Shrinkage of De Morgan Formulae from Quantum Query Complexity

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 $f : \{-1,1\}^n \rightarrow \{-1,1\}$, setting each variable out of $x_1, \ldots, x_n$ 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], $A : \{-1,1\}^n \rightarrow \{-1,1\}$, which is in $P$, has formula size at least $\Omega(\frac{n^3}{\log^2 n \log \log n})$. This lower bound is tight up to the $\log^3 n$ factor, and is the best known lower bound for functions in $P$. In addition, the functions defined in [KRT13], $h_r : \{-1,1\}^n \rightarrow \{-1,1\}$, 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 $\frac{n^3}{r^2 \text{poly} \log n}$, for any parameter $r \leq n^{1/3}$.

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

The problem of \( P \) vs. \( \text{NC}^1 \) 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 \( \{x_1, \ldots, x_n, \neg x_1, \ldots, \neg x_n\} \) 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 \( f : \{-1,1\}^n \to \{-1,1\} \) 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 \( P \not\subseteq \text{NC}^1 \). \(^2\)

While 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 \( \text{NEXP} \)), polynomial lower bounds for \( P \) were achieved during the years. 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 \( \text{AC}^0 \). Subbotovskaya showed that the minimal formula size of a given function is shrunk by a factor of \( O(p^{1.5}) \) under a random restriction keeping each variable alive with probability \( p \), and fixing it to a uniformly chosen random bit otherwise. We call such restrictions \( p \)-random restrictions, and 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 \( \rho \). 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 \( \Omega(n^{1-o(1)}) \) where \( \Gamma \) is the shrinkage exponent of De Morgan formulae - the maximal constant such that any De Morgan formula is shrunk be a factor of \( O(p^\Gamma) \) under \( p \)-random restrictions. Impagliazzo and Nisan [IN93] showed \( \Gamma \geq 1.55 \); 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 \( p > 0 \),

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

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 [Lee07]). Applying a \( p \)-random restriction on the parity function yields a smaller parity function (or its negation) on \( k \) variables where \( k \sim \text{Bin}(n, p) \). By Khrapchenko’s argument, the formula size of the restricted function is \( \geq k^2 \), thus the expected formula size is at least \( \mathbb{E}[k^2] = p^2n^2 + p(1-p)n = \Omega\left(p^2L(f) + p\sqrt{L(f)}\right) \).

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. [GMWW13] suggested an information theoretical approach to further understand the formula size of this function.

---

\(^1\)We identify the truth values \( \text{true} \) and \( \text{false} \) with \(-1 \) and \( 1 \) respectively.

\(^2\)Here we think of the non-uniform version of \( \text{NC}^1 \): the class of languages \( L \subseteq \{-1,1\}^* \) such that for each length \( n \) there exists a Boolean formula \( F_n \) of size \( \text{poly}(n) \) which decides whether strings of length \( n \) are in the language.
Another recent line of work ([San10],[IMZ12],[KR13],[KRT13],[CKK+13],[CKS13]) 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 $h_r : \{-1,1\}^n \rightarrow \{-1,1\}$ such that any formula that computes this function on a fraction $\frac{1}{2} + 2^{-r}$ must be of size at least $n^{3-o(1)}$ 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 $p > 0$,

$$\mathbb{E}_{\rho \sim R_p} [L(f|\rho)] = O \left( p^2 L(f) + p \sqrt{L(f)} \right).$$

Note that both terms in Theorem 1.2 are essential as demonstrated by the parity function above. This improves the worst-case lower bound Håstad gave to Andreev’s function to $\Omega \left( \frac{n^3}{\log^2 n (\log \log n)^3} \right)$ immediately (following the proof of Theorem 8.1 in [Hås98]). 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 $r \leq n^{1/3}$ there is an explicit (computable in polynomial time) Boolean function $h_r : \{-1,1\}^n \rightarrow \{-1,1\}$ such that any formula of size $\frac{n^3}{r^2 \cdot \text{poly log}(n)}$ 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. Hoyer, 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 $\deg(f) = O(\sqrt{L(f)})$ is known – it might be interesting to find such.

