Electronic Colloquium on Computational Complexity, Report No. 179 (2016)

Computing Requires Larger Formulas than Approximating

Avishay Tal *

Abstract

A de Morgan formula over Boolean variables $x_1, \ldots, x_n$ is a binary tree whose internal nodes are marked with AND or OR gates and whose leaves are marked with variables or their negation. We define the size of the formula as the number of leaves in it. Proving that some explicit function (in **P** or **NP**) requires large formula is a central open question in computational complexity. While we believe that some explicit functions require exponential formula size, currently the best lower bound for an explicit function is the $\widetilde{\Omega}(n^3)$ lower bound for Andreev's function [And87, Hås98, Tal14].

In this work, we show how to trade average-case hardness in exchange for size. More precisely, we show that if a function f cannot be computed correctly on more than $\frac{1}{2} + 2^{-k}$ of the inputs by any formula of size at most s, then computing f exactly requires formula size at least $\widetilde{\Omega}(k) \cdot s$. The proof relies on a result from quantum query complexity by Reichardt [Rei11]. Due to the work of Impagliazzo and Kabanets [IK16], this tradeoff is essentially tight.

As an application, we improve the state of the art lower bounds for explicit functions by a factor of $\widetilde{\Omega}(\log n)$.

Additionally, we present candidates for explicit simple functions that we believe have formula complexity $\widetilde{\Omega}(n^4)$. In particular, one such function was studied in [GT16] and is given by $F(x, y, z) = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i y_j z_{i+j} \mod 2$. Based on our main theorem, we give non-trivial super-quadratic lower bounds for these functions.

[*] Institute for Advanced Study, Princeton, NJ. Email: avishay.tal@gmail.com. Supported by the Simons Collaboration on Algorithms and Geometry, and by the National Science Foundation grant No. CCF-1412958.

ISSN 1433-8092

1 Introduction

Does $\mathbf{P} = \mathbf{NC^1}$? Can any computational task be perfectly parallelized? The answer is still unknown. The question can be rephrased as finding a function in $\mathbf{P}$ that does not have a de Morgan formula (a binary tree with AND and OR gates on internal nodes and variables or their negations on the leaves) of polynomial size.

Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. The (de Morgan) formula size of f, denoted by $L(f)$, is the size of the smallest de Morgan formula that computes f. The study of formula lower bounds dates back to 1961 with the work of Subbotovskaya [Sub61] that proved an $\Omega(n^{1.5})$ lower bound on the formula size of the parity function. Khrapchencko[Khr71] improved the lower bound for the parity function to n^2, which is tight. Prior to this work, the best known formula lower bound for an explicit function is the $\Omega\left(\frac{n^3}{\log^2 n \cdot \log\log n}\right)$ lower bound for Andreev's function, $A : \{0,1\}^n \to \{0,1\}$. This is the result of a long line of research [And87, PZ93, IN93, Hås98, Tal14]. The n^3 barrier (or cubic-barrier) in formula lower bounds has stood for many years now, despite efforts to break it [KRW95, GMWW14, DM16].

The recent study of average-case hardness of formulas [KR13, IMZ12, KRT13] managed to almost match the known worst-case bounds. We say a Boolean function $f : \{0,1\}^n \to \{0,1\}$ is $1/2 + \varepsilon$ hard for formulas of size s if any de Morgan formula of size at most s agrees with f on at most $1/2 + \varepsilon$ fraction of the inputs (under the uniform distribution). [San10] showed that the parity function is $1/2 + \exp(-\Omega(n))$ hard for linear-size formulas. This was recently extended by Impagliazzo and Kabanets [IK13], who showed that the parity function is $1/2 + \exp(-n^2/s^{1+o(1)})$ hard for formulas of size s. Their result is tight up to the $o(1)$ factor. Komargodski et al. [KRT13] presented for any parameter $r \le n^{1/3}$ an explicit function $\mathrm{KRT}_r : \{0,1\}^{O(n)} \to \{0,1\}$ which is $1/2 + 2^{-r}$ hard for size $\frac{n^3}{r^2 \cdot \mathrm{polylog}(n)}$. In their paper, Komargodski et al. asks whether this trade-off between the size of the formula and the approximation quality is necessary. More specifically, they posed the challenge to prove $1/2 + \exp(-n^{\Omega(1)})$ hardness for size $n^{3-o(1)}$ for an explicit function. Though the challenge is natural, no implications of it were given in [KRT13].

In this work, we show that meeting this challenge would imply breaking the cubic-barrier in formula lower bounds. We establish the following reduction.

Theorem 1.1 (Main Theorem). *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. If f is $1/2 + 2^{-r}$ hard for formulas of size s, then $L(f) \ge \Omega\left(\min\{s^2, \frac{sr}{\log r}\}\right)$.*

As a motivating choice of parameters, Theorem 1.1 implies that if f is $1/2 + \exp(-\Omega(n))$ hard for size $n^{3-o(1)}$, then $L(f) \ge n^{4-o(1)}$.

