Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 48 (2014)

Shrinkage of de Morgan Formulae by Spectral Techniques

Avishay Tal*

Abstract

We give a new and improved proof that the shrinkage exponent of de Morgan formulae is 2. Namely, we show that for any Boolean function $f : \{0,1\}^n \to \{0,1\}$, setting each variable out of $x_1, \dots, x_n$ with probability $1-p$ to a randomly chosen constant, reduces the expected formula size of the function by a factor of $O(p^2)$. This result is tight and improves the work of Håstad [Hås98] by removing logarithmic factors.

As a consequence of our results, the function defined by Andreev [And87], $A : \{0,1\}^n \to \{0,1\}$, which is in **P**, has formula size at least $\Omega(\frac{n^3}{\log^2 n \log\log n})$. This lower bound is tight (for the function A) up to the $\log\log n$ factor, and is the best known lower bound for functions in **P**. In addition, we strengthen the average-case hardness result of Komargodski et al. [KRT13]; we show that the functions defined in [KRT13], $h_r : \{0,1\}^n \to \{0,1\}$, which are also in **P**, cannot be computed correctly on a fraction greater than $1/2+2^{-r}$ of the inputs, by de Morgan formulae of size at most $\frac{n^3}{r^2 \mathrm{poly}\log n}$, for any parameter $r \le n^{1/3}$.

The proof relies on a result from quantum query complexity by [LLS06, HLS07, Rei11]: for any Boolean function f, $Q_2(f) \le O(\sqrt{L(f)})$, where $Q_2(f)$ is the bounded-error quantum query complexity of f, and $L(f)$ is the minimal size de Morgan formula computing f.

*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

The problem of **P** vs. $\mathbf{NC^1}$ is a major open-problem in computational complexity. It asks whether any function computable by a polynomial time Turing machine can also be computed by a formula of polynomial size. A *de Morgan formula* is a binary tree in which each leaf is labeled with a literal from $\{x_1, \dots, x_n, \neg x_1, \dots, \neg x_n\}$ and each internal node is labeled with either a Boolean AND or OR gate. Such a tree naturally describes a Boolean function on n variables by propagating values from leaves to root, and returning the root's value. The *formula size* is the number of leaves in the tree; for a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$[1] we denote by $L(f)$ the minimal size formula which computes f. Showing that some language in **P** requires formulae of super-polynomial size would imply that $\mathbf{P} \not\subseteq \mathbf{NC^1}$. [2]

Showing super-polynomial formula size lower bounds for problems in **P** would be a major breakthrough in complexity theory, and such lower bounds are not even known for **NEXP**. However, lower bounds of the form $\Omega(n^c)$, for a fixed constant c, were achieved during the years for problems in **P**. This line of research began with the work of Subbotovskaya [Sub61] who gave an $\Omega(n^{1.5})$ lower bound for the parity function. Subbotovskaya introduced the technique of random restrictions in her proof; a method which was applied successfully to solve other problems such as giving lower bounds for $\mathbf{AC^0}$. Subbotovskaya showed that the minimal formula size of a given function is shrunk, on expectation, by a factor of $O(p^{1.5})$ under *p-random restrictions*. These are restrictions to the function variables keeping each variable "alive" with probability p (independently of other choices) and fixing it to a uniformly chosen random bit otherwise. We denote the distribution of p-random restrictions by $\mathcal{R}_p$; If $\rho \sim \mathcal{R}_p$, then $f|_\rho$ denotes the restriction of the function f by ρ. Since the parity function does not become constant after fixing less than all of its input bits, this implies that its size is at least $\Omega(n^{1.5})$. Khrapchenko [Khr71] used a different method to give a tight $\Omega(n^2)$ lower bound for the parity function. Andreev [And87] constructed a function in **P** and showed that its formula size is at least $\Omega(n^{2.5-o(1)})$. In fact, he got a lower bound of $\Omega(n^{1+\Gamma-o(1)})$ where Γ is the *shrinkage exponent* of de Morgan formulae - the maximal constant such that any de Morgan formula is shrunk by a factor of $O(p^\Gamma)$ under p-random restrictions. Impagliazzo and Nisan [IN93] showed that $\Gamma \geq 1.55$; Paterson and Zwick [PZ93] improved this bound to $\Gamma \geq 1.63$; and finally Håstad [Hås98] showed that $\Gamma \geq 2 - o(1)$. More precisely, Håstad proved the following result.

Theorem 1.1 ([Hås98])**.** *Let f be a Boolean function. For every $p > 0$,*

$$\mathop{\mathbf{E}}_{\rho \sim \mathcal{R}_p} [L(f|_\rho)] \leq O\left(p^2 \left(1 + \log^{3/2} \min\left\{\frac{1}{p}, L(f)\right\}\right) L(f) + p\sqrt{L(f)}\right).$$

This result is essentially tight up to the logarithmic terms as exhibited by the parity function. The formula size of the parity function of n variables is $\Theta(n^2)$ (see [Yab54, Khr71]). Applying a p-random restriction on the parity function yields a smaller parity function (or its negation) on k variables where $k \sim \mathrm{Bin}(n,p)$. By Khrapchenko's argument, the formula size of the restricted function is $\geq k^2$, thus the expected formula size is at least $\mathbf{E}[k^2] = p^2n^2 + p(1-p)n = \Omega\left(p^2 L(f) + p\sqrt{L(f)}\right)$.

Other efforts have been made to give a function in **P** that requires super-polynomial formula size: Karchmer, Raz and Wigderson [KRW95] suggested a function in **P** that might require super-polynomial formula size. Recently, Gavinsky et al. [GMWW14] suggested an information theoretical approach to further understand the formula size of this function.

[1] We identify the truth values **true** and **false** with -1 and 1 respectively.

[2] Here we think of the non-uniform version of $\mathbf{NC^1}$: the class of languages $L \subseteq \{-1,1\}^*$ such that for each length n there exists a Boolean formula F_n of size $\mathrm{poly}(n)$ which decides whether strings of length n are in the language.

Another recent line of work ([San10, IMZ12, KR13, KRT13, CKK+14, CKS14]) concentrated on giving average-case formula lower bounds for problems in **P**. These works also explored applications of shrinkage properties of formulae to: pseudo-random generators, compression algorithms and non-trivial #SAT algorithms for small formulae. The state of the art average-case lower bound for de Morgan formulae is the result of Komargodski, Raz and Tal [KRT13] who gave an explicit $h_r : \{-1,1\}^n \to \{-1,1\}$ such that any formula that computes this function on a fraction $\frac{1}{2} + 2^{-r}$ must be of size at least $\frac{n^{3-o(1)}}{r^2}$ where r is an arbitrary parameter smaller than $n^{1/3}$.

1.1 Our Results

In this work, we give a new proof of Håstad's result. In fact, we obtain a tight result showing that the shrinkage exponent is exactly 2.

Theorem 1.2. *Let f be a Boolean function. For every $p > 0$,*

$$\mathop{\mathbf{E}}_{\rho\sim\mathcal{R}_p}[L(f|_\rho)] = O\left(p^2 L(f) + p\sqrt{L(f)}\right) .$$

Note that both terms in Theorem 1.2 (i.e., $p^2L(f)$ and $p\sqrt{L(f)}$) are needed as demonstrated by the parity function above. This improves the worst-case lower bound Håstad gave to Andreev's function from $\Omega(\frac{n^3}{(\log n)^{7/2}(\log\log n)^3})$ to $\Omega(\frac{n^3}{(\log n)^2(\log\log n)^3})$ immediately, following the proof of Theorem 8.1 in [Hås98]. A more careful choice of distribution over restrictions gives a slightly better bound on Andreev's function, $\Omega\left(\frac{n^3}{(\log n)^2 \log\log n}\right)$ (see Section 7). This is tight up to the $\log\log n$ factor. In addition, replacing Theorem 1.1 with Theorem 1.2 improves the analysis of the average-case lower bound in [KRT13].

