Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 174 (2014)

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

Avishay Tal*

May 7, 2015

Abstract

We show that $\mathbf{AC^0}$ circuits of depth d and size m have at most $2^{-\Omega(k/(\log m)^{d-1})}$ 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 (CACM, 2011). Braverman showed that any $r(m, d, \epsilon)$-wise independent distribution ϵ-fools $\mathbf{AC^0}$ circuits of size m and depth d, for

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

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

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

In contrast, an example by Mansour (appearing in Luby and Velickovic - Algorithmica, 1996) shows that there is a $\log(m)^{d-1} \cdot \log(1/\epsilon)$-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

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.

The study of bounded depth circuits (where d is constant) was one of the most exciting areas in computational complexity in the 1980s. Perhaps the most notable achievement was the tight $\exp(n^{1/(d-1)})$ size lower bound for the parity function in this model, proven by Håstad [Hås86], following the work of [Ajt83], [FSS84] and [Yao85].[1] Håstad introduced the switching lemma, which uses random restrictions to decrease the depth of $\mathbf{AC^0}$ circuits by one. The main idea was the following - $\mathbf{AC^0}$ circuits with size m and depth d become constant w.h.p. 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 for $m = \exp(o(n^{1/(d-1)}))$.

In an inspiring paper, Linial, Mansour and Nisan [LMN93] showed that $\mathbf{AC^0}$ circuits can be learned in quasipolynomial time, $n^{\log(n)^d}$, 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 $O(\log(n)^d)$. Boppana [Bop97] improved their bound on the degree to $O(\log(n)^{d-1})$, which is optimal for constant error. The existence of an approximating low degree polynomial implies a learning algorithm from random examples. For polynomial size **DNFs** (depth 2 circuits), Mansour [Man95] showed that only $n^{O(\log\log n)}$ out of the $\binom{n}{\leq \log n}$ monomials are needed to approximate the **DNF**, and achieved a $n^{O(\log\log n)}$ time learning algorithm 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 $\mathbf{AC^0}$ circuits. They showed that for a 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 weight at level at least k of f*. This was improved by Håstad [Hås01] to

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

which is tight for $k \leq \log(m)^d$, however not for larger values of k. Indeed, quite recently Håstad [Hås14] and Impagliazzo, Matthews and Paturi [IMP12] showed that any $\mathbf{AC^0}$ circuit f agrees with parity on at most a $1/2 + 2^{-\Omega(n/\log(m)^{d-1})}$ fraction of the inputs, i.e., $\hat{f}([n]) \leq 2^{-\Omega(n/\log(m)^{d-1})}$.

1.1 Our Results

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

Theorem 1.1 (Main Theorem)**.** *Let f be an $\mathbf{AC^0}$ circuit with depth d and size m, then*

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

[1] Lower bounds for the **DNF**-size of the parity function were long known [Lup61] and are much less involved.

A few things to note. Increasing k from 0 to n, the first time that Theorem 1.1 is meaningful is at $k = O(\log(m)^{d-1})$, which is roughly the same as in [LMN93] (and exactly the same as in [Hås01]). However, for values larger than this threshold, the decay in our bound is much faster, and in particular for $m = \text{poly}(n)$ we get $2^{-n/\text{poly}\log(n)}$ at level $k = \Omega(n)$ as opposed to $2^{-\Omega((n/\log n)^{1/(d-1)})}$ by [Hås01]. In addition, while [Hås14] and [IMP12] give bounds on individual $\hat{f}(S)^2$, we give bounds on the sum of $\exp(n)$ such squares (e.g. for $k = n/2$).