We note that Theorem 1.1 easily implies that the formula size of the parity function on n variables, denoted $\oplus_n$, is at least $\Omega(n^2)$. Indeed, it is easy to see that any formula of size at most $s := n - 1$ agrees with parity on exactly 1/2 of the inputs (and not more than that). Thus, the parity on n variables is $1/2 + 2^{-n^2}$ hard for size $s = n - 1$, and Theorem 1.1 gives $L(\oplus_n) \ge \Omega\left(\min\{s^2, \frac{sn^2}{\log n^2}\}\right) \ge \Omega(n^2)$.

From Average-Case Hardness to Worst-Case Hardness! Our reduction is unusual in its nature. Most hardness reductions try to get from worst-case hardness to average-case hardness. We do the opposite. This seems trivial at first glance, since any average-case hard function is also worst-case hard. However, we gain in the size/complexity of the device. Thus, we view our reduction as a size-amplification reduction. We trade the average-case hardness in exchange for a better size lower bound. Most reductions in hardness-amplification reduce the size/complexity of the device being fooled (e.g. Yao's XOR lemma). This is due to a contrapositive argument that given a device that violates the conclusion (e.g., approximates the function) builds a bigger device that violates the assumption (e.g., computes the function on any input). Our reduction shrinks the device. Furthermore, the reduction is a white-box reduction: using the formula structure, a new smaller formula is built that approximates the bigger formula.

The reduction is in the spirit of the discriminator lemma of Hajnal et al. [HMP+93] that shows that average-case lower bounds for thresholds circuits of depth d implies worst-case lower bounds for thresholds circuits of depth $d+1$.

1.1 Proof Technique

Our proof is based on the beautiful result from quantum query complexity:

Theorem 1.2 ([Rei11])**.** *Let F be a de Morgan formula of size s on variables $x_1, \dots, x_n$. Then, there is a polynomial $p \in \mathbb{R}[x_1, \dots, x_n]$ of degree $O(\sqrt{s})$, such that for any $x \in \{0,1\}^n$, the value $p(x)$ is in the range $[F(x) - 1/6, F(x) + 1/6]$.*

Theorem 1.2 gives the very strong notion of approximation in ℓ_∞. That is, the polynomial is accurate up to small constant error in *every* point in the hypercube. Theorem 1.2 is the conclusion of long line of work in quantum query complexity [BBC+01, LLS06, HLS07, FGG08, Rei09, ACR+10, RS12]. It demonstrates the quantum method, giving a quantum based proof for a classical theorem.

Theorem 1.1 is proved via a contrapositive argument. We show that for any formula F of size s and any $k \leq \sqrt{s}$ there exists a formula F' of size $O(s/k)$ that agrees with F on at least $1/2 + 2^{-O(k \cdot \log k)}$ of the inputs. In other words, any de Morgan formula may be approximated by smaller formulas.

Decomposing F. The proof first decomposes a formula F into a top formula F' with $O(k^2)$ input leaves, where each leaf of F' is fed by a subformula of size $O(s/k^2)$. The correctness of this step was proved in [Tal14].

Approximating F' by a polynomial. The uniform distribution on the inputs to F induces some distribution on the inputs to F'. We approximate the top formula F' under this induced distribution. To do so, we apply Theorem 1.2 and get a polynomial p of degree $O(k)$ that approximates F'. We crucially use the fact that Theorem 1.2 gives a point-wise approximation, as this approximation is good with respect to any distribution.

Approximating F' by a small-Parity. Next, using the multilinear expansion of p we show that F' is $1/2 + 2^{-O(k \cdot \log k)}$ approximated (under the induced distribution) by a parity function of size $O(k)$ in the inputs of F'. However, note that the formula complexity of a parity on $O(k)$ inputs is $O(k^2)$, which is the same as the size of F'. It seems that we got nowhere with this argument. However, this is not the case.

Approximating F' by a small-AND. We use the fact that a parity of $O(k)$ inputs may be written as a linear combination of AND-type functions (i.e., an AND-function up to negations to the inputs and output) on $O(k)$ inputs. We deduce that some AND-type function on $O(k)$ inputs has agreement $1/2 + 2^{-O(k \cdot \log k)}$ (under the induced distribution) with F'. Thus, an AND-type function of $O(k)$ of the subformulas of size $O(s/k^2)$ has agreement $1/2 + 2^{-O(k \cdot \log k)}$ with F under the uniform distribution. This completes the proof, as the AND of $O(k)$ subformulas of size $O(s/k^2)$ is a formula of size $O(s/k)$.