Corollary 1.3. *Let n be large enough, then for any parameter $r \le n^{1/3}$ there is an explicit (computable in polynomial time) Boolean function $h_r : \{-1,1\}^{6n} \to \{-1,1\}$ such that any formula of size $\frac{n^3}{r^2\cdot \mathrm{poly}\log n}$ computes h_r correctly on a fraction of at most $1/2 + 2^{-r}$ of the inputs.*

1.2 Proof Outline

The proof comes from a surprising area: quantum query complexity. The connection between de Morgan formulae and quantum query complexity was first noted in the work of Laplante, Lee and Szegedy [LLS06]. They showed that the *quantum adversary bound* is at most the square root of the formula size of a function. Høyer, Lee and Špalek [HLS07] replaced the quantum adversary bound by the *negative weight adversary bound*, achieving a stronger relation. The long line of works [FGG08, Rei09, ACR+10, RS12, Rei11] showed that the negative weight adversary bound is equal up to a constant to the *bounded-error quantum query complexity* of a function, $Q_2(f)$. Combining all these results yields $Q_2(f) = O(\sqrt{L(f)})$. By the connection of quantum query complexity to the approximate degree [3] , $\widetilde{\deg}(f) = O(Q_2(f))$, established by Beals et al. [BBC+01], we get a classical result: $\widetilde{\deg}(f) = O(\sqrt{L(f)})$ for any Boolean function f. To our best knowledge, no classical proof that $\widetilde{\deg}(f) = O(\sqrt{L(f)})$ is known – it might be interesting to find such a proof.

[3]Let $f : \{-1,1\}^n \to \{-1,1\}$, we say that a polynomial $p(x)$ ϵ-approximates f pointwise if $|p(x) - f(x)| < \epsilon$ for all $x \in \{-1,1\}^n$. The approximate degree of a function f, denoted by $\widetilde{\deg}(f)$, is the minimal degree of a polynomial p which 1/3-approximates f pointwise.

Small formulae have exponentially small Fourier tails. We obtain a somewhat simpler proof of our main theorem, compared to Håstad's original proof, by taking the result $\widetilde{\deg}(f) = O(\sqrt{L(f)})$ as a given. First, we note that by using amplification there exists a polynomial of degree $\tilde{d} = O(\sqrt{L(f)}\log(1/\epsilon))$ which ϵ-approximates f pointwise. Using standard arguments this implies that the Fourier mass above degree $\tilde{d}$, i.e. $\sum_{S:|S|>\tilde{d}} \hat{f}(S)^2$, is at most ϵ. In other words, the Fourier mass above $O(\sqrt{L(f)} \cdot t)$ is at most 2^{-t}, and we call this property exponentially small tails of the Fourier spectrum of f above level $O(\sqrt{L(f)})$.[4]

Exponentially small Fourier tails imply a "switching lemma" type property. Our next step is novel. We show that exponentially small Fourier tails imply a strong behavior under random restrictions. If for all t, f has at most 2^{-t} of the mass above level $m \cdot t$, then under a p-random restriction we have

$$\forall d : \Pr_{\rho \sim \mathcal{R}_p}[\deg(f|_\rho) \geq d] \leq (8pm)^d \, . \quad (1)$$

[5]

In particular, if we take p to be $\leq \frac{1}{cm}$ for a large enough constant c we get that the degree of the restricted function is d with probability $\exp(-10d)$. [6]

We call such a property a *"switching lemma" type property* since the switching lemma ([Hås86]) states something similar for DNF formulae: If f can be computed by a DNF formula where each term is the logical AND of w literals, then

$$\forall d : \Pr_{\rho \sim \mathcal{R}_p}[\mathrm{DT}(f|_\rho) \geq d] \leq (5pw)^d \, .$$

Our conclusion is somewhat analogous for functions with exponentially small tails, replacing the decision tree complexity with the degree as a polynomial. We think that the relation between exponentially small Fourier tails and the "switching lemma" type property is of independent interest.

Proving the case $p = O(1/\sqrt{L(f)})$. Using the fact that functions with small formula size have exponentially small tails above level $\sqrt{L(f)}$, we get that for $p = O(1/\sqrt{L(f)})$, applying a p-random restriction yields a function with degree d with probability at most $\exp(-10d)$. In particular, with high probability the function becomes a constant. As the formula size of a degree d polynomial is at most 32^d we get that for some large enough constant c, applying a p-random restriction with $p = \frac{1}{c\sqrt{L(f)}}$, yields a function with expected formula size at most 1. This completes our proof for the case $p = \Theta(1/\sqrt{L(f)})$, and in fact the case $p = O(1/\sqrt{L(f)})$ as well.

Proving the general case. In order to establish the case where $p = \Omega(1/\sqrt{L(f)})$, we use an idea from Impagliazzo, Meka and Zuckerman's work ([IMZ12]). They showed how to decompose a large formula into $O(L(f)/\ell)$ many small formulae, each of size $O(\ell)$. Furthermore, applying any restriction, the formula size of the restricted function is at most the sum of formula sizes of the restricted sub-functions represented by the sub-formulae. Taking ℓ to be $1/p^2$ and using linearity of expectation we get the required result for general p.

[4] Of course, this is meaningless when $L(f) \geq n^2$, since there is no Fourier mass above level n.

[5] For technical reasons, it is more convinent for us to argue about the probablity of having degree exactly d. We actually show $\Pr_{\rho \sim \mathcal{R}_p}[\deg(f|_\rho) = d] \leq (4pm)^d$ and this implies the statement above by simple arithmetics.

[6] This is essentially the opposite of a key step in the proof of Linial, Mansour and Nisan [LMN93] which showed that $\mathbf{AC^0}$ circuits have Fourier spectrum concentrated on the poly $\log(n)$ first levels.

1.3 Related Work

The recent work of Impagliazzo and Kabanets [IK14] shows that shrinkage properties imply Fourier concentration. In some sense, our result shows the opposite, although we need exponential small Fourier tails to begin with.

2 Preliminaries

2.1 Formulae

A *de Morgan formula* F on n variables $x_1, \ldots, x_n$ is a binary tree whose leaves are labeled with variables or their negations, and whose internal nodes are labeled with either $\vee$ or $\wedge$ gates. The *size* of a de Morgan formula F, denoted by $L(F)$, is the number of leaves in the tree. The *formula size* of a function $f : \{-1,1\}^n \to \{-1,1\}$ is the size of the minimal formula which computes the function, and is denoted by $L(f)$. A de Morgan formula is called *read-once* if every variable appears at most once in the tree.

2.2 Restrictions

Definition 2.1 (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 $f|_\rho$ the function f restricted according to ρ in the following sense: if $\rho_i = *$ then the i-th input bit of f is unassigned and otherwise the i-th input bit of f is assigned to be ρ_i.*

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 $\mathcal{R}_p$.*

2.3 Fourier Analysis of Boolean Functions

For any Boolean function $f : \{-1,1\}^n \to \{-1,1\}$ there is a unique Fourier representation:

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

The coefficients $\hat{f}(S)$ are given by $\hat{f}(S) = \mathbf{E}_x[f(x) \cdot \prod_{i \in S} x_i]$. Parseval's equality states that $\sum_S \hat{f}(S)^2 = \mathbf{E}_x[f(x)^2] = 1$. Note that the Fourier representation is the unique multilinear polynomial which agrees with f on $\{-1,1\}^n$. The polynomial degree is denoted by $\deg(f)$ and is equal to $\max\{|S| : \hat{f}(S) \neq 0\}$. 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 by $\mathbf{W}^{\geq k}[f] \triangleq \sum_{S \subseteq [n], |S| \geq k} \hat{f}(S)^2$. The following fact relates the Fourier coefficients of f and of $f|_\rho$ where ρ is a p-random restriction. In fact, the result holds for any distribution over restrictions which is *random-valued*, as defined next.

Definition 2.3. *A distribution $\mathcal{D}$ over restrictions is* random-valued *if for $\rho \sim \mathcal{D}$, given $J = \{i \in [n] : \rho(i) = *\}$, the values of ρ on $\bar{J}$ are uniform independent bits.*

By definition, $\mathcal{R}_p$ is random-valued.

Fact 2.4 (Proposition 4.17,[O'D14]). *Let $\mathcal{D}$ be a random-valued distribution of restrictions. Then,*

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

For the case of $\mathcal{D} = \mathcal{R}_p$, summing over all coefficients of size d, we get the following corollary.