We point out that the results of [Hås14], [IMP12] and ours are quite surprising considering that most proofs for $\mathbf{AC^0}$ circuits follow by induction on the depth d; reducing the depth by 1 in each step using Håstad's switching lemma. Our main theorem is equivalent to saying that degree $O(\log(m)^{d-1}\log(1/\epsilon))$ polynomials ϵ-approximates an $\mathbf{AC^0}$ circuit of size m and depth d, as opposed to $O(\log(m/\epsilon)^d)$ by [LMN93]. It seems at first glance that one must pay a factor of $\log(m/\epsilon)$ for each step in the induction to ensure error at most ϵ, giving degree at least $\log(m/\epsilon)^{d-1}$. However, Håstad and Impagliazzo et al. manage to avoid that. Håstad performs random restrictions, with parameter $p = 1/O(\log m)$ that does not depend on ϵ. This only guarantee that the switching will succeed with probability $1 - 1/\text{poly}(m)$, as opposed to probability of $1 - \epsilon/m$ in the original proof of [LMN93]. However, in the cases where the restrictions "fail", Håstad fixes D additional variables using a decision tree of depth D. Under these additional fixings, the probability that the switching does not succeed reduces to $m \cdot 2^{-D}$. We show that the parameters p and D translate into a multiplicative $1/p$ term and an additive D term in the degree, correspondingly. Choosing D to be roughly $\log(m/\epsilon)$ and applying induction gives the desired dependency on m and ϵ.

Theorem 1.1 shows that the Fourier tails above level k decreases exponentially in k. In Section 4 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 show that for any $\mathbf{AC^0}$ circuit of size m and depth d the spectral norm of f at level k is at most

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

In Section 5, we prove that Equation (1) implies the following two *known* results:

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

2. $\mathbf{AC^0}$ circuits cannot distinguish between fair coins and coins with bias at most $\frac{1}{O(\log(m)^{d-1})}$. 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 $\mathbf{AC^0}$ circuits. The most significant improvement is to the work of Braverman [Bra11] who proved a longstanding conjecture, showing that poly-logarithmic independent distributions fool (polynomial size) $\mathbf{AC^0}$ circuits. To be more precise, Braverman showed

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

that any k-wise independent distribution, where $k = \log(m/\epsilon)^{2d^2+7d+3}$, ϵ-fools circuits of size m and depth d. In addition, it was long known [LV96] that k must be larger than $\log(m)^{d-1}\log(1/\epsilon)$; 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 $k = \log(m/\epsilon)^{3d+3}$, which comes much closer to matching the lower bound from [LV96]. 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 calculated by $\mathbf{AC^0}$ circuits of depth $O(\log(n)/\log\log(n))$ and polynomial size, giving a non trivial PRG for $d = O(\log(n)/\log\log(n))$ 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.

Conjecture 1.2. *Any k-wise independence ϵ-fools circuits of size m and depth d, for*

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

k-wise independence fools DNFs. We improve in Section 7 the earlier result of Bazzi [Baz09], who showed that $\log(m/\epsilon)^2$-wise independence fools **DNFs** of size m. We improve the dependence on ϵ and get that $\log(m)\log(m/\epsilon)$-wise suffices. Note that by [LV96] this is optimal for $\epsilon \leq 1/m^{\Omega(1)}$. The range $\epsilon \geq 1/m^{o(1)}$ is still not tightly understood.

PRGs for $\mathbf{AC^0}$ and DNFs. We improve the results of De et al. [DETT10] and of Trevisan and Xue [TX13] which gives the best known PRGs for **DNFs** and $\mathbf{AC^0}$ circuits respectively. We leave verifying the details here to the reader.

Sparse polynomial approximations of $\mathbf{AC^0}$ circuits. We show in Corollary 4.8 that any $\mathbf{AC^0}$ circuit f of size m and depth d can be ϵ-approximated in L_2 by a polynomial $p(x)$ of sparsity $\log(m)^{O(\log(m)^{d-1}\log(1/\epsilon))}$, improving the results of [LMN93] and [Man95]. As the inner product on $k = \log(m)^{d-1}$ 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 $2^{o(\log(m)^{d-1})}$.

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

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 $\|f\|_p = \left(\mathbf{E}_{x\in\{-1,1\}^n}[|f(x)|^p]\right)^{1/p}$.

Theorem 2.1 (The generalized binomial theorem)**.** *Let $|x| < 1$, and $k \in \mathbb{N}$, then $\sum_{n=0}^{\infty}\binom{k+n-1}{k-1}\cdot x^n = \frac{1}{(1-x)^k}$.*

Multiplying both sides by x^k one get the following corollary.