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

\(^3\)The approximate degree of a function $f : \{-1,1\}^n \rightarrow \{-1,1\}$ is the minimal degree of a polynomial $p(x)$ such that $|p(x) - f(x)| \leq 1/3$ for all $x \in \{-1,1\}^n$. 


degree $\tilde{d} = O(\sqrt{L(f)} \log(1/\epsilon))$ which $\epsilon$-approximates $f$ pointwise. Using standard arguments this implies that the Fourier mass above degree $\tilde{d}$, i.e. $\sum_{|S| > \tilde{d}} \hat{f}(S)^2$, is at most $\epsilon$. 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

$$\forall d : \Pr_{\rho \sim R_p} [\deg(f|\rho) = d] \leq (4pm)^d . \quad (1)$$

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 $\exp(-10d)$. \(^5\)

We call such a property a “switching lemma” type property since the switching lemma 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

$$\forall d : \Pr_{\rho \sim R_p} [DT(f|\rho) \geq d] \leq (5pw)^d .$$

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 $p = O(1/\sqrt{L(f)})$.** Combining 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 $\ell$ to be $1/p^2$ and using linearity of expectation we get the required result for general $p$.

### 1.3 Related Work

The recent work of Impagliazzo and Kabanets [IK13] 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.

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

\(^5\)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.
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 $\lor$ or $\land$ 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 $\rho$ is a vector of length $n$ of elements from $\{0,1,*\}$. We denote by $f|_{\rho}$ the function $f$ restricted according to $\rho$ 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 $\rho_i$.

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 $1 - p^2$ set $\rho_i$ to be 0 and 1, respectively. We denote this distribution of restrictions by $R_p$.

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 $\hat{f}(S) = \mathbb{E}_x[f(x) \cdot \prod_{i \in S} x_i]$. Parseval’s equality states that $\sum_S \hat{f}(S)^2 = \mathbb{E}_x[f(x)^2] = 1$. 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

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

the Fourier weight at level $k$ of $f$. Similarly, we denote by $W^k[f] \triangleq \sum_{S \subseteq [n], |S| \geq k} \hat{f}(S)^2$. The following fact relates the Fourier coefficients of $f$ and of $f|_{\rho}$ where $\rho$ is a $p$-random restriction.

Fact 2.3 (Proposition 4.17,[O’D12]).

$$\mathbb{E}_{\rho \sim R_p} [\hat{f}_\rho(S)^2] = \sum_{U \subseteq [n]} \hat{f}(U)^2 \cdot \mathbb{P}_{\rho \sim R_p} \{i \in U : \rho(i) = *\} = |S|$$

Summing over all coefficients of size $d$, we get the following corollary.
Corollary 2.4.

\[
\mathbb{E}_{p \sim R_p} \left[ \sum_{S : |S| = d} \hat{f}_p(S)^2 \right] = \sum_{k=d}^{n} W^{=k} \Pr[\text{Bin}(k, p) = d]
\]

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.

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

Let \( p(y) = \sum_{T \subseteq [n]} a_T \cdot \prod_{i \in T} y_i \) be the unique multi-linear polynomial 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

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

It is clear from Equation (3) that \( \deg(p) = \deg(f) \), hence the definition of degree does not depend whether we are talking about the \( \{-1, 1\} \) or the \( \{0, 1\} \) representation of the function. Note that since \( \tilde{f} \) is Boolean, the coefficients \( a_T \) are integers, since 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 multi-linear 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 \( 2^{1-d} \) granular.