Corollary 2.5.

$$\mathop{\mathbf{E}}_{\rho\sim\mathcal{R}_p}\left[\sum_{S:|S|=d} \widehat{f|_\rho}(S)^2\right] = \sum_{k=d}^{n} \mathbf{W}^{=k}[f]\cdot\mathbf{Pr}[\mathrm{Bin}(k,p)=d]$$

One can represent a Boolean function also as $\tilde{f}:\{0,1\}^n\to\{0,1\}$. Identifying $\{0,1\}$ with $\{1,-1\}$ by $b\mapsto 1-2b$ we get the following relation between the $\{0,1\}$ and the $\{-1,1\}$ representation of the same function.

$$\tilde{f}(y) = \frac{1-f(1-2y_1,\ldots,1-2y_n)}{2} = \frac{1}{2}-\frac{1}{2}\sum_{S\subseteq[n]}\hat{f}(S)\cdot\prod_{i\in S}(1-2y_i) \tag{2}$$

Let $p(y)=\sum_{T\subseteq[n]} a_T\cdot\prod_{i\in T} y_i$ be the unique multilinear polynomial over the reals, which agrees with $\tilde{f}(y)$ on $\{0,1\}^n$. Using Equation (2) gives $a_\emptyset = 1/2 - 1/2\cdot\sum_S \hat{f}(S)$ and

$$\forall T\neq\emptyset : a_T = (-2)^{|T|-1}\cdot\sum_{S\supseteq T}\hat{f}(S)\ . \tag{3}$$

It is clear from Equation (3) that $\deg(p)=\deg(f)$, hence the definition of degree does not depend whether we are considering the $\{-1,1\}$ or the $\{0,1\}$ representation of the function. Note that since $\tilde{f}$ is Boolean, the coefficients a_T are integers, as we can write

$$\tilde{f}(y) = \sum_{z\in\{0,1\}^n}\tilde{f}(z)\cdot\prod_{i:z_i=0}(1-y_i)\cdot\prod_{i:z_i=1}y_i$$

which opens up to a multilinear polynomial over y with integer coefficients.

An immediate consequence of the above discussion is the following fact, which states that the Fourier coefficients of a degree d polynomial are 2^{-d} "granular", i.e. integer multiples of 2^{-d}.

Fact 2.6 (Granularity). *Let $f:\{-1,1\}^n\to\{-1,1\}$ with $\deg(f)=d$, then $\hat{f}(S)=k_S\cdot 2^{-d}$ where $k_S\in\mathbb{Z}$ for any $S\subseteq[n]$.*

Proof. We prove by contradiction. Let T be a maximal set with respect to inclusion for which $\hat{f}(T)$ is not an integer multiple of 2^{-d}. We first handle the case $T\neq\emptyset$. Equation (3) gives $a_T=(-2)^{|T|-1}\sum_{S\supseteq T}\hat{f}(S)$. Multiplying both sides by $(-2)^{d-|T|+1}$ we get

$$(-2)^{d-|T|+1}\cdot a_T = (-2)^d\sum_{S\supseteq T}\hat{f}(S)\ .$$

By the assumption on maximality of T, all coefficients on the RHS except $\hat{f}(T)$ are integer multiples of 2^{-d}, hence the RHS is not an integer. On the other hand, the LHS is an integer since a_T is an integer, and we reach a contradiction.

For the case $T=\emptyset$, we have $a_\emptyset = 1/2-1/2\cdot\sum_S\hat{f}(S)$. Multiplying both sides by 2^{d+1} gives $2^{d+1}a_\emptyset = 2^d - 2^d\sum_S\hat{f}(S)$. Again, the RHS is not an integer, while the LHS is an integer. □

Definition 2.7. *We define the sparsity of* $f : \{-1,1\}^n \to \{-1,1\}$ *as* $\mathrm{sparsity}(f) \triangleq |\{S : \hat{f}(S) \neq 0\}|$.

Corollary 2.8. *Let* $f : \{-1,1\}^n \to \{-1,1\}$ *with* $\deg(f) = d$*, then* $\mathrm{sparsity}(f) \leq 2^{2d}$.

Proof. By Parseval, $1 = \sum_S \hat{f}(S)^2 \geq \mathrm{sparsity}(f) \cdot \left(2^{-d}\right)^2$. □

Claim 2.9. *Let* $\tilde{f} : \{0,1\}^n \to \{0,1\}$ *be a Boolean function with* $\deg(\tilde{f}) = d$ *then* $\tilde{f}$ *can be written as*

$$\tilde{f}(x) = \sum_{i=1}^{\mathrm{sparsity}(f)} g_i(x)$$

where each $g_i : \{0,1\}^n \to \mathbb{Z}$ *is a* d*-junta, i.e. depends only on at most* d *coordinates.*

Proof. Write $\tilde{f}(x) = \sum_{T \subseteq [n]} a_T \prod_{i \in T} x_i$. By Equation (3) any $T \subseteq [n]$ such that $a_T \neq 0$ is contained in some subset $S \subseteq [n]$ for which $\hat{f}(S) \neq 0$. Arbitrarily order the sets $\{S : \hat{f}(S) \neq 0\}$ as $S_1, \ldots, S_{\mathrm{sparsity}(f)}$ and let

$$g_i(x) = \sum_{T \subseteq S_i, \forall j < i: T \not\subseteq S_j} a_T \cdot \prod_{i \in T} x_i \ .$$

Then, by definition $\tilde{f}(x) = \sum_{i=1}^{\mathrm{sparsity}(f)} g_i(x)$. By the integrality of a_T, each g_i takes integer values. Moreover, each g_i depends only on the variables in the set S_i, i.e. on at most d coordinates. □

2.4 Approximate Degree

Let $f : \{-1,1\}^n \to \{-1,1\}$. Given an $\epsilon \geq 0$ we define the ϵ*-approximate degree*, denoted by $\widetilde{\deg}_\epsilon(f)$, as the minimal degree of a multilinear polynomial p such that for all $x \in \{-1,1\}^n$, $|f(x) - p(x)| \leq \epsilon$. We denote $\widetilde{\deg}_{1/3}(f)$ by $\widetilde{\deg}(f)$.

When defining approximate degree the choice of $1/3$ may seem arbitrary. The next fact (essentially proved in [BNRdW07], Lemma 1) shows how approximate degree for different errors relate. We prove this fact in Appendix B for completeness.

Fact 2.10. *Let* $f : \{-1,1\}^n \to \{-1,1\}$ *be a Boolean function and let* $0 < \epsilon < 1$ *then:* $\widetilde{\deg}_\epsilon(f) \leq \widetilde{\deg}(f) \cdot \lceil 8 \cdot \ln(2/\epsilon) \rceil$.

Relating the approximate degree to the Fourier transform one gets the following fact.

Fact 2.11. *Let* $f : \{-1,1\}^n \to \{-1,1\}$ *be a Boolean function,* $0 < \epsilon < 1$ *and* $d = \widetilde{\deg}_\epsilon(f)$*, then* $\mathbf{W}^{>d}[f] \leq \epsilon^2$.

Proof. Let p be a polynomial of degree d which ϵ approximates f pointwise. Obviously $\mathbf{E}_x[(f(x) - p(x))^2] \leq \epsilon^2$. Let q be the best ℓ_2 approximation of f by a degree d polynomial, namely the polynomial of degree d which minimizes $\|f - q\|_2^2 \triangleq \mathbf{E}_x[(f(x) - q(x))^2]$. Obviously, $\|f - q\|_2^2 \leq \|f - p\|_2^2 \leq \epsilon^2$ by the choice of p and q. Using Parseval's equality $\|f - q\|_2^2 = \sum_S \left(\hat{f}(S) - \hat{q}(S)\right)^2$, and it is easy to see that the minimizer of this expression among degree d polynomials is the Fourier expansion of f truncated above degree d:

$$q(x) = \sum_{S \subseteq [n] : |S| \leq d} \hat{f}(S) \cdot \prod_{i \in S} x_i \ .$$

Overall, we get that $\epsilon^2 \geq \|f - q\|_2^2 = \sum_{S : |S| > d} \hat{f}(S)^2$. □

Our proof relies heavily on the following result from quantum query complexity.

Theorem 2.12 ([BBC+01, HLS07, Rei11])**.** *There exists a universal constant $C_1 \geq 1$ such that for any $f : \{-1,1\}^n \to \{-1,1\}$ we have $\widetilde{\deg}(f) \leq C_1 \cdot \sqrt{L(f)}$.*

The next claim states that functions have exponentially small fourier tails above level $\sqrt{L(f)}$.