1.2 Applications

A function harder than Andreev's function. We give a lower bound for an explicit function which is provenly $\widetilde{\Omega}(\log n)$ harder than the Andreev's function. For the explicit function suggested by Komargodski and Raz [KR13][1] , KR : $\{0,1\}^n \to \{0,1\}$, we give a lower bound of

$$L(\mathrm{KR}) \geq \Omega\left(\frac{n^3}{\log n \cdot (\log\log n)^2}\right) .$$

Prior to this work, the best known formula size lower bounds were given to Andreev's function, $A : \{0,1\}^n \to \{0,1\}$. It was established in [Tal14] that $L(A) \geq \Omega(\frac{n^3}{\log^2 n \cdot \log\log n})$. This lower bound is tight for Andreev's function, up to a factor of $\log\log n$, as $L(A) \leq O(\frac{n^3}{\log^2 n})$ via a straight-forward formula. Both the upper and lower bound for Andreev's function are subsumed by the lower bound for KR by a factor of at least $\Omega(\log n/(\log\log n)^2)$.

Non-trivial lower bounds for random t-linear functions. In Section 4.2 we study the formula complexity of a random t-linear function, for any constant t. That is, we study functions of the form $f : (\{0,1\}^n)^t \to \{0,1\}$ which are $\mathbb{F}_2$-linear in each of their t-blocks of input. For example, the case $t = 2$ corresponds to bilinear functions of the form $f(x,y) = \sum_{i=1}^n \sum_{j=1}^n A_{i,j} x_i y_j$ and is naturally associated with a matrix $A \in \mathbb{F}_2^{n\times n}$. While a simple counting argument ensures that most t-linear functions require size $\widetilde{\Omega}(n^t)$, the straight-forward upper bound is quadratically bigger, $O(n^{2t} \cdot t)$ (recall that the addition of m bits modulo 2 requires de Morgan formula of size $\Omega(m^2)$.)

We believe that the upper bound is tight. Using our main theorem we are able to deduce the non-trivial $\widetilde{\Omega}(n^{t+1})$ lower bound on the formula size of most t-linear functions. Using a result of Kaufman, Lovett and Porat [KLP12], we show that most t-linear functions are $1/2 + \exp(-\Omega(n))$ hard for size $\widetilde{\Omega}(n^t)$, which is then converted by our main theorem to a $\widetilde{\Omega}(n^{t+1})$ worst-case lower bound. We are unaware of any other proof for this fact that does not use our reduction.

We then move to convert such bounds to bounds on explicit functions. We give three explicit constructions:

- An explicit $(t+1)$-linear function based on Andreev's function, whose formula complexity is $\widetilde{\Omega}(n^{3-1/t})$.
- The explicit 3-linear function $f(x,y,z) = \sum_i \sum_j x_i y_j z_{i+j} \mod 2$, based on random Hankel matrices [GW13, GT16] whose formula complexity is $\widetilde{\Omega}(n^{2.4})$.
- An explicit 4-linear function based on 2^{-n}-biased matrices [GT16], whose formula complexity is $\widetilde{\Omega}(n^{2.5})$.

All these lower bounds are proved using the $\widetilde{\Omega}(n^{t+1})$ worst-case lower bound for random t-linear functions. We further conjecture that all three constructions have formula complexity $\widetilde{\Omega}(n^{3+\Omega(1)})$; We give some evidence for the latter two from the rigidity bounds in [GT16].

Direct Products and Sums Similar to the standard reductions in hardness-amplification we wish to show that if a function is somewhat hard for small formulas than taking the XOR of many copies of it is very hard for small formulas. Such a black-box reduction is not known with good parameters.

[1] The function is a variant of Andreev's function that combines an error-correcting code in the construction.

Since formulas cannot recycle computation it makes sense to conjecture a strong direct sum and product theorems, as we do in Section 4.3. As a corollary of our main theorem we show that such natural conjecture implies another conjecture by [GMWW14] (which is a special case of the KRW conjecture [KRW95]).

1.3 Comparison to Other Works in Terms of Usage of Theorem 1.2

In [Tal14], the author used Theorem 1.2 to give a new proof for Håstad's result [Hås98], showing that the shrinkage exponent of de Morgan formula is 2. Namely, it is shown that under p-random restrictions, fixing each variable to a randomly chosen constant with probability $1-p$ and keeping it alive otherwise, every de Morgan formula shrinks by a factor of $O(p^2)$ in expectation.

There, the already known shrinkage result was given a new proof using different techniques - namely, the quantum method in addition to Fourier analysis. Furthermore, the new proof shaved log factors from the previous proof of Håstad and gave the tight $O(p^2)$ shrinkage factor.

In the first step of the proof in [Tal14], the strong ℓ_∞ approximation, ensured by Theorem 1.2, is replaced with a weaker ℓ_2-approximation guarantee, which is then used to deduce the result. Indeed, the proof does not use all the power of this "sledgehammer".

In Theorem 1.1, we are relying on the fact that the approximation is in ℓ_∞, and thus holds with respect to any distribution on the inputs to F'. Moreover, using Theorem 1.1, we deduce results that have no classical proof: the formula lower bounds for the function KR, and the $\widetilde{\Omega}(n^{t+1})$ formula lower bound for most t-linear functions. We leave as an open problem to give a classical proof for Theorem 1.2.