**Fact 2.5** (Granularity). Let \( f : \{-1, 1\}^n \to \{-1, 1\} \) with \( \deg(f) = d \), then \( \hat{f}(S) = k_S \cdot 2^{1-d} \) 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 \( 2^{1-d} \). Equation (3) gives \( a_T = (-2)^{|T| - 1} \sum_{S \supseteq T} \hat{f}(S) \). Multiplying both sides by \( (-2)^{d - |T|} \) we get

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

By the assumption on maximality of \( T \), all coefficients on the RHS except \( \hat{f}(T) \) are integer multiples of \( 2^{1-d} \), hence the RHS is not an integer. On the other hand, the LHS is an integer since \( a_T \) is an integer.

**Definition 2.6.** We define the sparsity of \( f : \{-1, 1\}^n \to \{-1, 1\} \) as

\[
\text{sparsity}(f) \overset{\Delta}{=} |\{S : \hat{f}(S) \neq 0\}|.
\]

**Corollary 2.7.** Let \( f : \{-1, 1\}^n \to \{-1, 1\} \) with \( \deg(f) = d \), then \( \text{sparsity}(f) \leq 2^{2d-2} \).

**Proof.** By Parseval, \( 1 = \sum_S \hat{f}(S)^2 \geq \text{sparsity}(f) \cdot (2^{1-d})^2 \).
Claim 2.8. Let \( \tilde{f} : \{0,1\}^n \rightarrow \{0,1\} \) be a Boolean function with \( \deg(\tilde{f}) = d \) then \( \tilde{f} \) can be written as

\[
\tilde{f}(x) = \sum_{i=1}^{\text{sparsity}(f)} g_i(x)
\]

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

Proof. Write \( \tilde{f}(x) = \sum_{T \subseteq [n]} a_T \prod_{j \in T} x_j \). By Equation (3) any \( T \subseteq [n] \) such that \( a_T \neq 0 \) is contained in some subset \( S \subseteq [n] \) for which \( \tilde{f}(S) \neq 0 \). Order the sets \( \{S : \tilde{f}(S) \neq 0\} \) according to some arbitrary order: \( \{S_1, \ldots, S_{\text{sparsity}(f)}\} \) and let

\[
g_i(x) = \sum_{T \subseteq S_i, \forall j \in i : T \nsubseteq S_j} a_T \cdot \prod_{j \in T} x_j.
\]

Then, by definition \( \tilde{f}(x) = \sum_{i=1}^{\text{sparsity}(f)} g_i(x) \). 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. \( \square \)

2.4 Approximate Degree

Let \( f : \{-1,1\}^n \rightarrow \{-1,1\} \). Given an \( \epsilon \) we define the \( \epsilon \)-approximate degree, denoted by \( \deg_\epsilon(f) \), as the minimal degree of a multi-linear polynomial \( p \) such that for all \( x \in \{-1,1\}^n \), \( |f(x) - p(x)| \leq \epsilon \). We denote \( \deg_{1/3}(f) \) by \( \deg(f) \).

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 2.9. Let \( f : \{-1,1\}^n \rightarrow \{-1,1\} \) be a Boolean function and let \( 0 < \epsilon < 1 \) then:

\[
\deg_\epsilon(f) \leq \deg(f) \cdot [8 \cdot \ln(2/\epsilon)].
\]

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

Fact 2.10. Let \( f : \{-1,1\}^n \rightarrow \{-1,1\} \) be a Boolean function, \( 0 < \epsilon < 1 \) and \( d = \deg_\epsilon(f) \), then \( \mathbf{W}^d(f) \leq \epsilon^2 \).

Proof. Let \( p \) be a polynomial of degree \( d \) which \( \epsilon \) approximates \( f \). Obviously \( \mathbf{E}_x[(f(x) - p(x))^2] \leq \epsilon^2 \).