Claim 2.13. *There exists a constant $C > 0$ such that for any $f : \{-1,1\}^n \to \{-1,1\}$ and $k \in \mathbb{N}$,*

$$\mathbf{W}^{\geq k}[f] \leq e \cdot \exp\left(\frac{-k}{C\sqrt{L(f)}}\right) .$$

Proof. Let $t = \frac{k}{C\sqrt{L(f)}}$ where C is some constant we shall set later. We prove that $\mathbf{W}^{\geq k}[f] \leq e \cdot e^{-t}$. Assume without loss of generality that $t \geq 1$ or else the claim is trivial since $\mathbf{W}^{\geq k}[f] \leq 1 \leq e \cdot e^{-t}$. Put $\epsilon = e^{-t/2}$, and combine Theorem 2.12 and Fact 2.10 to get

$$\widetilde{\deg}_\epsilon(f) \leq \sqrt{L(f)} \cdot C_1 \cdot \lceil 8\ln(2/\epsilon)\rceil = \sqrt{L(f)} \cdot C_1 \cdot \lceil 4t + 8\ln(2)\rceil \underset{(t\geq 1)}{\leq} \sqrt{L(f)} \cdot C_1 \cdot 11t .$$

Using Fact 2.11 we get $\mathbf{W}^{>\sqrt{L(f)}\cdot C_1 \cdot 11t}[f] \leq e^{-t}$. Hence $\mathbf{W}^{\geq\sqrt{L(f)}\cdot C_1 \cdot 12t}[f] \leq e^{-t}$. Setting $C := C_1 \cdot 12$ completes the proof. □

2.5 The Generalized Binomial Theorem

Theorem 2.14 (The generalized binomial theorem)**.** *Let $|x| < 1$, and $d \in \mathbb{N}$, then*

$$\sum_{n=0}^{\infty} \binom{d+n-1}{d-1} \cdot x^n = \frac{1}{(1-x)^d} .$$

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

Corollary 2.15. *Let $|x| < 1$, and $d \in \mathbb{N}$ then $\sum_{k=d}^{\infty} \binom{k-1}{d-1} \cdot x^k = \frac{x^d}{(1-x)^d}$.*

3 Exponentially Small Tails and The Switching Lemma

In this section we prove the main technical part of our proof by showing a close relation between two properties of Boolean functions:

1. Having exponentially small Fourier tails above level t: $\forall k : \mathbf{W}^{\geq k}[f] \leq e^{-k/t}$.
2. A "switching lemma" type property with parameter t': $\forall p, d : \mathbf{Pr}_{\rho\sim\mathcal{R}_p}[\deg(f|_\rho) \geq d] \leq (t'p)^d$.

Linial, Mansour and Nisan proved that Property 2 implies Property 1. For completeness we include a proof of their theorem in Appendix A.

Theorem 3.1 ([LMN93], restated slightly)**.** *Let $f : \{-1,1\}^n \to \{-1,1\}$ and assume there exists $t > 0$ such that for all $d \in \mathbb{N}, p \in (0,1)$, $\mathbf{Pr}_{\rho\sim\mathcal{R}_p}[\deg(f|_\rho) \geq d] \leq (tp)^d$; then for any k, $\mathbf{W}^{\geq k}[f] \leq 2e \cdot e^{-k/te}$.*

Next, we prove a converse to Theorem 3.1.

Theorem 3.2. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean function, let $t, C > 0$ such that for all k, $\mathbf{W}^{\geq k}[f] \leq C \cdot e^{-k/t}$ and let ρ be a p-random restriction; then for all d, $\mathbf{Pr}[\deg(f|_\rho) = d] \leq C \cdot (4pt)^d$.*

Proof Sketch If a function f has exponentially small Fourier tails above level t then on expectation the restricted function $f|_\rho$ will have exponentially small Fourier tails above level pt, since the Fourier spectrum of f roughly squeezes by a factor of p under a p-random restriction (see Corollary 2.5). However, the Fourier mass above level d of a Boolean function of degree d cannot be smaller than 4^{-d} by the granularity property. We get that if $pt \ll 1$, then with high probability the restricted function is not a degree d polynomial.

Proof. Our proof strategy is as follows: we bound the value of $\mathbf{E}_\rho\left[\sum_{S:|S|=d}\widehat{f|_\rho}(S)^2\right]$ from below and above showing

$$\mathop{\mathbf{E}}_{\rho}\left[\sum_{S:|S|=d}\widehat{f|_\rho}(S)^2\right] \geq \mathbf{Pr}[\deg(f|_\rho)=d]\cdot 4^{-d} \tag{4}$$

and

$$\mathop{\mathbf{E}}_{\rho}\left[\sum_{S:|S|=d}\widehat{f|_\rho}(S)^2\right] \leq C\,(pt)^d\ . \tag{5}$$

Combining the two estimates will complete the proof.

We begin by proving Equation (4). Conditioning on the event that $\deg(f|_\rho)=d$, Fact 2.6 implies that any nonzero Fourier coefficient of $f|_\rho$ is of magnitude $\geq 2^{-d}$. Hence, $\sum_{S:|S|=d}\widehat{f|_\rho}(S)^2 \geq 4^{-d}$, and we get

$$\mathop{\mathbf{E}}_{\rho}\left[\sum_{S:|S|=d}\widehat{f|_\rho}(S)^2\right] \geq \mathbf{Pr}[\deg(f|_\rho)=d]\cdot\mathop{\mathbf{E}}_{\rho}\left[\sum_{S:|S|=d}\widehat{f|_\rho}(S)^2 \,\middle|\, \deg(f|_\rho)=d\right] \geq \mathbf{Pr}[\deg(f|_\rho)=d]\cdot 4^{-d}\ .$$

Next, we turn to prove Equation (5).

$$\begin{aligned}
\mathop{\mathbf{E}}_{\rho}\left[\sum_{S:|S|=d}\widehat{f|_\rho}(S)^2\right] &= \sum_{k\geq d}\mathbf{W}^{=k}[f]\cdot\binom{k}{d}\cdot p^d\cdot(1-p)^{k-d} && \text{(Corollary 2.5)}\\
&\leq \sum_{k\geq d}\mathbf{W}^{=k}[f]\cdot\binom{k}{d}\cdot p^d\\
&= p^d\cdot\sum_{k\geq d}\left(\mathbf{W}^{\geq k}[f]-\mathbf{W}^{\geq k+1}[f]\right)\cdot\binom{k}{d}
\end{aligned}$$

We can rearrange the RHS of the above equation, gathering terms according to $\mathbf{W}^{\geq k}[f]$. We denote $\binom{d-1}{d}=0$, and get:

$$\begin{aligned}
\mathop{\mathbf{E}}_{\rho}\left[\sum_{S:|S|=d}\widehat{f|_\rho}(S)^2\right] &= p^d\cdot\sum_{k\geq d}\mathbf{W}^{\geq k}[f]\cdot\left(\binom{k}{d}-\binom{k-1}{d}\right)\\
&= p^d\cdot\sum_{k\geq d}\mathbf{W}^{\geq k}[f]\cdot\binom{k-1}{d-1}\ .
\end{aligned}$$

Let $a := e^{-1/t}$. The assumption on the Fourier tails of f, $\mathbf{W}^{\geq k}[f]\leq C\cdot a^k$, gives

$$\mathop{\mathbf{E}}_{\rho}\left[\sum_{S:|S|=d}\widehat{f|_\rho}(S)^2\right] \leq p^d\cdot\sum_{k\geq d}C\cdot a^k\cdot\binom{k-1}{d-1}\ .$$

Next we use Corollary 2.15 with $x := a$ to get

$$\mathop{\mathbf{E}}_{\rho}\left[\sum_{S:|S|=d} \widehat{f|_\rho}(S)^2\right] \le C\left(\frac{ap}{1-a}\right)^d = C\left(\frac{p}{1/a-1}\right)^d .$$

Substituting a with $e^{-1/t}$ gives

$$\mathop{\mathbf{E}}_{\rho}\left[\sum_{S:|S|=d} \widehat{f|_\rho}(S)^2\right] \le C\left(\frac{p}{e^{1/t}-1}\right)^d \le C\,(pt)^d ,$$

where the last inequality follows since $e^x - 1 \ge x$ for any $x \ge 0$. $\square$