2 Preliminaries

We start with some general notation. We denote by $[n]$ the set of numbers $\{1,2,\ldots,n\}$. We denote by $\binom{n}{\leq k} = \binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{k}$. We denote by $\mathcal{U}_k$ the uniform distribution over $\{0,1\}^k$. For a distribution $\mathcal{D}$ we denote by $x \sim \mathcal{D}$ a random element sampled according to $\mathcal{D}$. For a finite set X, we denote by $x \in_R X$ a random element sampled according to the uniform distribution over X.

For two functions $f\colon \{0,1\}^n \to \{0,1\}$ and $g\colon \{0,1\}^m \to \{0,1\}$, we denote by $f \circ g\colon \{0,1\}^{nm} \to \{0,1\}$ the (Boolean function) composition of f and g, defined by

$$f(x_{1,1}, x_{1,2}, \ldots, x_{n,m}) \triangleq f\left(g(x_{1,1}, \ldots, x_{1,m}), g(x_{2,1}, \ldots, x_{2,m}), \ldots, g(x_{n,1}, \ldots, x_{n,m})\right) ,$$

for all $x \in \{0,1\}^{n \times m}$. In words, $f \circ g$ views the input as an n-by-m matrix of bits, applies g to each row of this matrix to get a column vector of length n, on which she applies f to get a single bit.

Boolean Formulas

Definition 2.1. *A* de Morgan formula *is a binary tree with* OR *and* AND *gates with fan in 2 on the internal nodes, and variables or their negations on the leaves.*

Definition 2.2. *The* size *of a de Morgan formula F is the number of leaves in it and is denoted by $L(F)$. For a function $f\colon \{0,1\}^n \to \{0,1\}$, we denote by $L(f)$ the size of the smallest de Morgan formula computing the function f.*

Definition 2.3 (Restriction)**.** *Let $f\colon \{0,1\}^n \to \{0,1\}$ be a Boolean function. A* restriction *ρ is a vector of length n of elements from $\{0,1,*\}$. We denote by $f|_\rho$ 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 2.4 (Average-Case Hardness). *Let $\mathcal{D}$ be a distribution over $\{0,1\}^n$, $s \in \mathbb{N}$ and $\varepsilon \in [0,1/2]$. A function $f\colon \{0,1\}^n \to \{0,1\}$ is said to be $1/2+\varepsilon$* hard for formulas of size s under $\mathcal{D}$ *if for any formula of size at most s, it holds that*

$$\mathop{\mathbf{Pr}}_{x\sim\mathcal{D}}[F(x) = f(x)] \leq \frac{1}{2} + \varepsilon \; .$$

When $\mathcal{D} = \mathcal{U}_n$ we may omit mentioning $\mathcal{D}$ and just write that f is $1/2+\varepsilon$ hard for formulas of size s.

2.1 Previous Results

In this work, we are using the following previous results. We denote the parity function on m variables by $\oplus_m$. That is, $\oplus_m : \{0,1\}^m \to \{0,1\}$ is defined by $\oplus_m(x_1,\dots,x_m) = x_1 \oplus x_2 \oplus \dots \oplus x_m$, or alternatively $\oplus_m(x_1,\dots,x_m) = x_1 + x_2 + \dots + x_m \mod 2$.

Theorem 2.5 ([Khr71]). *$L(\oplus_m) = \Theta(m^2)$.*

Theorem 2.6 ([Tal14, Section 7]). *$L(f \circ \oplus_m) = \Theta(L(f)\cdot L(\oplus_m)) = \Theta(L(f)\cdot m^2)$.*

In fact, to prove Theorem 2.6, the following result is implicitly proved in [Tal14].

Theorem 2.7 ([Tal14, Section 7]). *Let $B_1,\dots B_k$ be a partition of $[n]$ to sets of size at least m each. Then, for a random restriction ρ keeping exactly one variable alive from each B_i, and fixing all other variables to values chosen uniformly at random, it holds that*

$$\mathop{\mathbf{E}}_{\rho}[L(f|_\rho)] = O\left(1 + \tfrac{1}{m^2}\cdot L(f)\right) \; .$$

Lemma 2.8 ([Tal14, Claim 6.2]). *Let F be any de Morgan formula of size s, and $\ell \leq s$ be some parameter. Then, F is equivalent to $F'(G_1,\dots,G_m)$ where F' is a read-once de Morgan formula, each G_i is a de Morgan formula of size at most ℓ, and $m = O(s/\ell)$.*

3 Any De-Morgan Formula Can Be Approximated by a Smaller Formula

Theorem 3.1 (Any de Morgan formula can be approximated by a smaller formula). *Let F be a de Morgan formula of size s. Let $k \leq \sqrt{s}$. Let E be any distribution on $\{0,1\}^n$. Then, there exist de Morgan formulas $F_1,\dots,F_{O(k)}$, each of size at most s/k^2, such that either*

$$\mathop{\mathbf{Pr}}_{x\sim E}\left[F(x) = \bigwedge_{i=1}^{O(k)} F_i(x)\right] \geq \frac{1}{2} + \frac{1}{k^{O(k)}} \; .$$

or

$$\mathop{\mathbf{Pr}}_{x\sim E}\left[F(x) = \neg\bigwedge_{i=1}^{O(k)} F_i(x)\right] \geq \frac{1}{2} + \frac{1}{k^{O(k)}} \; .$$

In particular, in both cases, there exists a de Morgan formula F'' of size $O(s/k)$ that agrees with F with probability at least $1/2 + 1/k^{O(k)}$ under the distribution E.