Corollary 2.2. *Let $|x| < 1$, and $k \in \mathbb{N}$ then $\sum_{d=k}^{\infty}\binom{d-1}{k-1}\cdot x^d = \frac{x^k}{(1-x)^k}$.*

2.1 Restrictions

Definition 2.3 (Restriction)**.** *Let $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean function. A restriction ρ is a vector of length n of elements from $\{-1,1,*\}$. We denote by $f|_\rho : \{-1,1\}^n \to \{-1,1\}$ the function f restricted according to ρ, defined by*

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

Task	Ref.	Bound
k-wise ind. fooling **DNFs**	[Baz09]	$k = O(\log(m/\epsilon)^2)$
	This Work	$k = O(\log(m/\epsilon)\log(m))$
	Lower Bound	$k \geq \log(m)\log(1/\epsilon)$
k-wise ind. fooling $\mathbf{AC^0}$	[Bra11]	$k = O(\log(m/\epsilon)^{d^2+3d}\log(m)^{d^2+4d+3})$
	This Work	$k = O(\log(m/\epsilon)^{d}\log(m)^{2d+3})$
	Lower Bound	$k \geq \log(m)^{d-1}\log(1/\epsilon)$
sparse polynomial approximating **DNFs** in L_2	[Man95]	sparsity $= (m/\epsilon)^{O(\log\log(m/\epsilon)\log(1/\epsilon))}$
	This Work	sparsity $= m^{O(\log\log(m)\log(1/\epsilon))}$
sparse polynomial approximating $\mathbf{AC^0}$ in L_2	[LMN93]	sparsity $= 2^{O(\log(n)\log(m/\epsilon)^d)}$
	[Hås01]	sparsity $= 2^{O(\log(n)\log(m/\epsilon)^{d-2}\log(m)\log(1/\epsilon))}$
	This Work	sparsity $= 2^{O(\log\log(m)\log(m)^{d-1}\log(1/\epsilon))}$
	Lower Bound	sparsity $\geq 2^{\Omega(\log(m)^{d-1})}$
PRGs for **DNFs**	[DETT10]	seed $= O(\log n + \log^2(m/\epsilon)\log\log(m/\epsilon))$
	This Work	seed $= O(\log n + \log(m/\epsilon)\log(m)\log\log m)$
PRGs for $\mathbf{AC^0}$	[TX13]	seed $= \widetilde{O}(\log(m/\epsilon)^{d+4})$
	This Work	seed $= \widetilde{O}(\log(m/\epsilon)^{d+3})$

Figure 1: Summary of Applications

When fixing only one bit to a constant, we will denote the restricted function by $f|_{x_i=b}$.

Definition 2.4 (p-Random Restriction)**.** *A p-random restriction is a restriction as in Definition 2.3 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.

Fact 2.5 (Proposition 4.17, [O'D14])**.** *Let $f : \{-1,1\}^n \to \mathbb{R}$, let $S \subseteq [n]$, and $p > 0$, then*

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

and

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

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

Fact 2.6. *Denote by* $\mathrm{Bin}(k,p)$ *a binomial random variable with parameters k, 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 2.7 (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 $L_{1,k}(f) \triangleq \sum_{S:|S|=k} |\hat{f}(S)|$.*

Fact 2.8 (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. $L_1(f) \leq 2^d$.

2.3 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 2.9 (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 2.10 (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 implies that D_T is well defined, i.e. that the function $D_T f$ does not depend on the order of indices we chose, and gives equivalent formulations for the function $D_T f$.

Claim 2.11.

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

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.

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 multiple of $2^{-|T|}$. This holds 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 2.12.

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

Definition 2.13 (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 2.12 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 2.14. $\mathrm{Inf}^k(f) = \sum_{d\geq k} \mathbf{W}^{\geq d}[f] \cdot \binom{d-1}{k-1}$ *for all $k \in \mathbb{N}$.*

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

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

□

3 Exponentially Small Tails for $\mathbf{AC^0}$

We generalize the proof of Håstad ([Hås14]), who showed that the correlation between the parity function and any $\mathbf{AC^0}$ circuit of depth d and size m is at most $2^{-\Omega(n/\log(m)^{d-1})}$. This bound is tight up to constants in the exponent, as shown by an example in [Hås14], and improves upon previous bounds from [LMN93, Hås01].