4 Degree vs. Formula Size

We use the following fact about the formula size of the parity function

Fact 4.1 ([Yab54]). *$L(\mathrm{PARITY}_m) \le 9/8 \cdot m^2$. Furthermore, if $m = 2^k$ for some integer k, then $L(\mathrm{PARITY}_m) \le m^2$.*

Claim 4.2. *Let $\tilde{f} : \{0,1\}^n \to \{0,1\}$ such that $\deg(\tilde{f}) = d$, then $L(\tilde{f}) \le 2 \cdot 32^d$.*

Proof. According to Claim 2.9, $\tilde{f}$ can be written as $\sum_{i=1}^{4^d} g_i(x)$, where the functions $g_i(x)$ take integer values, and each of them depends on at most d variables. Since $\tilde{f}(x) \in \{0,1\}$ we may perform all operations modulo 2 and get $\tilde{f}(x) = \bigoplus_{i=1}^{4^d} h_i(x)$, where $h_i(x) = g_i(x) \mod 2$. Taking a formula for the parity of $m = 4^d$ variables, $y_1, \ldots, y_m$, and replacing each instance of a variable y_i with a formula computing $h_i(x)$ gives a formula for $\tilde{f}$. The size of the formula computing each h_i is at most 2^{d+1} since any function on d variables can be computed by a formula of such size. Thus, the size of the combined formula is $\le L(\mathrm{PARITY}_m) \cdot 2^{d+1} = 16^d \cdot 2^{d+1} = 2 \cdot 32^d$. $\square$

5 The Case $p = O(1/\sqrt{L(f)})$

Claim 5.1. *There exists a constant $C > 0$ such that for any function $f : \{-1,1\}^n \to \{-1,1\}$ and any $p \le \frac{1}{C\sqrt{L(f)}}$ the following hold. Let ρ be a p-random restriction, then $\mathbf{E}_\rho[L(f|_\rho)] = O(p\sqrt{L(f)})$. In particular, in this regime of parameters, $\mathbf{E}_\rho[L(f|_\rho)] = O(1)$.*

Proof of Claim 5.1. From Claim 2.13, there exists a constant $C > 0$ such that

$$\forall k : \mathbf{W}^{\ge k}[f] \le e \cdot e^{-k/(C\sqrt{L(f)})} .$$

This implies, using Theorem 3.2, that $\mathbf{Pr}_{\rho\sim\mathcal{R}_p}[\deg(f|_\rho) = d] \le e \cdot \left(4pC\sqrt{L(f)}\right)^d$. Using Claim 4.2, if $\deg(f|_\rho) = d$ then $L(f|_\rho) \le 2 \cdot 32^d$. For $p \le \frac{1}{64 \cdot 4C\sqrt{L(f)}}$ we get

$$
\begin{aligned}
\mathop{\mathbf{E}}_{\rho\sim\mathcal{R}_p}[L(f|_\rho)] &= \sum_{d=0}^{n} \mathop{\mathbf{Pr}}_{\rho}[\deg(f|_\rho)=d] \cdot \mathop{\mathbf{E}}_{\rho}[L(f|_\rho) | \deg(f|_\rho)=d] \\
&= \sum_{d=1}^{n} \mathop{\mathbf{Pr}}_{\rho}[\deg(f|_\rho)=d] \cdot \mathop{\mathbf{E}}_{\rho}[L(f|_\rho) | \deg(f|_\rho)=d] \quad (\deg(f|_\rho)=0 \text{ implies } L(f|_\rho)=0) \\
&\le \sum_{d=1}^{\infty} e \cdot \left(4pC\sqrt{L(f)}\right)^d \cdot 2 \cdot 32^d \\
&\le O(p\sqrt{L(f)}) \cdot \sum_{d=1}^{\infty} \left(4pC\sqrt{L(f)}\right)^{d-1} \cdot 32^{d-1} \\
&\le O(p\sqrt{L(f)}) \cdot \sum_{d=1}^{\infty} (1/64)^{d-1} \cdot 32^{d-1} \\
&= O(p\sqrt{L(f)}) \ . \qquad \square
\end{aligned}
$$

6 The General Case

In Section 5 we have proved Theorem 1.2 for the case $p = O(1/\sqrt{L(f)})$. In this section we give a reduction from the case where p is larger, i.e. $p = \Omega(1/\sqrt{L(f)})$, to the case where p is small, i.e. $p = \Theta(1/\sqrt{L(f)})$. We use the tree decompsition of Impagliazzo, Meka and Zuckerman [IMZ12] to establish this reduction.[7]

The next lemma states that every binary tree can be decomposed into smaller subtrees with some small overhead. Its proof can be found in [IMZ12].

Lemma 6.1 ([IMZ12])**.** *Let $\ell \in \mathbb{N}$. Any binary tree with $s \ge \ell$ leaves can be decomposed into at most $6s/\ell$ subtrees, each with at most ℓ leaves, such that each subtree has at most two other subtree children. Here subtree T_1 is a child of subtree T_2 if there exists nodes $t_1 \in T_1$, $t_2 \in T_2$, such that t_1 is a child of t_2.*

Claim 6.2. *Let F be a formula over the set of variables $X = \{x_1, \dots, x_n\}$, and $\ell \in \mathbb{N}$ be some parameter; then, there exist $m \le 36 \cdot L(F)/\ell$ formulae over X, denoted by $G_1, \dots, G_m$, each of size at most ℓ, and there exists a read-once formula F' of size m such that $F'(G_1(x), \dots, G_m(x)) = F(x)$ for all $x \in \{-1, 1\}^n$.*

Proof. Consider the decomposition promised by Lemma 6.1 with parameter ℓ. Let $T_1, \dots, T_{m'}$ be the subtrees in this decomposition where $m' \le 6n/\ell$. We will show by induction on m', that one can construct a read-once formula F' of size $m \le 6m'$ along with m sub-formulae $G_1, \dots, G_m$ of size ℓ such that $F \equiv F'(G_1, \dots, G_m)$. For $m' = 1$ the statement holds trivially.

Assume that the root of the formula F is a node in the subtree T_1, and that the subtree T_1 has two subtree children: T_2 and T_3 (the case where T_1 has one subtree child can be handled similarly, and is in fact slightly simpler). We now add two special leaves to the tree T_1. Let $t_2 \in T_2, t_1 \in T_1$ (respectively $t_3 \in T_3, t'_1 \in T_1$) be the nodes such that t_2 (t_3, resp.) is a child of t_1 (t'_1, resp.) in the tree represented by F, and add a leaf labeled by the "special" variable z_2 (z_3, resp.) as a child of

[7] Another approach to prove the general case is to follow Håstad original proof, changing the estimates when $p = O(1/\sqrt{L(F)})$ with what we showed in Section 5. The reduction we suggest simplifies this approach significantly.

t_1 (t'_1, resp.). Call the new subtree T. Note that since T is a de Morgan formula, the value of T is monotone in z_2 and z_3. Let T' be the minimal subtree of T which contains both leaves marked by z_2 and z_3. By minimality $T' = T'_2$ **op** T'_3, for **op** $\in \{\wedge, \vee\}$, where T'_2 contains z_2 and not z_3, and T'_3 contains z_3 and not z_2.