Proof. We start with some notation. Throughout the proof we treat Boolean functions as functions mapping $\{0,1\}^m \to \{-1,1\}$. We do so by transforming any function $f : \{0,1\}^m \to \{0,1\}$ to the function $f' = 1 - 2f : \{0,1\}^m \to \{-1,1\}$. We note that under this standard manipulation, for any two functions $f, g : \{0,1\}^m \to \{-1,1\}$ we have

$$\mathop{\mathbf{Pr}}_{x \in_R \{0,1\}^m}[f(x) = g(x)] = \frac{1}{2} + \frac{1}{2} \cdot \mathop{\mathbf{E}}_{x \in_R \{0,1\}^m}[f(x) \cdot g(x)].$$

We also note that if $f : \{0,1\}^m \to \{0,1\}$ has a polynomial $p(x)$ of degree d such that for all $x \in \{0,1\}^m$, we have $|p(x) - f(x)| \le 1/6$ (as in Theorem 1.2), then $f' = 1 - 2f$ has a polynomial $p' = 1 - 2p$ of degree d such that for all $x \in \{0,1\}^m$, we have $|p(x) - f(x)| \le 1/3$.

For any set $S \subseteq [n]$, we denote by $\chi_S(x) = (-1)^{\sum_{i \in S} x_i}$ the S-Fourier character. χ_S is a $\{0,1\}^m \to \{-1,1\}$ function that corresponds to the parity function on the bits in S. It is well known that any function $f : \{0,1\}^m \to \mathbb{R}$ may be written uniquely as a linear combination of the characters (this is usually referred to as the Fourier transform of f)

$$f(x) = \sum_{S \subseteq [n]} \hat{f}(S) \cdot \chi_S(x),$$

where $\hat{f}(S) \in \mathbb{R}$ is the S-Fourier coefficient given by $\hat{f}(S) = \mathbf{E}_{x \in_R \{0,1\}^m}[f(x) \cdot \chi_S(x)]$.

We are ready to start the actual proof. Applying Lemma 2.8 on F with parameter $\ell = s/k^2$, we get that there exist formulas $G_1, \ldots, G_m$ and F' such that $F \equiv F'(G_1, \ldots, G_m)$, each G_i is of size at most s/k^2, $m = O(k^2)$, and F' is read-once. Let D be the distribution over $\{0,1\}^m$ induced by the distribution of $(G_1(x), \ldots, G_m(x))$ for $x \sim E$.

Let $f' : \{0,1\}^m \to \{-1,1\}$ be the function defined by the formula F'. By Theorem 1.2 applied to F', we have a polynomial $p(y)$ with degree $d = O(\sqrt{m}) = O(k)$, such that for any $y \in \{0,1\}^m$

$$|p(y) - f'(y)| \le 1/3 .$$

Since $f'(y) \in \{-1,1\}$, for any $y \in \{0,1\}^m$, we have $p(y) \cdot f'(y) \in [2/3, 4/3]$. In particular, under the distribution D (and in fact under any distribution), it holds that $\mathbf{E}_{y \sim D}[p(y) \cdot f'(y)] \ge 2/3$. We write the Fourier expansion of $p(y)$ as a function over $\{0,1\}^m$

$$p(y) = \sum_{S \subseteq [m] : |S| \le d} \hat{p}(S) \cdot \chi_S(y) .$$

Since $\hat{p}(S) = \mathbf{E}_{y \in_R \{0,1\}^m}[\chi_S(y) \cdot p(y)]$, we have that

$$|\hat{p}(S)| \le \mathop{\mathbf{E}}_{y \in_R \{0,1\}^m}[|\chi_S(y) \cdot p(y)|] \le 4/3.$$

Hence,

$$\begin{aligned} 2/3 &\le \mathop{\mathbf{E}}_{y \sim D}[p(y) \cdot f'(y)] \\ &= \mathop{\mathbf{E}}_{y \sim D}\left[\sum_{S \subseteq [m] : |S| \le d} \hat{p}(S) \cdot \chi_S(y) \cdot f'(y)\right] \\ &\le \sum_{S \subseteq [m] : |S| \le d} (4/3) \cdot \left|\mathop{\mathbf{E}}_{y \sim D}[\chi_S(y) \cdot f'(y)]\right| \end{aligned}$$