We will use two simple Lemmas which explains the behaviour 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 $p \in [0,1]$*

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

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

$$\begin{aligned}
\mathop{\mathbf{E}}_{\rho\sim\mathcal{R}_p}\left[\mathbf{W}^{\geq \lfloor kp \rfloor}[f|_\rho]\right] &= \sum_{\ell \geq \lfloor kp \rfloor} \mathbf{W}^{\ell}[f] \cdot \mathbf{Pr}[\mathrm{Bin}(\ell,p) \geq \lfloor kp \rfloor] && \text{(Fact 2.6)}\\
&\geq \sum_{\ell \geq k} \mathbf{W}^{\ell}[f] \cdot \mathbf{Pr}[\mathrm{Bin}(\ell,p) \geq \lfloor kp \rfloor]\\
&\geq \sum_{\ell \geq k} \mathbf{W}^{\ell}[f] \cdot 1/2 && (\mathrm{median}(\mathrm{Bin}(\ell,p)) \geq \lfloor \ell p \rfloor \geq \lfloor kp \rfloor, \text{[KB80]})\\
&= 1/2 \cdot \mathbf{W}^{\geq k}[f]. && \square
\end{aligned}$$

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*

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

Proof. Rearranging the Fourier expression for f gives

$$f(x) = \sum_{S\subseteq[n]} \hat{f}(S) \prod_{j\in S} x_j = \sum_{S\subseteq[n]\setminus\{i\}} \left(\hat{f}(S) + \hat{f}(S\cup\{i\})\cdot x_i\right) \prod_{j\in S} x_j \ .$$

By the uniqueness of the Fourier transform we see that $\widehat{f|_{x_i=b}}(S) = \hat{f}(S) + b \cdot \hat{f}(S\cup\{i\})$ for $b \in \{-1,1\}$, and $S \subseteq [n] \setminus \{i\}$. Hence,

$$\widehat{f|_{x_i=1}}(S)^2 + \widehat{f|_{x_i=-1}}(S)^2 = 2\left(\hat{f}(S)^2 + \hat{f}(S\cup\{i\})^2\right) \ .$$

Summing over all sets $S \subseteq [n] \setminus \{i\}$ of size at least $k-1$ gives

$$\mathbf{W}^{\geq k-1}[f|_{x_i=1}] + \mathbf{W}^{\geq k-1}[f|_{x_i=-1}] = 2 \cdot \sum_{\substack{S\subseteq[n]\setminus\{i\}:\\ |S|\geq k-1}} \hat{f}(S)^2 + \hat{f}(S\cup\{i\})^2 \geq 2 \cdot \sum_{\substack{T\subseteq[n]:\\ |T|\geq k}} \hat{f}(T)^2 \ . \qquad \square$$

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 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 paths in the tree, then we have Fourier tail bounds for the unrestricted function. 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 $\le D$ such that for any leaf ℓ, under the corresponding restriction $\mathbf{W}^{\ge k}[f|_{\tau_\ell}] \le \epsilon$, then $\mathbf{W}^{\ge k+D}[f] \le \epsilon$.*

Proof. Apply induction on the depth of the restriction tree. For depth 0 this obviously holds. For depth D, consider both subtrees which 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 the each subtree to get $\mathbf{W}^{\ge k+(D-1)}[f|_{x_i=1}] \le \epsilon$ and $\mathbf{W}^{\ge k+(D-1)}[f|_{x_i=-1}] \le \epsilon$. Finally, applying Lemma 3.2 gives $\mathbf{W}^{\ge k+D}[f] \le \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$. □

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 $m \cdot (24pt)^D$.*

We are ready to prove the Fourier tail bounds for $\mathbf{AC^0}$. We define the *effective size* of an $\mathbf{AC^0}$ circuit as the number of gates in the circuit which are of distance at least 2 from the inputs. We assume the circuits have alternating gates at odd and even depths.

Theorem 3.7. *Let f be an $\mathbf{AC^0}$ circuit of depth d, effective size m, and bottom fan-in t, then $\mathbf{W}^{\ge k}[f] \le 8^{d-1} \cdot 2^{-k/\left(20t(96\log(2m))^{d-2}\right)}$.*

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

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

For the induction step, we apply a p-random restriction with $p = 1/48t$. Consider the gates at distance 2 from the inputs: $f_1, \dots, f_m$. These gates compute functions given by depth-2 circuits with bottom fan-in $\le t$. Setting $D = \lfloor kp/2 \rfloor$ and using Lemma 3.6 gives that with probability at least $1 - m \cdot 2^{-D} \ge 1 - 2^{\log(m)-D}$ 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 $\mathbf{W}^{\ge k}[f] \le 2 \cdot \mathbf{E}_\rho[\mathbf{W}^{\ge \lfloor kp \rfloor}[f|_\rho]]$. Since $\mathbf{W}^{\ge \lfloor kp \rfloor}[f|_\rho]$ is a random variable bounded in $[0,1]$ we have