We will construct a formula equivalent to T' by finding equivalent formulae for T'_2 and T'_3. We claim that $T'_2 = (T'_2|_{z_2=\mathbf{false}}) \vee (T'_2|_{z_2=\mathbf{true}} \wedge z_2)$. This follows since T'_2 is monotone in z_2: if $T'_2|_{z_2=\mathbf{false}} = \mathbf{true}$ then $T'_2 = \mathbf{true}$, otherwise $T'_2 = \mathbf{true}$ only if both $T'_2|_{z_2=\mathbf{true}}$ and z_2 are **true**. Same goes for T'_3, and we get

$$T' \equiv \left((T'_2|_{z_2=\mathbf{false}}) \vee (T'_2|_{z_2=\mathbf{true}} \wedge z_2)\right) \ \mathbf{op}\ \left((T'_3|_{z_3=\mathbf{false}}) \vee (T'_3|_{z_3=\mathbf{true}} \wedge z_3)\right) \ .$$

Replacing T' with a leaf labeled with z, where z is a new "special" variable, and doing the same trick we get: $T \equiv T|_{z=\mathbf{false}} \vee (T|_{z=\mathbf{true}} \wedge z)$. Combining both formulae, we get the following equivalence:

$$T \equiv T|_{z=\mathbf{false}} \vee \left(T|_{z=\mathbf{true}} \wedge \left((T'_2|_{z_2=\mathbf{false}}) \vee (T'_2|_{z_2=\mathbf{true}} \wedge z_2)\right) \ \mathbf{op}\ \left((T'_3|_{z_3=\mathbf{false}}) \vee (T'_3|_{z_3=\mathbf{true}} \wedge z_3)\right)\right) \ .$$

Note that the RHS of the equation above can be written as $F''(G_1(x), \dots, G_6(x), z_2, z_3)$ where F'' is read-once and $G_1(x), \dots, G_6(x)$ are formulae of size ℓ, defined on the variables in X.

Let m_2, m_3 be the number of subtrees which are descendants of T_2, T_3 in the tree-decomposition given by Lemma 6.1. By induction, the subformula of F rooted at t_2 is equivalent to $F'_2(G^2_1(x), \dots, G^2_{6m_2}(x))$ where F'_2 is read-once and $G^2_i(x)$ are formulae of size $\le \ell$. Similarly for t_3. We thus get that

$$F(x) = F''\left(G_1(x), \dots, G_6(x),\ F'_2(G^2_1(x), \dots, G^2_{6m_2}(x)),\ F'_3(G^3_1(x), \dots, G^3_{6m_3}(x))\right) \ .$$

Rearranging the RHS, we get a read-once formula of size $m \le 6 + 6m_2 + 6m_3 = 6m'$ alongside m sub-formulae, each of size ℓ, such that their composition is equivalent to F. □

We now turn to complete the proof of our main theorem.

Theorem (Theorem 1.2, restated). *Let $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean function, and let $p > 0$, then* $\mathbf{E}_{\rho\sim\mathcal{R}_p}[L(f|_\rho)] = O\left(p^2 L(f) + p\sqrt{L(f)}\right)$.

Proof. The case $p \le \frac{1}{C\sqrt{L}}$ is implied by Claim 5.1. Therefore, it is enough to show the statement holds when $p > \frac{1}{C\sqrt{L}}$. Let F be the smallest de Morgan formula computing f. Applying Claim 6.2 with $\ell := \frac{1}{p^2 \cdot C^2}$, we get a read-once de Morgan formula F' of size $m = O(L(F)/\ell)$ along with formulae $G_1, \dots, G_m$, each of size at most ℓ, such that $f(x) = F'(G_1(x), \dots, G_m(x))$ for all $x \in \{-1,1\}^n$. Denote the functions that $G_1, \dots, G_m$ compute by $g_1, \dots, g_m$ respectively. Applying a restriction ρ we get $f|_\rho \equiv F'(g_1|_\rho, \dots, g_m|_\rho)$, hence $L(f|_\rho) \le \sum_{i=1}^m L(g_i|_\rho)$. By linearity of expectation,

$$\mathop{\mathbf{E}}_{\rho}[L(f|_\rho)] \le \mathop{\mathbf{E}}_{\rho}\left[\sum_{i=1}^m L(g_i|_\rho)\right] \le m \cdot O(p \cdot \sqrt{\ell}) = m \cdot O(1) = O(p^2 \cdot L(f)) \ . \qquad \square$$

7 Lower Bound for Andreev's Function

In this section, we prove a $\Omega\left(\frac{n^3}{\log^2 n \log\log n}\right)$ formula lower bound for Andreev's function.

For two Boolean functions $f : \{0,1\}^n \to \{0,1\}$ and $g : \{0,1\}^m \to \{0,1\}$, the *composition* of f and g is defined as $f \circ g : \{0,1\}^{nm} \to \{0,1\}$, where

$$(f \circ g)(x_{1,1}, x_{1,2}, \ldots, x_{n,m}) = f(g(x_{1,1}, \ldots, x_{1,m}), \ldots, g(x_{n,1}, \ldots, x_{n,m}))\ .$$

In words, $f \circ g$ is a function whose value is the value of f on n input bits, each of them is the calculation of g on an independent set of m bits.

The next lemma shows that the size of $h \circ \oplus_m$, where $\oplus_m$ is the parity function on m variables, is equal, up to a constant factor, to the product of the formula sizes of h and $\oplus_m$.

Lemma 7.1. *For any $h : \{0,1\}^r \to \{0,1\}$, let $f = h \circ \oplus_m$, then*

$$L(f) = \Theta(L(h) \cdot L(\oplus_m)) = \Theta(L(h) \cdot m^2)$$

Proof. Recall that by Fact 4.1 and Khrapchenko's Theorem [Khr71], $m^2 \le L(\oplus_m) \le 9/8 \cdot m^2$.

Think of the input to f as an $r \times m$ matrix $\{y_{i,j}\}_{i\in[r],j\in[m]}$, and of the input to h as a vector $z = (z_1, \ldots, z_r)$. The upper bound, $L(f) \le L(h) \cdot L(\oplus_m)$, is easy, since replacing each leaf marked by a variable z_i (or its negation) in the formula for h with a formula computing $\oplus_{j\in[m]} y_{i,j}$ (or its negation), gives a formula for f whose size is at most $L(h) \cdot L(\oplus_m) = O(L(h) \cdot m^2)$.

For the lower bound, $L(f) = \Omega(L(h) \cdot L(\oplus_m))$, we can assume without loss of generality that $L(h) \ge 2C$ for a large enough constant C. This is without loss of generality since:

1. if $L(h) = 0$, then we have nothing to prove.

2. We show that if $1 \le L(h) < 2C$ then $L(f)$ is at least $L(\oplus_m)$. Since h is not the constant function, there is an input bit z_k of h and a restriction fixing $z_1, \ldots, z_{k-1}, z_{k+1}, \ldots, z_r$ under which h becomes the dictatorship function z_k or the anti-dictatorship function $\neg z_k$. Fixing each row in $\{y_{i,j}\}$ except the k-th row, such that the parity of the i-th row equals the required value for z_i, gives a restriction ρ under which $f|_\rho$ is the parity of $y_{k,1}, \ldots, y_{k,m}$ or its negation. Hence, $L(f) \ge L(f|_\rho) \ge L(\oplus_m) \ge \frac{L(h)L(\oplus_m)}{2C} \ge \Omega(L(h)L(\oplus_m))$.

For larger values of $L(h)$, we shall establish the lower bound $L(f) \ge \Omega\left(L(h) \cdot m^2\right)$ using a tailored distribution of random restrictions, which is not a distribution of p-random restrictions. For each row in the matrix $\{y_{i,j}\}$, we pick one variable uniformly, keep it alive, and fix all the rest uniformly. This leaves us with a function on r variables which is equivalent to h, up to negations to the inputs, hence its formula size is $L(h)$.

We want to analyze the shrinkage factor due to this distribution of random restrictions. Noting that our distribution is *random-valued* (Recall Def. 2.3), as required in Fact 2.4, we get

$$\mathop{\mathbf{E}}_{\rho}\left[\widehat{f|_\rho}(S)^2\right] = \sum_{U \supseteq S} \hat{f}(U)^2 \mathop{\mathbf{Pr}}_{\rho}[S = \{i \in U : \rho(i) = *\}]\ .$$

By the definition of the distribution of random restrictions, $\mathbf{Pr}_\rho[S = \{i \in U : \rho(i) = *\}] = 0$ if S contains more than one coordinate in a certain row. Thus, we may restrict our attention to sets S which contain at most one variable from each row. Since the probability that ρ restricts U to S is at most the probability that ρ keeps alive all the variables in S, and since each variable in S is in its own row, we get

$$\mathop{\mathbf{Pr}}_{\rho}[S = \{i \in U : \rho(i) = *\}] \le \frac{1}{m^{|S|}}\ .$$