Let \( q \) be the best \( \ell_2 \) approximation of \( f \) by a degree \( d \) polynomial, namely the polynomial of degree \( d \) which minimizes \( \|f - q\|_2^2 \equiv \mathbf{E}_x[(f(x) - q(x))^2] \). 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 \( d \):

\[
q(x) = \sum_{S \subseteq [n] : |S| \leq d} \hat{f}(S) \cdot \chi_S(x).
\]

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 2.11 ([BBC+01, HLS07, Rei11]). There exists a universal constant \( C_1 \geq 1 \) such that for any \( f : \{-1,1\}^n \rightarrow \{-1,1\} \) we have \( \deg(f) \leq C_1 \cdot \sqrt{L(f)} \).
Theorem 3.2. Let \( t > 0 \) (\cite{LMN93}, restated slightly) a proof of this theorem in Appendix A.

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

Claim 2.12. There exists a constant \( C > 0 \) such that for any \( f : \{-1,1\}^n \to \{-1,1\} \) and \( k \in \mathbb{N} \),

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

Proof. Let \( t = \frac{k}{C \sqrt{L(f)}} \) where \( C \) is some constant we shall set later. We prove that \( W^{\geq k}[f] \leq e \cdot e^{-t} \).

Assume without loss of generality that \( t \geq 1 \) or else the claim is trivial since \( W^{\geq k}[f] \leq 1 \leq e \cdot e^{-t} \).

Put \( \epsilon = e^{-t/2} \), and combine Theorem 2.11 and Fact 2.9 to get

\[
\overline{\deg}_t(f) \leq \sqrt{L(f)} \cdot C_1 \cdot |8 \ln(2/\epsilon)| = \sqrt{L(f)} \cdot C_1 \cdot [4t + 8 \ln(2)] \leq \sqrt{L(f)} \cdot C_1 \cdot 11t.
\]

Using Fact 2.10 we get \( W^{> \sqrt{L(f)} \cdot C_1 \cdot 11t}[f] \leq e^{-t} \). Hence \( W^{\geq \sqrt{L(f)} \cdot C_1 \cdot 12t}[f] \leq e^{-t} \). Setting \( C := C_1 \cdot 12 \) completes the proof.

\[\square\]

2.5 The Generalized Binomial Theorem

Theorem 2.13 (The generalized binomial theorem). Let \( |x| < 1 \), and \( s \in \mathbb{N} \) then

\[
\frac{1}{(1 - x)^s} = \sum_{k=0}^{\infty} \binom{s + k - 1}{s - 1} \cdot x^k.
\]

Rearranging this equality one get the following corollary:

Corollary 2.14. Let \( |x| < 1 \), and \( m \in \mathbb{N} \cup \{0\} \) then \( \sum_{n=m}^{\infty} x^n \cdot \binom{n}{m} = \frac{x^m}{(1 - x)^{m+1}} \).

Proof. By the generalized binomial theorem \( \frac{x^m}{(1 - x)^{m+1}} = \sum_{k=0}^{\infty} \binom{m+k}{m} \cdot x^{m+k} \). The RHS can be rewritten as \( \sum_{n=m}^{\infty} \binom{n}{m} \cdot x^n \), which completes the proof.

\[\square\]

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

2. A “switching lemma” type property with parameter \( t' \): \( \forall p, d : \Pr_{p \sim \mathcal{R}_p}[\deg(f|_p) \geq d] \leq (t'p)^d \).

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

Theorem 3.1 (\cite{LMN93}, restated slightly). Let \( f : \{-1,1\}^n \to \{-1,1\} \) and assume the existence of \( t > 0 \) such that for all \( d \in \mathbb{N}, p \in (0,1) \), \( \Pr_{p \sim \mathcal{R}_p}[\deg(f|_p) \geq d] \leq (tp)^d \). Then for any \( k \), \( W^{\geq k}[f] \leq 2e \cdot e^{-k/t} \).

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 \( k \), \( W^{\geq k}[f] \leq C \cdot e^{-k/t} \). Let \( p \) be a \( p \)-random restriction, then for all \( d \), \( \Pr[\deg(f|_p) = d] \leq \frac{C}{4}(4pt)^d \).
**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 2.4). 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 $E_{\rho} \left[ \sum_{S:|S|=d} \hat{f}_{\rho}(S)^2 \right]$ from below and above showing

$$E_{\rho} \left[ \sum_{S:|S|=d} \hat{f}_{\rho}(S)^2 \right] \geq \Pr[\deg(f|_{\rho}) = d] \cdot 4^{1-d} \quad (4)$$

and

$$E_{\rho} \left[ \sum_{S:|S|=d} \hat{f}_{\rho}(S)^2 \right] \leq C (pt)^d. \quad (5)$$

Combining the two estimates will complete the proof.