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

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

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

Proof of Claim 3.8. If $b \le a$ then the conclusion is trivial since $X \le 1 \le 2^{1-b/a}$. Otherwise $b > a \ge 1$ and we have $a - b \le (a-b)/a$, thus $X \le 2^{a-b} \le 2^{(a-b)/a}$ as required. □

We are left to analyze $\mathbf{E}[\mathbf{W}^{\ge \lfloor kp \rfloor}[f|_\rho] \mid \rho \text{ is good}]$. Fix some ρ which is good, we will bound $\mathbf{W}^{\ge \lfloor kp \rfloor}[f|_\rho]$ 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 $\le \log(2m)$, hence as a **CNF**/**DNF** formula of width $\log(2m)$. This means that applying the restriction $\rho \circ \tau_\ell$, the circuit f collapses to a depth $d-1$ $\mathbf{AC^0}$ circuit with bottom fan-in $t' \le \log(2m)$ and effective size at most m.[3] By the induction hypothesis, for any k' we have $\mathbf{W}^{\ge k'}[f|_\rho|_{\tau_\ell}] \le 8^{d-2} \cdot 2^{-\Omega(k'/(t' \log(2m)^{d-3}))}$. Setting $k' = \lfloor kp \rfloor - D \ge \lfloor kp/2 \rfloor \ge \frac{k}{96t} - 1$ and applying Lemma 3.4 we have

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

and

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

Theorem 3.9 (Theorem 1.1, restated)**.** *Let f be an $\mathbf{AC^0}$ circuit of depth d and size m, for $m > 1$, then $\mathbf{W}^{\ge k}[f] \le 2 \cdot e^{-k/(c_d \log(m)^{d-1})}$ where $c_d = \log_2(e) \cdot 3d \cdot 20 \cdot 96^{d-1} \cdot 2^{d-1} \le 230^d$.*

Proof. Let f be a function computed by an $\mathbf{AC^0}$ 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 $\mathbf{AC^0}$ circuit of depth $d+1$, effective size m and bottom fan-in 1. Plugging this into Theorem 3.7 gives $\mathbf{W}^{\ge k}[f] \le 2^{3d - k/(20 \cdot 96^{d-1} \cdot \log(2m)^{d-1})}$. Hence, by Claim 3.8 we get

$$\mathbf{W}^{\ge k}[f] \le 2 \cdot 2^{-k/(3d \cdot 20 \cdot 96^{d-1} \cdot \log(2m)^{d-1})} .$$

Changing the base of the exponent from 2 to e we get

$$\mathbf{W}^{\ge k}[f] \le 2 \cdot e^{-k/(\log_2(e) \cdot 3d \cdot 20 \cdot 96^{d-1} \cdot \log(2m)^{d-1})} \le 2 \cdot e^{-k/(\log_2(e) \cdot 3d \cdot 20 \cdot 96^{d-1} \cdot 2^{d-1} \cdot \log(m)^{d-1})},$$

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

4 Connections between Fourier Spectrum Attributes of Boolean Functions

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.

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

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

- **SLTP**: Switching lemma type property / degree shrinkage
$$\forall d, p : \mathop{\mathbf{Pr}}_{\rho \sim \mathcal{R}_p} [\deg(f|_\rho) = d] \leq O(pt)^d$$
- **L1**: Bounded spectral norm of the kth 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 Lemmas 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, we show the following connections:

$$\begin{tikzcd}
\mathbf{ESFT} \arrow[r, Leftrightarrow, "4.3", "4.4"'] \arrow[d, Leftrightarrow, "4.1"', "4.2"] & \mathbf{SLTP} \arrow[d, Rightarrow, "4.6"] \\
\mathbf{InfK} \arrow[r, Rightarrow, "4.5"] & \mathbf{L1}
\end{tikzcd}$$

We remark that Lemma 4.6 is due to Mansour [Man95], and Lemma 4.4 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 Lemmas 4.1, 4.2, 4.3, 4.4, 4.5, 4.6 more accurately and prove them.