Summing over all sets S of size d we get

$$\mathop{\mathbf{E}}_{\rho}\left[\sum_{|S|=d} \widehat{f|_\rho}(S)^2\right] = \sum_U \hat{f}(U)^2 \sum_{\substack{S\subseteq U:\\ |S|=d}} \mathop{\mathbf{Pr}}_{\rho}[S = \{i \in U : \rho(i) = *\}] \le \sum_U \hat{f}(U)^2 \binom{|U|}{d} \frac{1}{m^d} .$$

Plugging this in the analysis of Theorem 3.2 we get $\mathbf{Pr}[\deg(f|_\rho) = d] \le C(4t/m)^d$. From here we can continue the proofs of Claim 5.1 and Theorem 1.2 by replacing p with $1/m$. We get that $\mathbf{E}_\rho[L(f|_\rho)] = O\left(\frac{L(f)}{m^2} + 1\right)$. Conversely,

$$L(f) \ge \Omega\left(m^2 \cdot (\mathop{\mathbf{E}}_{\rho}[L(f|_\rho)] - C)\right)$$

for some universal constant C. Since

$$\mathop{\mathbf{E}}_{\rho}[L(f|_\rho)]) - C \ge L(h) - C \ge L(h)/2 ,$$

where we used the assumption $L(h) \ge 2C$ in the last inequality, we get $L(f) \ge \Omega(m^2 L(h))$. □

We now describe Andreev's function $A : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. A gets two inputs $x, y \in \{0,1\}^n$. Let $r = \log n$ and $m = n/\log n$. We interpret the second input y as an $r \times m$ matrix $\{y_{i,j}\}_{i\in[r],j\in[m]}$. Let $z_1, \ldots z_r \in \{0,1\}$ be the parities of each row, i.e., $z_i = \oplus_{j=1}^m y_{i,j}$. Then $A(x,y) = x_{\mathrm{bin}(z)}$ where $\mathrm{bin}(z)$ is the integer between 1 and 2^r represented by the string $z_1, \ldots, z_r$. An alternative way to view the function is to think of x as the truth table of a Boolean function on $\log n$ bits, and then take the value of this function on the input z.

Theorem 7.2.
$$L(A) \ge \Omega\left(\frac{n^3}{\log^2 n \log\log n}\right) .$$

Proof. Let $h : \{0,1\}^{\log n} \to \{0,1\}$ be the function on $\log n$ variables with largest formula size. It is well known (Theorem 1.23, [Juk12]) that $L(h) = \Omega(n/\log\log n)$. Define $A_h : \{0,1\}^n \to \{0,1\}$ as $A_h(y) = A(\mathrm{tt}(h), y) = (h \circ \oplus_m)(y_{1,1}, \ldots, y_{r,m})$ where $\mathrm{tt}(h)$ stands for the truth table of h. Using Lemma 7.1, $L(A_h) = L(h \circ \oplus_m) = \Theta(L(h) \cdot m^2) = \Theta\left(\frac{n^3}{\log^2 n \log\log n}\right)$. Since A_h is a subfunction of A, $L(A) \ge L(A_h)$, which completes the proof. □

8 Open Ends

An interesting open question raised by Håstad in [Hås98] is

> What is the shrinkage exponent of monotone de Morgan formulae?

In particular, this has strong connections with understanding the monotone formula size of Majority. The analysis in Section 6 implies that it is enough to find the critical probability p_c for which $\mathbf{E}_{\rho\sim\mathcal{R}_{p_c}}[L(f|_\rho)] = 1$, and then use the tree decomposition to argue for $p \ge p_c$ (note that the decomposition done in Section 6 respects monotonicity). Hence, in order to show Γ shrinkage, i.e. that formulae of size s shrink to expected size $O(p^\Gamma s + 1)$ after applying a p-random restriction, it is necessary and sufficient to show that for $p = \frac{1}{L(f)^{1/\Gamma}}$, the expected size of the minimal monotone formula computing $f|_\rho$ is $O(1)$.

Acknowledgement

I wish to thank my advisor Ran Raz for his guidance and encouragement. I thank Zeev Dvir and Ilan Komargodski for helpful discussions. I thank Robin Kothari, Igor Sergeev and the anonymous referees for their helpful comments.

References

[ACR+10] A. Ambainis, A. M. Childs, B. Reichardt, R. Spalek, and S. Zhang. Any AND-OR formula of size n can be evaluated in time $n^{1/2+o(1)}$ on a quantum computer. *SIAM J. Comput.*, 39(6):2513–2530, 2010.

[And87] A. E. Andreev. On a method for obtaining more than quadratic effective lower bounds for the complexity of π-schemes. *Moscow Univ. Math. Bull.*, 42:63–66, 1987. In Russian.

[BBC+01] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. Quantum lower bounds by polynomials. *J. ACM*, 48(4):778–797, 2001.

[BNRdW07] H. Buhrman, I. Newman, H. Röhrig, and R. de Wolf. Robust polynomials and quantum algorithms. *Theory Comput. Syst.*, 40(4):379–395, 2007.

[CKK+14] R. Chen, V. Kabanets, A. Kolokolova, R. Shaltiel, and D. Zuckerman. Mining circuit lower bound proofs for meta-algorithms. In *CCC*, 2014.

[CKS14] R. Chen, V. Kabanets, and N. Saurabh. An improved deterministic #SAT algorithm for small De Morgan formulas. In *MFCS*, pages 165–176, 2014.

[FGG08] E. Farhi, J. Goldstone, and S. Gutmann. A quantum algorithm for the hamiltonian nand tree. *Theory of Computing*, 4(1):169–190, 2008.

[GMWW14] D. Gavinsky, O. Meir, O. Weinstein, and A. Wigderson. Toward better formula lower bounds: an information complexity approach to the KRW composition conjecture. In David B. Shmoys, editor, *STOC*, pages 213–222. ACM, 2014.

[Hås86] Johan Håstad. Almost optimal lower bounds for small depth circuits. In Juris Hartmanis, editor, *STOC*, pages 6–20. ACM, 1986.

[Hås98] J. Håstad. The shrinkage exponent of De Morgan formulas is 2. *SIAM J. Comput.*, 27(1):48–64, 1998.

[HLS07] P. Høyer, T. Lee, and R. Spalek. Negative weights make adversaries stronger. In David S. Johnson and Uriel Feige, editors, *STOC*, pages 526–535. ACM, 2007.

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

[IMZ12] R. Impagliazzo, R. Meka, and D. Zuckerman. Pseudorandomness from shrinkage. In *FOCS*, pages 111–119. IEEE Computer Society, 2012.

[IN93] R. Impagliazzo and N. Nisan. The effect of random restrictions on formula size. *Random Struct. Algorithms*, 4(2):121–134, 1993.

[Juk12] S. Jukna. *Boolean Function Complexity: Advances and Frontiers*. Springer Berlin Heidelberg, 2012.

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

[Khr71] V. M. Khrapchenko. A method of determining lower bounds for the complexity of π schemes. *Matematischi Zametki*, 10:83–92, 1971. In Russian.

[KR13] I. Komargodski and R. Raz. Average-case lower bounds for formula size. In D. Boneh, T. Roughgarden, and J. Feigenbaum, editors, *STOC*, pages 171–180. ACM, 2013.

[KRT13] I. Komargodski, R. Raz, and A. Tal. Improved average-case lower bounds for De Morgan formula size. In *FOCS*, pages 588–597. IEEE Computer Society, 2013.

[KRW95] M. Karchmer, R. Raz, and A. Wigderson. Super-logarithmic depth lower bounds via the direct sum in communication complexity. *Computational Complexity*, 5(3/4):191–204, 1995.

[LLS06] S. Laplante, T. Lee, and M. Szegedy. The quantum adversary method and classical formula size lower bounds. *Computational Complexity*, 15(2):163–196, 2006.

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

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

[PZ93] M. Paterson and U. Zwick. Shrinkage of De Morgan formulae under restriction. *Random Struct. Algorithms*, 4(2):135–150, 1993.

[Rei09] B. Reichardt. Span programs and quantum query complexity: The general adversary bound is nearly tight for every boolean function. In *FOCS*, pages 544–551. IEEE Computer Society, 2009.

[Rei11] B. Reichardt. Reflections for quantum query algorithms. In Dana Randall, editor, *SODA*, pages 560–569. SIAM, 2011.

[RS12] B. Reichardt and R. Spalek. Span-program-based quantum algorithm for evaluating formulas. *Theory of Computing*, 8(1):291–319, 2012.

