed under } w] + \varepsilon_0 \cdot (d-1) \\ &= (1 - p^{d-2}/64)^{|T|} + \varepsilon_0 \cdot (d-1) \,. \qquad \square\end{aligned}$$

Theorem D.5 ([TX13, Theorem 12, restated slightly])**.** *Let C be a size M, depth d circuit, and $\varepsilon > 0$. Then, there exists a pseudorandom generator $\mathcal{G}_1$ of seed length $\tilde{O}(\log^3(M/\varepsilon))$ such that:*

- $|\mathbf{Pr}_{\rho \sim \mathcal{G}_1, x \sim U_n}[C_\rho(x) = 1] - \mathbf{Pr}_{y \sim U_n}[C(y) = 1]| < \varepsilon$.
- *Let p be the largest power of $1/2$ less than $1/(64\log(8M/\varepsilon))$. Then, each set of variables $T \subseteq [n]$ has probability at most $(1 - p^{d-2}/64)^{|T|} + \varepsilon_0 \cdot (d-1)$ of being unassigned by ρ.*

Proof. We initiate the generator from Thm. D.4 based on the generator from Thm. C.6. We choose parameters so that the bound we get from Thm. D.4 is at most $\varepsilon/2$. Choosing s to be a power of 2 between $\log(8M/\varepsilon)$ to $2\log(8M/\varepsilon)$, $p = 1/64s$ and $\varepsilon_0 = 2^{-9s^2}$ guarantees that

$$M \cdot (2^{2s+1} \cdot \max\{(5ps)^s, (5/64)^s\} + \varepsilon_0 \cdot 2^{(s+1)\cdot(3s+\log M)}) \leq \varepsilon/2 \,.$$

The choice also guarantees that $\varepsilon_0 \leq \varepsilon/2$. In order to apply Thm. D.4 with these parameters, the generator $\mathcal{G}$ in Thm. C.6 should ε_0-fool circuits of size $M' = M \cdot 2^{(q+2)s} = 2^{O(\log(M/\varepsilon)\log\log(M/\varepsilon))}$. Theorem C.6 guarantees that seed-length $r = \tilde{O}(\log(M') \cdot \log(M'/\varepsilon_0)) = \tilde{O}(\log^3(M/\varepsilon))$ is enough.

The generator in Thm. D.4, $\mathcal{G}_0$, selects a set of coordinates $J = \{i \in [n] : \theta(i) = *\}$. Thm. D.4 guarantees that with probability at least $(1 - \varepsilon/2)$ over the choice of J and the restriction of J^c by random bits, C reduces to an s-**DNF** of size at most $M \cdot 2^s$. We then assign values to the variables indexed by J according to De et al. generator, $\mathcal{G}$. Overall, we

need seed-length $(d-1)\cdot r + r = dr$, where the first term comes from sampling from $\mathcal{G}_0$ and the second from sampling according to $\mathcal{G}$.

For any fixed choice of θ we have:

$$\mathop{\mathbf{Pr}}_{y\sim U_n}[C(y)=1] = \mathop{\mathbf{Pr}}_{x\sim U_J, z\sim U_{J^c}}[C(x,z)=1] = \mathop{\mathbf{E}}_{z}[\mathop{\mathbf{Pr}}_{x}[C(x,z)=1]]$$

Hence also for $\mathcal{G}_0$ which is a distribution over selections θ the following holds

$$\mathop{\mathbf{Pr}}_{y\sim U_n}[C(y)=1] = \mathop{\mathbf{E}}_{\theta}\mathop{\mathbf{E}}_{z}[\mathop{\mathbf{Pr}}_{x}[C(x,z)=1]]$$