Lemma 4.1. *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.*

Proof. We shall prove for $C = 1$, the proof generalizes for all C. Denote $a := e^{-1/t}$. Using Lemma 2.14 we bound the total degree-k influence:
$$\mathrm{Inf}^k(f) = \sum_{d \geq k} \mathbf{W}^{\geq d}[f] \cdot \binom{d-1}{k-1} \leq \sum_{d \geq k} a^d \cdot \binom{d-1}{k-1}$$
Using Corollary 2.2 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.

Lemma 4.2. *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 2.14, $\mathbf{W}^{\geq d}[f] \cdot \binom{d-1}{k-1} \leq \mathrm{Inf}^k[f] \leq t^k$. 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, the following relation was established.

Lemma 4.3 ([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 2.8.

We are left to prove Equation (3). Using Fact 2.6, 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 4.1. $\square$

Lemma 4.4 ([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

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

Lemma 4.5. *If f is Boolean, then $L_{1,k}[f] \leq 2^k \cdot \mathrm{Inf}^k[f]$.*

Proof. It is easy to see from Claim 2.11 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, and we have $\forall x : |D_T f(x)| \leq 2^{|T|}(D_T f(x))^2$. Hence,

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

Summing over all sets of size k completes the proof. $\square$

Remark: It is necessary that f is Boolean in Lemma 4.5, since otherwise we can have the function

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

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

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

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

Lemma 4.6 ([Man95]). *Let $t > 0$, and $f : \{-1,1\}^n \to \{-1,1\}$, if for all d, p, $\mathbf{Pr}_{\rho\sim\mathcal{R}_p}[\deg(f|_\rho) = d] \leq C(pt)^d$, then $\forall k : L_{1,k}[f] \leq 2C(4t)^k$.*

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 $L_{1,k}(f) \leq \frac{1}{p^k}\mathbf{E}_{\rho\sim\mathcal{R}_p}[L_{1,k}[f|_\rho]]$.

$$\begin{aligned} L_{1,k}[f] &= \sum_{S:|S|=k} |\hat{f}(S)| = \sum_{S:|S|=k} \left|\frac{1}{p^k} \mathop{\mathbf{E}}_{\rho\sim\mathcal{R}_p}\left[\widehat{f|_\rho}(S)\right]\right| && \text{(Fact 2.5)} \\ &\leq \sum_{S:|S|=k} \frac{1}{p^k} \mathop{\mathbf{E}}_{\rho\sim\mathcal{R}_p}\left[|\widehat{f|_\rho}(S)|\right] = \frac{1}{p^k} \mathop{\mathbf{E}}_{\rho\sim\mathcal{R}_p}\left[\sum_{S:|S|=k} |\widehat{f|_\rho}(S)|\right] \\ &= \frac{1}{p^k} \mathop{\mathbf{E}}_{\rho\sim\mathcal{R}_p}\left[L_{1,k}[f|_\rho]\right] . && (5) \end{aligned}$$

Next, we show that for $f : \{-1,1\}^n \to \{-1,1\}$, if there exists $t > 0$ such that for all d, p, $\mathbf{Pr}[\deg(f|_\rho) = d] \leq (pt)^d$, then $\mathbf{E}_{\rho\sim\mathcal{R}_p}[L_1[f|_\rho]] \leq 2$ for $p = 1/4t$. Conditioning on $\deg(f|_\rho) = d$ and using Fact 2.8, we have $L_1[f|_\rho] \leq 2^d$. Hence,

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

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 4.7. *Let $t > 0$ 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 $t^{O(t\log(1/\epsilon))}$ 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 $1-\epsilon$ of the Fourier mass of f, i.e. $\sum_{i=1}^r \hat{f}(S_i)^2 \geq 1-\epsilon$.