[San10] R. Santhanam. Fighting perebor: New and improved algorithms for formula and QBF satisfiability. In *FOCS*, pages 183–192. IEEE Computer Society, 2010.

[Sub61] B. A. Subbotovskaya. Realizations of linear function by formulas using $+, \cdot, -$. *Doklady Akademii Nauk SSSR*, 136:3:553–555, 1961. In Russian.

[Yab54] S. V. Yablonskii. Realization of the linear function in the class of π-schemes. In *Dokl. Akad. Nauk SSSR*, volume 94, pages 805 – 806, 1954. In Russian.

A A Theorem of Linial, Mansour and Nisan

Theorem (Theorem 3.1, restated)**.** *Let $f : \{-1,1\}^n \to \{-1,1\}$ and assume there exists $t \in \mathbb{R}$ such that for all d, p, $\mathbf{Pr}_{\rho\sim\mathcal{R}_p}[\deg(f|_\rho) \ge d] \le (tp)^d$. Then for any k, $\mathbf{W}^{\ge k}[f] \le 2e \cdot e^{-k/(te)}$.*

Proof. For any $d \in \mathbb{N}$ and $p \in (0,1]$ we have

$$\begin{aligned}
\mathop{\mathbf{E}}_{\rho\sim\mathcal{R}_p}\left[\sum_{S:|S|\ge d} \widehat{f|_\rho}(S)^2\right] &= \sum_{k\ge d} \mathbf{W}^{=k}[f]\cdot \mathbf{Pr}[\mathrm{Bin}(k,p)\ge d] && \text{(Corollary 2.5)}\\
&\ge \sum_{k\ge d/p} \mathbf{W}^{=k}[f]\cdot \mathbf{Pr}[\mathrm{Bin}(k,p)\ge d]\\
&\ge \sum_{k\ge d/p} \mathbf{W}^{=k}[f]\cdot 1/2 && (\textit{median}(\mathrm{Bin}(k,p)) \ge \lfloor kp\rfloor \ge d,\ \text{[KB80]})\\
&= 1/2\cdot \mathbf{W}^{\ge d/p}[f]
\end{aligned}$$

Overall we got

$$\mathbf{W}^{\ge d/p}[f] \le 2\cdot \mathop{\mathbf{E}}_{\rho\sim\mathcal{R}_p}\left[\sum_{S:|S|\ge d} \widehat{f|_\rho}(S)^2\right] \le 2 \mathop{\mathbf{Pr}}_{\rho\sim\mathcal{R}_p}[\deg(f|_\rho)\ge d] \le 2(tp)^d. \tag{6}$$

Given k and t we choose $p := 1/(te)$ and $d := \lfloor kp\rfloor$. Substituting d and p in Equation (6) we get $\mathbf{W}^{\ge k}[f] \le 2\cdot e^{-\lfloor k/(te)\rfloor} \le 2e\cdot e^{-k/(te)}$. □

B Amplification of Approximate Degree

The proof in this section is essentially the same as the one in [BNRdW07]; we present it here for completeness.

Definition B.1. *For $q \in [-1,1]$ we say that x is a q-biased bit, denoted by $x \sim N_q$, if $\mathbf{Pr}[x=1] = \frac{1+q}{2}$ and $\mathbf{Pr}[x=-1] = \frac{1-q}{2}$. In other words, x is a random variable taking values from $\{-1,1\}$ with $\mathbf{E}[x] = q$.*

The next lemma connects the value of a polynomial representing a Boolean function on non-Boolean inputs with a product-measure distribution.

Lemma B.2. *Let $f : \{-1,1\}^n \to \mathbb{R}$ and let $p \in \mathbb{R}[x_1,\dots,x_n]$ be the unique multilinear polynomial agreeing with f on $\{-1,1\}^n$. Let $q_1,\dots,q_n \in [-1,1]$ then*

$$\mathop{\mathbf{E}}_{x_i\sim N_{q_i}}[f(x_1,\dots,x_n)] = p(q_1,\dots,q_n)$$

where the x_is are drawn independently.

Proof. We write $p(x_1,\dots,x_n) = \sum_{S\subseteq[n]} \hat{f}(S)\cdot\prod_{i\in S} x_i$. We first show the lemma for a single monomial:

$$\mathop{\mathbf{E}}_{x_i\sim N_{q_i}}\left[\prod_{i\in S} x_i\right] \underset{x_i \text{ are ind.}}{=} \prod_{i\in S}\mathop{\mathbf{E}}_{x_i\sim N_{q_i}}[x_i] = \prod_{i\in S} q_i\,.$$

By linearity of expectation we have:

$$\mathop{\mathbf{E}}_{x_i\sim N_{q_i}}[p(x_1,\dots,x_n)] = \mathop{\mathbf{E}}_{x_i\sim N_{q_i}}\left[\sum_{S\subseteq[n]}\hat{f}(S)\cdot\prod_{i\in S}x_i\right] = \sum_{S\subseteq[n]}\hat{f}(S)\cdot\prod_{i\in S}q_i = p(q_1,\dots,q_n)\,. \qquad \square$$

We now turn to prove Fact 2.10, restated next.

Fact B.3. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be Boolean function and let $0 < \epsilon < 1$ then:* $\widetilde{\deg}_\epsilon(f) \leq \widetilde{\deg}(f) \cdot \lceil 8 \cdot \ln(2/\epsilon) \rceil$.

Proof. Let m be some parameter we will set later. Take $\mathrm{MAJ}_m : \{-1,1\}^m \to \{-1,1\}$ to be the majority of m inputs, and denote by $p_{\mathrm{MAJ}} \in \mathbb{R}[x_1, \ldots, x_m]$ the multilinear polynomial agreeing with MAJ_m on $\{-1,1\}^m$. Let $q \in (0,1]$ (the case $q \in [-1,0)$ is similar), then by Lemma B.2 we have

$$p_{\mathrm{MAJ}}(q, q, \ldots, q) = \mathop{\mathbf{E}}_{x_i \sim N_q}[\mathrm{MAJ}_m(x_1, \ldots, x_m)] = \mathop{\mathbf{Pr}}_{x_i \sim N_q}\left[\sum_i x_i \geq 0\right] - \mathop{\mathbf{Pr}}_{x_i \sim N_q}\left[\sum_i x_i < 0\right] .$$

Let $X = \sum_i x_i$, then by Chernoff-Hoeffding bound we have

$$\mathbf{Pr}\left[X \geq 0\right] = \mathbf{Pr}\left[X - \mathbf{E}[X] \geq -q \cdot m\right] \;\geq\; 1 - e^{-(qm)^2/2m} \;=\; 1 - e^{-mq^2/2} ,$$

which implies

$$p_{\mathrm{MAJ}}(q, q, \ldots, q) \;\geq\; 1 - 2e^{-mq^2/2} . \tag{7}$$

By definition there exists a polynomial p of degree $\widetilde{\deg}(f)$ such that $p(x) \in [-4/3, -2/3]$ if $f(x) = -1$ and $p(x) \in [2/3, 4/3]$ if $f(x) = 1$. Take $p'(x) = \frac{p(x)}{4/3}$, then $p'(x) \in [1/2, 1]$ if $f(x) = 1$ and $p'(x) \in [-1, -1/2]$ if $f(x) = -1$. Consider the polynomial

$$g(x) = p_{\mathrm{MAJ}}(p'(x), p'(x), \ldots, p'(x)),$$

then $\deg(g) \leq \deg(p_{\mathrm{MAJ}}) \cdot \deg(p') = m \cdot \widetilde{\deg}(f)$. On the other hand, for x such that $f(x) = 1$ (the case where $f(x) = -1$ is analogous) we have $g(x) = p_{\mathrm{MAJ}}(q, q, \ldots, q)$ for some $q \in [1/2, 1]$. Since p_{MAJ} is monotone and using Equation (7), we have

$$1 \geq g(x) = p_{\mathrm{MAJ}}(q, \ldots, q) \geq p_{\mathrm{MAJ}}(1/2, \ldots, 1/2) \geq 1 - 2e^{-m/8} .$$

Picking $m = \lceil 8 \cdot \ln(2/\epsilon) \rceil$ completes the proof. □

ECCC ISSN 1433-8092

http://eccc.hpi-web.de