We begin by proving Equation (4). Conditioning on the event that $\deg(f|_{\rho}) = d$, Fact 2.5 implies that any nonzero Fourier coefficient of $f|_{\rho}$ is of magnitude $\geq 2^{1-d}$. Hence, $\sum_{S:|S|=d} \hat{f}(S)^2 \geq 4^{1-d}$, and we get

$$E_{\rho} \left[ \sum_{S:|S|=d} \hat{f}_{\rho}(S)^2 \right] \geq \Pr[\deg(f|_{\rho}) = d] \cdot E_{\rho} \left[ \sum_{S:|S|=d} \hat{f}_{\rho}(S)^2 \right] \geq \Pr[\deg(f|_{\rho}) = d] \cdot 4^{1-d}.$$

Next, we turn to prove Equation (5).

$$E_{\rho} \left[ \sum_{S:|S|=d} \hat{f}_{\rho}(S)^2 \right] = \sum_{k \geq d} W^k[f] \cdot \binom{k}{d} \cdot p^d \cdot (1-p)^{k-d} \quad \text{(Corollary 2.4)}$$

$$= \sum_{k \geq d} \left( W^{\geq k}[f] - W^{\geq k+1}[f] \right) \cdot \binom{k}{d} \cdot p^d \cdot (1-p)^{k-d}$$

$$= \left( \frac{p}{1-p} \right)^d \sum_{k \geq d} \left( W^{\geq k}[f] - W^{\geq k+1}[f] \right) \cdot \binom{k}{d} \cdot (1-p)^k$$

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

$$E_{\rho} \left[ \sum_{S:|S|=d} \hat{f}_{\rho}(S)^2 \right] = \left( \frac{p}{1-p} \right)^d \sum_{k \geq d} W^{\geq k}[f] \cdot \left( \binom{k}{d} \cdot (1-p)^k - \binom{k-1}{d} \cdot (1-p)^{k-1} \right)$$

$$\leq \left( \frac{p}{1-p} \right)^d \sum_{k \geq d} W^{\geq k}[f] \cdot \left( \binom{k}{d} \cdot (1-p)^k - \binom{k-1}{d} \cdot (1-p)^{k} \right)$$

$$= \left( \frac{p}{1-p} \right)^d \sum_{k \geq d} W^{\geq k}[f] \cdot \binom{k-1}{d-1} \cdot (1-p)^k.$$
Let $a := e^{-1/t}$. The assumption on the Fourier tails of $f$, $W^k[f] \leq C \cdot a^k$, gives

$$
E_{\rho} \left[ \sum_{|S| = d} \hat{f}_{\rho}(S)^2 \right] \leq \left( \frac{p}{1-p} \right)^d \sum_{k \geq d} C(a(1-p))^k \cdot \binom{k-1}{d-1} \\
= \left( \frac{p}{1-p} \right)^d \cdot C(a(1-p)) \cdot \sum_{k \geq d} (a(1-p))^{k-1} \cdot \binom{k-1}{d-1} \\
= \left( \frac{p}{1-p} \right)^d \cdot C(a(1-p)) \cdot \sum_{k' \geq d-1} (a(1-p))^{k'} \cdot \binom{k'}{d-1}
$$

Next we use Corollary 2.14 with $x := a(1-p)$ and $m := d-1$ to get

$$
E_{\rho} \left[ \sum_{|S| = d} \hat{f}_{\rho}(S)^2 \right] \leq \left( \frac{p}{1-p} \right)^d \cdot C(a(1-p)) \cdot \frac{(a(1-p))^{d-1}}{(1-a(1-p))^d} = C \left( \frac{ap}{1-a(1-p)} \right)^d \leq C \left( \frac{ap}{1-a} \right)^d
$$