Proof. We shall prove for $C = 1$, the proof generalizes for all constant C. Let $w := t \cdot \ln(2/\epsilon)$. First it is enough to consider Fourier coefficients of sets of size $\leq w$, since all the sum of squares of Fourier coefficients of larger sets is at most $\epsilon/2$. Now $\sum_{S:|S|\leq w} |\hat{f}(S)| = \sum_{i=0}^{w} L_{1,i}[f]$. Using Lemmas 4.1 and 4.5 we get

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

Letting $\mathcal{F} = \{S : |S| \leq w, |\hat{f}(S)| \geq \frac{\epsilon/2}{t^w 2^{w+1}}\}$ 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 $\sum_{|S|>w} \hat{f}(S)^2 \leq \epsilon/2$. To bound the last term

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

Hence, $\sum_{S\in\mathcal{F}} \hat{f}(S)^2 \geq 1 - \epsilon$. It remain to figure out the size of $\mathcal{F}$. Since every coefficient in $\mathcal{F}$ contributes at least $\frac{\epsilon/2}{t^w 2^{w+1}}$ 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 $2(t^w 2^{w+1})^2/\epsilon = O(t)^{2t\ln(1/\epsilon)}$, which completes the proof. □

Immediate from Theorem 3.9, Lemmas 4.1, 4.3, 4.5, and 4.7 we get the following corollary.

Corollary 4.8. *Let f be computed by an $\mathbf{AC^0}$ circuit of depth d and at most m gates, then*

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

5 Short Proofs for Known Results

In this section we show how Corollary 4.8 gives simple proofs for two theorems. The first states that (almost) balanced symmetric functions, and in particular the Majority function, cannot be well approximated by a small $\mathbf{AC^0}$ circuit.

Theorem 5.1. *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*

$$c_d \log(m)^{d-1} \leq (n/100\ln(n))^{1/3}$$

then

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

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

$$|\mathop{\mathbf{E}}_{x}[f(x)g(x)]| \le \sum_{S} |\hat{f}(S)\hat{g}(S)| = |\hat{f}(\emptyset)\hat{g}(\emptyset)| + \sum_{k=1}^{\ell} \sum_{S:|S|=k} |\hat{f}(S)\hat{g}(S)| + \sum_{S:|S|>\ell} |\hat{f}(S)\hat{g}(S)| \,. \qquad (7)$$

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 3.7, and Parseval's identity ($\sum_{S:|S|>\ell} \hat{g}(S)^2 \le 1$), to get

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

Picking $\ell := \ln(n) \cdot c_d \log(m)^{d-1}$ 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

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

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

$$\frac{2c_d \log(m)^{d-1}}{\sqrt{n}} \sqrt{\frac{(k+1)^{k+1}}{k^k}} \le \frac{2c_d \log(m)^{d-1}}{\sqrt{n}} \sqrt{e \cdot (k+1)} \le \frac{2c_d \log(m)^{d-1}}{\sqrt{n}} \sqrt{e \cdot \ell} \le \frac{1}{2} \,,$$

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

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

We remark that although our proof is Fourier analytical, it differs from the standard argument that is used to bound the correlation of $\mathbf{AC^0}$ 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 vote function, and f to be a poly(n) size $\mathbf{AC^0}$ 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.

Theorem 5.2. *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 $\le 4c_d p \log(m)^{d-1}$.*

Proof. We can assume $pc_d \log(m)^{d-1} \le 1/4$, since otherwise the result is trivial. For $-1 \le p \le 1$, 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. its 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

$$\begin{aligned} \mathit{Distinguishability}(f) &= \left| \mathop{\mathbf{E}}_{x\sim U_n}[f(x)] - \mathop{\mathbf{E}}_{x\sim B(p,n)}[f(x)] \right| = \left| \hat{f}(\emptyset) - \sum_S \hat{f}(S)p^{|S|} \right| \\ &= \left| \sum_{S\neq\emptyset} \hat{f}(S)p^{|S|} \right| \le \sum_{k=1}^{n} p^k \left(2c_d \log(m)^{d-1}\right)^k \\ &\le p\left(2c_d\log(m)^{d-1}\right)\sum_{k=1}^{n} 2^{1-k} \le 2p\left(2c_d\log(m)^{d-1}\right) . \end{aligned}$$

□

6 Improving Braverman's Analysis

Definition 6.1. *Denote by* $\mathsf{tail}(m,d,k)$ *the maximal* $\mathbf{W}^{\ge k}[F]$ *over all* $\mathbf{AC^0}$ *circuits* F *of size* $\le m$ *and depth* $\le d$.