Hence there must exists a set $S \subseteq [m]$ with size at most d such that $|\mathbf{E}_{y\sim D}[\chi_S(y) \cdot f'(y)]| \geq 1/(2 \cdot \binom{m}{\leq d})$. Now, substitute $\chi_S(y)$ with $\sum_{z\in\{0,1\}^S} \mathbb{1}_{\{y_S=z\}} \cdot \chi_S(z)$ to get

$$\begin{aligned}\frac{1}{2\cdot\binom{m}{\leq d}} &\leq \left|\mathop{\mathbf{E}}_{y\sim D}[\chi_S(y)\cdot f'(y)]\right| \\ &= \left|\mathop{\mathbf{E}}_{y\sim D}\left[\sum_{z\in\{0,1\}^S} \mathbb{1}_{\{y_S=z\}}\cdot\chi_S(z)\cdot f'(y)\right]\right| \\ &\leq 2^{|S|}\cdot \max_{z\in\{0,1\}^S}\left|\mathop{\mathbf{E}}_{y\sim D}[\mathbb{1}_{\{y_S=z\}}\cdot f'(y)]\right| .\end{aligned}$$

Hence, under the distribution D, there exists a set $S \subseteq [m]$ of size at most d, and a value $z \in \{0,1\}^S$, such that

$$\left|\mathop{\mathbf{E}}_{y\sim D}[\mathbb{1}_{\{y_S=z\}}\cdot f'(y)]\right| \geq \frac{1}{\binom{m}{\leq d}\cdot 2^{|S|+1}} \geq \frac{1}{\binom{m}{\leq d}\cdot 2^{d+1}}.$$

Denote by $b_1 \in \{-1,1\}$ the sign of $\mathbf{E}_{y\sim D}[\mathbb{1}_{\{y_S=z\}} \cdot f'(y)]$, and consider the two Boolean functions defined by $(b_1 \cdot \mathbb{1}_{\{y_S=z\}} + b_2 \cdot \mathbb{1}_{\{y_S\neq z\}})$ for $b_2 \in \{-1,1\}$. For a uniformly random $b_2 \in_R \{-1,1\}$, independent of all other choices, it holds that $\mathbf{E}_{b_2\in\{-1,1\},y\sim D}[b_2 \cdot \mathbb{1}_{\{y_S\neq z\}} \cdot f'(y)] = 0$. Thus, we get

$$\mathop{\mathbf{E}}_{b_2\in_R\{-1,1\}}[\mathop{\mathbf{E}}_{y\sim D}[(b_1\cdot\mathbb{1}_{\{y_S=z\}} + b_2\cdot\mathbb{1}_{\{y_S\neq z\}})\cdot f'(y)]] = \left|\mathop{\mathbf{E}}_{y\sim D}[\mathbb{1}_{\{y_S=z\}}\cdot f'(y)]\right| \geq \frac{1}{\binom{m}{\leq d}\cdot 2^{d+1}}.$$

By averaging, there must exists a choice of $b_2 \in \{-1,1\}$ that makes

$$\mathop{\mathbf{E}}_{y\sim D}[(b_1\cdot\mathbb{1}_{\{y_S=z\}} + b_2\cdot\mathbb{1}_{\{y_S\neq z\}})\cdot f'(y)] \geq \frac{1}{\binom{m}{\leq d}\cdot 2^{d+1}} .$$

Consider the function $h(y) \triangleq (b_1 \cdot \mathbb{1}_{\{y_S=z\}} + b_2 \cdot \mathbb{1}_{\{y_S\neq z\}})$. In the case where $b_2 = b_1$, this is the constant function $h \equiv b_1$. In the case where $b_2 \neq b_1$, the resulting function $h = b_2 \cdot (-1)^{y_S=z}$ is the AND function on variables in S up to possible negations of the inputs (determined by the values of $(z_i)_{i\in S}$) and of the output (determined by b_2).

Now, take H to be the formula that computes the function h. Since h and f' have correlation at least $1/(\binom{m}{\leq d} \cdot 2^{d+1})$ under D, we have

$$\begin{aligned}\frac{1}{2} + \frac{1}{\binom{m}{\leq d}\cdot 2^{d+2}} &\leq \mathop{\mathbf{Pr}}_{y\sim D}[H(y) = F'(y)] \\ &= \mathop{\mathbf{Pr}}_{x\sim E}[H(G_1(x),\ldots,G_m(x)) = F'(G_1(x),\ldots,G_m(x))] \\ &= \mathop{\mathbf{Pr}}_{x\sim E}[H(G_1(x),\ldots,G_m(x)) = F(x)]\end{aligned}$$

If $b_1 = b_2$ then F is approximated by a constant function. If $b_1 = -1$ and $b_2 = 1$, then F is approximated by the AND of at most $d = O(k)$ of the (possibly-negated) G_i's. If $b_1 = 1$ and $b_2 = -1$, then F is approximated by the NAND of at most $d = O(k)$ of the (possibly-negated) G_i's. □