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

$$
E_{\rho} \left[ \sum_{|S| = d} \hat{f}_{\rho}(S)^2 \right] \leq C \left( \frac{p}{1-1/a} \right)^d = C \left( \frac{p}{e^{1/t}-1} \right)^d \leq C (pt)^d,
$$

where the last inequality follows since $e^x - 1 \geq x$ for any $x \geq 0$. \hfill \Box

## 4 Degree vs. Formula Size

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

**Fact 4.1** ([Lee07]). $L(\text{PARITY}_m) \leq 9/8 \cdot m^2$. Furthermore, if $m = 2^k$ for some integer $k$, then $L(\text{PARITY}_m) = m^2$.

**Claim 4.2.** Let $\tilde{f} : \{0,1\}^n \to \{0,1\}$ such that $\deg(\tilde{f}) = d$, then $L(\tilde{f}) \leq 1/8 \cdot 32^d$.

**Proof.** According to Claim 2.8, $\tilde{f}$ can be written as $\sum_{i=1}^{4d-1} g_i(x)$, where the functions $g_i(x)$ take integer values, and each of them depends on at most $d$ variables. Since $\tilde{f}(\bar{x}) \in \{0,1\}$ we may perform all operations modulo 2 and get $\tilde{f}(x) = \bigoplus_{i=1}^{4d-1} h_i(x)$, where $h_i(x) = g_i(x) \mod 2$. Taking a formula for the parity of $m = 4^{d-1}$ 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 $\leq L(\text{PARITY}_m) \cdot 2^{d+1} = 16^{d-1} \cdot 2^{d+1} = 1/8 \cdot 32^d$. \hfill \Box

## 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 $p \leq \frac{1}{C \sqrt{L(f)}}$ the following hold. Let $\rho$ be a $p$-random restriction, then $E_{\rho}[L(f|\rho)] = O(p \cdot \sqrt{L(f)})$.

In particular, in this regime of parameters, $E_{\rho}[L(f|\rho)] = O(1)$.  

9
Proof of Claim 5.1. From Claim 2.12, there exists a constant $C > 0$ such that

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

This implies, using Theorem 3.2, that $\Pr_{\rho \sim \mathcal{R}_p}[\deg(f|\rho) = d] \leq \frac{1}{4} \cdot \left(4pC\sqrt{L(f)}\right)^d \leq \left(4pC\sqrt{L(f)}\right)^d$.

Using Claim 4.2, if $\deg(f|\rho) = d$ then $L(f|\rho) \leq 32^d$. For $p \leq 1 \over 644C\sqrt{L(f)}$ we get

$$
\mathbb{E}_{\rho \sim \mathcal{R}_p}[L(f|\rho)] = \sum_{d=0}^{\infty} \Pr_{\rho}[\deg(f|\rho) = d] \cdot \mathbb{E}_\rho[L(f|\rho)|\deg(f|\rho) = d]
$$

$$
= \sum_{d=1}^{\infty} \Pr_\rho[\deg(f|\rho) = d] \cdot \mathbb{E}_\rho[L(f|\rho)|\deg(f|\rho) = d] \, (\deg(f|\rho) = 0 \implies L(f|\rho) = 0)
$$

$$
\leq \sum_{d=1}^{\infty} \left(4pC\sqrt{L(f)}\right)^d \cdot 32^d
$$

$$
\leq O(p\sqrt{L(f)}) \cdot \sum_{d=1}^{\infty} \left(4pC\sqrt{L(f)}\right)^{d-1} \cdot 32^{d-1}
$$

$$
\leq O(p\sqrt{L(f)}) \cdot \sum_{d=1}^{\infty} (1/64)^{d-1} \cdot 32^{d-1}
$$

$$
= O(p\sqrt{L(f)}). \qedhere
$$

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 decomposition of Impagliazzo, Meka and Zuckerman [IMZ12] to establish this reduction.\footnote{Another approach to prove the general case is to follow H˚astad 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.}