For choices (θ, z) such that $C_z(x) := C(x,z)$ is an s-**DNF** of size at most $M\cdot 2^s$, we have $\mathbf{Pr}_{x\sim U_J}[C(x,z)=1] = \mathbf{Pr}_{x\sim\mathcal{G}}[C(x,z)=1]\pm\varepsilon_0$. In the case where $C_z(x)$ is not an s-**DNF** of size $M\cdot 2^s$ we trivially have $\mathbf{Pr}_{x\sim U_J}[C(x,z)=1] = \mathbf{Pr}_{x\sim\mathcal{G}}[C(x,z)=1]\pm 1$. However, this is a rare event that happens with probability at most $\varepsilon/2$. Overall, we have

$$\mathop{\mathbf{Pr}}_{y\sim U_n}[C(y)=1] = \mathop{\mathbf{E}}_{\theta}\mathop{\mathbf{E}}_{z}[\mathop{\mathbf{Pr}}_{x\sim U_J}[C(x,z)=1]] = \mathop{\mathbf{E}}_{\theta}\mathop{\mathbf{E}}_{z}[\mathop{\mathbf{Pr}}_{x\sim\mathcal{G}}[C(x,z)=1]] \pm (\varepsilon_0+\varepsilon/2)\ .$$

which completes the proof of the first item as $\varepsilon_0 \le \varepsilon/2$.

Note that the generator $\mathcal{G}_1$ selects J and assigns values to the variables in J. It does not assign any of the variables in J^c. In this way, Trevisan and Xue change roles between the fixed and alive parts of the restriction: starting with pseudorandom restriction where J^c is fixed randomly and J is kept alive, they end up with a pseudorandom restriction where J is fixed pseudorandomly and J^c is kept alive.

The second item follows by observing that a set of variables is fixed in Thm. D.4 iff it is unassigned here. □

Theorem D.6 ([TX13, Theorem 13, improved])**.** *For every M, d, n, ε there is a polynomial time computable ε-pseudorandom generator for circuits of size M and depth d on n variables, whose seed length is $\tilde{O}(\log^{d+1}(M/\varepsilon)\cdot\log n)$.*

Proof. If $d\le 2$ we apply Thm. C.6. Otherwise, we may assume $d\ge 3$. As in Thm. D.5, let p be the largest power of $1/2$ which is smaller than $1/(64\log(8M/\varepsilon))$. Let $p' = p^{d-2}/64$. The theorem follows by applying $R = 3\ln(n)/p'$ independent random restrictions from $\mathcal{G}_1$, each with parameter $\varepsilon/2R$. Let $t = \log(2/\varepsilon)$, and let $T\subseteq[n]$ be a set of size t. The probability T remains totally unfixed after R iterations is at most

$$((1-p')^t + \varepsilon_0\cdot(d-1))^{3\ln(n)/p'}$$

where recall that $\varepsilon_0 = 2^{-\Omega(\log^2(M/\varepsilon))}$. We have $(1-p')^t > 1-p't \ge 1-\frac{\log(2/\varepsilon)}{64\log(8/\varepsilon)} > 1/2$, so

$$(1-p')^t + \varepsilon_0\cdot(d-1) \le (1-p')^t\cdot(1+2\varepsilon_0\cdot(d-1))\ ,$$

and we get

$$\begin{aligned}((1-p')^t + \varepsilon_0\cdot(d-1))^{3\ln(n)/p'} &< \left((1-p')^t\cdot(1+2\varepsilon_0\cdot(d-1))\right)^{3\ln(n)/p'} \\ &\le e^{(-p't)\cdot 3\ln(n)/p'}\cdot e^{2\varepsilon_0\cdot(d-1)\cdot 3\ln(n)/p'} \qquad (1+x\le e^x) \\ &= n^{-3t}\cdot e^{o(1)} \ = O(n^{-3t})\ .\end{aligned}$$

As there are only at most n^t sets T of size t, applying union bound, with probability at most $O(n^{-2t}) < \varepsilon/2$ there exists a set of size t which is unassigned. In the complement event, there are less than t variables alive, and we may sample from a t-wise independent distribution to fool the remaining circuit. Overall the seed length is

$$\left(\frac{3\ln(n)}{p'} \cdot \tilde{O}(\log^3(M/\varepsilon)\right) + O\left(\log(1/\varepsilon) \cdot \log n\right)) = \tilde{O}(\log^{d+1}(M/\varepsilon) \cdot \log n) \; . \qquad \square$$

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il