An equivalent form of Theorem 3.1 is given by Yao's minimax theorem, that states that for any class of Boolean functions $\mathcal{C}$ (e.g., all functions that may be written as the AND or NAND of at most $O(k)$ formulas, each of size at most s/k^2) and any function $f : \{0,1\}^n \to \{0,1\}$.

$$\min_{E\text{: distribution over }\{0,1\}^n}\ \max_{c\in\mathcal{C}}\left(\mathop{\mathbf{Pr}}_{x\sim E}[f(x) = c(x)]\right) = \max_{\mathcal{F}\text{: distribution over }\mathcal{C}}\ \min_{x\in\{0,1\}^n}\left(\mathop{\mathbf{Pr}}_{c\sim\mathcal{F}}[f(x) = c(x)]\right) .$$

Theorem 3.2 (Equivalent form of Main Theorem)**.** *Let F be a de Morgan formula of size s. Let $k \leq \sqrt{s}$. Then, there exists a distribution $\mathcal{F}$ over de Morgan formulas which are either the* AND *or* NAND *of at most $O(k)$ sub-formulas of size at most s/k^2, such that for any $x \in \{0,1\}^n$ we have*

$$\mathop{\mathbf{Pr}}_{F''\sim\mathcal{F}}[F(x) = F''(x)] \geq \frac{1}{2} + \frac{1}{k^{O(k)}} \, .$$

Taking the contrapositive of Theorem 3.1 allows to trade average-case hardness in exchange for better formula size lower bounds in the worst-case.

Theorem 3.3 (Average-case lower bounds over any distribution implies worst-case lower bounds)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. Let E be any distribution on $\{0,1\}^n$. If for any de Morgan formula F of size at most s we have $\mathbf{Pr}_{x\sim E}[F(x) = f(x)] \leq \frac{1}{2} + 2^{-r}$, then $L(f) \geq \Omega(\min\{s^2, \frac{sr}{\log r}\})$.*

Proof. Let F be a formula of size s' that computes f. We apply Theorem 3.1 with parameter $k = \min\{\sqrt{s'}, \frac{r}{c\log r}\}$ for a constant $c > 0$ to be determined later. Then, there exists a formula F'' of size $O(s'/k)$ such that

$$\mathop{\mathbf{Pr}}_{x\sim E}[F''(x) = f(x)] = \mathop{\mathbf{Pr}}_{x\sim E}[F''(x) = F(x)] \geq \frac{1}{2} + \frac{1}{k^{O(k)}} \geq \frac{1}{2} + \frac{1}{2^r}$$

which holds for a suitable choice of the constant $c > 0$. Since F'' agrees with f on more than $1/2 + 2^{-r}$ of the inputs, by the assumption of f we have $L(F'') > s$. Hence, $L(F) = s' = \Omega(sk)$. Now, if $k = \sqrt{s'}$, then $L(F) = \Omega(s^2)$, and if $k = \frac{r}{c\log r}$, then $L(F) = \Omega(sr/\log r)$. Combining both cases gives $L(F) = \Omega(\min\{s^2, sr/\log r\})$. □

A particular natural choice for the distribution E is the uniform distribution over $\{0,1\}^n$, which gives Theorem 1.1 as an immediate corollary.

Corollary 3.4 (Average-case lower bounds over the uniform distribution implies worst-case lower bounds)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. If for any de Morgan formula F of size at most s we have $\mathbf{Pr}_{x\in_R\{0,1\}^n}[F(x) = f(x)] \leq \frac{1}{2} + 2^{-r}$, then $L(f) \geq \Omega\left(\min\{s^2, \frac{sr}{\log r}\}\right)$.*

Remarks. Impagliazzo and Kabanets [IK16] result shows that Corollary 3.4 is tight for the parity function on n variables. They show that any formula of size at most s has correlation at most $\exp(-n^2/s^{1+o(1)})$ with parity. Since the formula complexity of parity is $\Theta(n^2)$ this shows that both Corollary 3.4 and the result of [IK16] are tight up to the $o(1)$ term. For example, for any constant $\alpha > 0$ and for size $s = n^{2-\alpha}$, our result shows that there exists a de Morgan formula of size s with correlation at least $\exp(-n^\alpha \cdot O(\log n))$ with parity, while Impagliazzo and Kabanets show that the correlation with any such formula could not be more than $\exp(-n^{\alpha-o(1)})$.

The work of Komargodski, Raz and Tal [KRT13] gives, for any $n, r \leq n^{1/3}$, a Boolean function $\mathrm{KRT}_r : \{0,1\}^n \to \{0,1\}$ that is $1/2 + 2^{-r}$ hard for formulas of size $\frac{n^3}{r^2\cdot\mathrm{polylog}(n)}$ under the uniform distribution over $\{0,1\}^n$. Improving the size lower bound to, say, $\frac{n^3}{r^{0.99}\cdot\mathrm{polylog}(n)}$ for $r = n^\varepsilon$ would imply (using Corollary 3.4) that this function has formula size $n^{3+\Omega(1)}$ breaking the n^3 barrier.