By Theorem 3.9, $\mathsf{tail}(m,d,k) \le 2\cdot e^{-k/(c_d\log(m)^{d-1})}$. For convenience, we use the cruder estimate $\mathsf{tail}(m,d,k) \le 2\cdot 2^{-k/(c_d\log(m)^{d-1})}$ in the next two sections. Braverman's Theorem can be rephrased as follows (we show that this is indeed the case in Appendix B).

Theorem 6.2 ([Bra11])**.** *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*

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

then

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

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

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

Theorem 6.3. $r(m,d,\epsilon)$*-independence* ϵ*-fools* $\mathbf{AC^0}$ *circuits of depth* d *and size* m, *where*

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

7 Improving Bazzi's Analysis

Bazzi [Baz09] showed that $O(\log(m/\epsilon)^2)$ independence fools $\mathbf{AC^0}$. We show that $O(\log(m/\epsilon)\log(m))$ independence suffices. For $\epsilon \le 1/m^{\Omega(1)}$ this bound is tight, due to the example of Mansour from [LV96].

Theorem 7.1 ([Baz09], [Raz09])**.** *Let* F *be a* **DNF** *with* m *terms. Let* t *be some parameter, then* F *is* $m^3\cdot\mathsf{tail}(m,2,(k-3t)/2) + m2^{-t}$ *fooled by any* k*-wise independence.*

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

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

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.

References

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

[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.

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

[Bra11] M. Braverman. Poly-logarithmic independence fools bounded-depth Boolean circuits. *Commun. ACM*, 54(4):108–115, 2011.

[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.

[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.

[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.

[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.

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

[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.

[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 2.11, 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 (Lemma 8, [Bra11]). *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 (Proposition 9, [Bra11]). *In Lemma B.1, for $s \geq \log(m)$, $\|f\|_\infty < (2m)^{\deg(f)-2} = (2m)^{(s\log(m))^d - 2}$*

Lemma B.3 (Rephrasing of Lemma 10, [Bra11]). *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

$$\Pr_{\mu}[\mathcal{E}_\nu = 1], \Pr_{U_n}[\mathcal{E}_\nu = 1] \le 2 \cdot \Pr_{\nu}[\mathcal{E}_\nu = 1] \le 2 \cdot 0.82^{s_1} m \ .$$

Let f be the degree $(s_1 \log(m))^d$ 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

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

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 (Rephrasing of Main Theorem, [Bra11]). *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*

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

then

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

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

Proof of Theorem B.4. Denote by $\epsilon_1 := 0.82^{s_1} \cdot (2m)$ and

$$\epsilon_2 := 0.82^{s_1} \cdot (4m) + 2^{4(s_1 \cdot \log m)^d \log m} \cdot \mathrm{tail}(m^3, d+3, s_2) \ .$$

Applying Lemma B.3 with parameters s_1 and s_2 gives

$$\|F' - f'\|_2^2 \le \epsilon_2 \ .$$

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

$$\mathbf{E}[F'(x) - f'_\ell(x)] \le \|F' - f'\|_2^2 \le \epsilon_2 \ .$$

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

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

$$\begin{aligned}\mathop{\mathbf{E}}_{\mu}[F(x)] &\geq \mathop{\mathbf{E}}_{\mu}[F'(x)] - \epsilon_1 \geq \mathop{\mathbf{E}}_{\mu}[f'_\ell(x)] - \epsilon_1 =^* \mathbf{E}[f'_\ell(x)] - \epsilon_1 \\ &= \mathbf{E}[F'(x)] - \mathbf{E}[F'(x) - f'_\ell(x)] - \epsilon_1 \geq \mathbf{E}[F'(x)] - \epsilon_2 - \epsilon_1 \geq \mathbf{E}[F(x)] - \epsilon_2 - \epsilon_1\end{aligned}$$

where we used in * the fact that $\deg(f'_\ell) \leq 2(s_2 + (s_1 \log(m))^d)$ and μ is $\deg(f'_\ell)$-wise independent. In a similar way, one can show $\mathbf{E}_\mu[F(x)] \leq \mathbf{E}[F(x)] + \epsilon_1 + \epsilon_2$. Combining both cases we get

$$|\mathop{\mathbf{E}}_{\mu}[F] - E[F]| \leq \epsilon_1 + \epsilon_2 = \epsilon(s_1, s_2, d) \ . \qquad \Box$$

ECCC ISSN 1433-8092

http://eccc.hpi-web.de