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 $s \geq \ell$ leaves can be decomposed into at most $6s/\ell$ subtrees, each with at most $\ell$ 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 $X = \{x_1, \ldots, x_n\}$, and $\ell \in \mathbb{N}$ be some parameter. Then, there exist $m \leq O(L(F)/\ell)$ formulae over $X$, denoted by $G_1, \ldots, G_m$, each of size at most $\ell$, and there exists a read-once formula $F'$ of size $m$ such that $F'(G_1(x), \ldots, G_m(x)) = F(x)$ for all $x \in \{-1,1\}^n$.

**Proof.** Consider the decomposition promised by Lemma 6.1 with parameter $\ell$. Let $T_1, \ldots, T_{m'}$ be the subtrees in this decomposition where $m' \leq 6n/\ell$. We will show by induction on $m'$, that one
can construct a read-once formula $F'$ of size $m \leq 6m'$ alongside $m$ sub-formulae $G_1, \ldots, G_m$ of size $\ell$ such that $F \equiv F'(G_1, \ldots, G_m)$. 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 $t_1$ ($t_1'$, 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 $T' = T_2 \circ \circ T_3'$, for $\circ \circ \in \{\land, \lor\}$, 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 $T_2' = (T_2'|z_2=\text{false}) \lor (T_2'|z_2=\text{true} \land z_2)$. This follows since $T_2'$ is monotone in $z_2$: if $T_2'|z_2=\text{false} = \text{true}$ then $T_2' = \text{true}$, otherwise $T_2' = \text{true}$ only if both $T_2'|z_2=\text{true}$ and $z_2$ are $\text{true}$. Same goes for $T_3'$, and we get

$$T' \equiv \left((T_2'|z_2=\text{false}) \lor (T_2'|z_2=\text{true} \land z_2)\right) \circ \circ \left((T_3'|z_3=\text{false}) \lor (T_3'|z_3=\text{true} \land z_3)\right).$$

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

$$T \equiv T'|z=\text{false} \lor (T'|z=\text{true} \land z) \circ \circ \left((T_2'|z_2=\text{false}) \lor (T_2'|z_2=\text{true} \land z_2)\right) \circ \circ \left((T_3'|z_3=\text{false}) \lor (T_3'|z_3=\text{true} \land z_3)\right).$$

Note that the RHS of the above can be written as $F''(G_1(x), \ldots, G_6(x), z_2, z_3)$ where $F''$ is read-once and $G_1(x), \ldots, G_6(x)$ are formulae of size $\ell$, 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 $F_2'(G_1^2(x), \ldots, G_6^{2m_2}(x))$ where $F_2'$ is read-once and $G_i^2(x)$ are formulae of size at most $\ell$. Similarly for $T_3$. We thus get that

$$F(x) = F''(G_1(x), \ldots, G_6(x), F_2'(G_1^2(x), \ldots, G_6^{2m_2}(x)), F_3'(G_1^3(x), \ldots, G_6^{3m_3}(x))).$$

Rearranging the RHS, we get a read-once formula of size $m \leq 6 + 6m_2 + 6m_3 = 6m'$ alongside $m$ sub-formulae, each of size $\ell$, 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 $E_{\rho \sim \mathcal{P}_p}[L(f|\rho)] = O\left(p^2 L(f) + p \sqrt{L(f)}\right)$.  

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

$$E_{\rho}[L(f|\rho)] \leq E_{\rho}\left[\sum_{i=1}^m L(g_i|\rho)\right] \leq m \cdot O(p \cdot \sqrt{\ell}) = m \cdot O(1) = O(p^2 \cdot L(f)).$$
7 Open Ends

An interesting open question raised by H˚astad in [H˚as98] 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 \( \mathbb{E}_{\rho \sim \mathcal{R}_p}[L(f|_\rho)] = 1 \), and then use the tree decomposition to argue for \( p \geq p_c \) (note that the decomposition done in Section 6 respects monotonicity). Hence, in order to show \( \Gamma \) 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.

References


A A Theorem of Linial, Mansour and Nisan

Theorem (Theorem 3.1, restated). Let \( f : \{-1,1\}^n \rightarrow \{-1,1\} \) and assume there exists \( t \in \mathbb{R} \) such that for all \( d, p \), \( \Pr_{\rho \sim \mathcal{R}_p}[\deg(f_\rho) \geq d] \leq (tp)^d \). Then for any \( k, W^{\geq k}[f] \leq 2e \cdot e^{-k/(te)} \).

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

\[
E_{\rho \sim \mathcal{R}_p} \left[ \sum_{S : |S| \geq d} \hat{f}_\rho(S)^2 \right] = \sum_{k \geq d} W^{=k}[f] \cdot \Pr[\text{Bin}(k, p) \geq d] \\
\geq \sum_{k \geq d/p} W^{=k}[f] \cdot \Pr[\text{Bin}(k, p) \geq d] \\
\geq \sum_{k \geq d/p} W^{=k}[f] \cdot 1/2 \quad (\text{median(\text{Bin}(k, p))} \geq \lfloor kp \rfloor \geq d, [KB80]) \\
= \frac{1}{2} \cdot W^{\geq d/p}[f]
\]

Overall we got

\[
W^{\geq d/p}[f] \leq 2 \cdot E_{\rho \sim \mathcal{R}_p} \left[ \sum_{S : |S| \geq d} \hat{f}_\rho(S)^2 \right] \leq 2 \Pr_{\rho \sim \mathcal{R}_p}[\deg(f_\rho) \geq d] \leq 2(tp)^d. \quad (6)
\]

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

B Amplification of Approximate Degree

Definition B.1. For \( q \in [-1,1] \) we say that \( x \) is a \( q \)-biased bit, denoted by \( x \sim N_q \), if \( \Pr[x = 1] = \frac{1+q}{2} \) and \( \Pr[x = -1] = \frac{1-q}{2} \). In other words, \( x \) is a random variable taking values from \( \{-1,1\} \) with \( E[x] = q \).
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 \( p \in \mathbb{R}[x_1, \ldots, x_n] \) be the unique multi-linear polynomial agreeing with \( f \) on \( \{−1, 1\}^n \). Let \( q_1, \ldots, q_n \in [−1, 1] \) then

\[
\mathbb{E}_{x_i \sim N_{q_i}} [f(x_1, \ldots, x_n)] = p(q_1, \ldots, q_n)
\]

where the \( x_i \)'s are drawn independently.

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

\[
\mathbb{E}_{x_i \sim N_{q_i}} \left[ \prod_{i \in S} x_i \right] = \prod_{i \in S} \mathbb{E}_{x_i \sim N_{q_i}} [x_i] = \prod_{i \in S} q_i .
\]

By linearity of expectation we have:

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

We now turn to prove Fact 2.9, restated next.

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

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

\[
p_{\text{MAJ}}(q, q, \ldots, q) = \mathbb{E}_{x_i \sim N_q} \left[ \text{MAJ}_m(x_1, \ldots, x_m) \right] = \Pr_{x_i \sim N_q} \left[ \sum_i x_i \geq 0 \right] - \Pr_{x_i \sim N_q} \left[ \sum_i x_i < 0 \right] .
\]

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

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

which implies

\[
p_{\text{MAJ}}(q, q, \ldots, q) \geq 1 - 2e^{-mq^2/2} . \tag{7}
\]

By definition there exists a polynomial \( p \) of degree \( \deg(f) \) 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 \( p'(x) = p(x)/4^{m/3} \), 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

\[
g(x) = p_{\text{MAJ}}(p'(x), p'(x), \ldots, p'(x)) ,
\]

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

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

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