All applications in this paper use the less general version of our main theorem, namely Corollary 3.4. That is, we exploit only the size of F'' and not its structure (as an AND or NAND of $O(k)$ smaller sub-formulas), and we use only the uniform distribution.

4 Applications

4.1 Slightly Improving Andreev's Lower Bound

The currently best formula lower bound for an explicit function is given in [Tal14] for the Andreev's function [And87], $A : \{0,1\}^n \to \{0,1\}$. There, it is shown that $L(A) \geq \Omega(\frac{n^3}{\log^2 n \log\log n})$, almost matching the straight-forward upper bound of $L(A) \leq O(\frac{n^3}{\log^2 n})$. In this section, we show that combining error-correcting codes with Andreev's argument (as done in [KR13]) allows to get an explicit function f with $L(f) \geq \Omega(\frac{n^3}{\log n \cdot (\log\log n)^2})$. In particular, f is provenly harder than Andreev's function by a factor of at least $\Omega(\log(n)/(\log\log n)^2)$.

Theorem 4.1. *Let $n \in \mathbb{N}$ and $8\log n \leq r \leq n/4$. Let $\mathcal{C}$ be an error correcting code of length 2^r, dimension n, and relative distance $1/2 - 1/2^{r/4}$. Then, there exists $x_0 \in \{0,1\}^n$ such that $L(\mathrm{Enc}_{x_0}) \geq \Omega(n \cdot \frac{r}{\log^2 r})$.*

Proof. By Johnson Bound, every ball of radius $1/2 - 1/2^{r/8}$ contains at most $2 \cdot n \cdot 2^r \leq 2 \cdot 2^{2r}$ many codewords. Let $s := c \cdot n/\log r$, where $c > 0$ is a constant to be determined later. For any formula F of size at most s, on variables $y_1, \ldots, y_r$, let $tt(F) \in \{0,1\}^{2^r}$ be the truth-table of F. A counting argument shows that there are at most $(9r)^s$ formulas of size s on r variables (see [Juk12, Theorem 1.23] for a proof). Take a ball of radius $1/2 - 1/2^{r/8}$ around the truth-table of each formula of size at most s. These balls contain at most $(9r)^s \cdot 2 \cdot 2^{2r} < 2^n$ many codewords, where the inequality holds for a small enough constant $c > 0$. Hence, there is a codeword that avoids all such balls. Denote this codeword by Enc_{x_0} where $x_0 \in \{0,1\}^n$. We know that Enc_{x_0} is a function on $y_1, \ldots, y_r$ which is $\frac{1}{2} + 2^{-r/8}$ hard to compute by any de Morgan formula of size at most s. Hence,

$$L(\mathrm{Enc}_{x_0}) \geq \Omega\left(\min\left\{s^2, \frac{sr}{\log r}\right\}\right) \geq \Omega\left(\frac{nr}{\log^2 r}\right). \qquad \square$$

Theorem 4.2. *Let $8\log n \leq r \leq n/4$. Let $\mathcal{C}$ be an error correcting code of length 2^r, dimension n, and relative distance $1/2 - 1/2^{r/4}$. Let $f : \{0,1\}^{n+n} \to \{0,1\}$ be the following function $f(x,y) = \mathrm{Enc}_x(z_1, \ldots, z_r)$ where each $z_i = \oplus_{j=1}^{n/r} y_{i,j}$. Then, $L(f) \geq \Omega(\frac{n^3}{r\log^2 r})$. In particular, if $r = \Theta(\log n)$, we get $L(f) \geq \Omega(\frac{n^3}{\log n \cdot \log^2 \log n})$.*

We remark that the code in the hypothesis of Theorem 4.2 may be constructed explicitly, as explained in [KRT13, Appendix B].

Proof. Let $x_0 \in \{0,1\}^n$ be as promised by Theorem 4.1. The proof follows since $L(f) \geq L(\mathrm{Enc}_{x_0} \circ \oplus_{n/r})$, and the composition of the function Enc_{x_0} and the parity function on n/r variables has size at least $\Omega(L(\mathrm{Enc}_{x_0}) \cdot n^2/r^2)$ (see Theorem 2.6). $\square$

4.2 Formula Lower Bounds for Random Low-Degree Polynomials (or Candidates for Multi-Linear Functions with Super-Cubic Lower Bounds)

In this section, we investigate the formula complexity of t-linear functions, which is closely related to the formula complexity of degree-t polynomials over $\mathbb{F}_2$. We wish to understand the formula complexity of a random such t-linear functions / degree-t polynomials. We will also be satisfied with lower bounding the hardest such polynomial, as this implies that most such polynomials are hard. We focus on the regime where t is some fixed constant and n grows to infinity.

First, let us set some notation. A t-linear function is a function of the form $f : (\{0,1\}^n)^t \to \{0,1\}$ that is defined over t blocks of variables, $x^{(1)}, x^{(2)}, \ldots, x^{(t)}$, each consists of n Boolean variables, and may be written as

$$f(x^{(1)}, \ldots, x^{(t)}) = \sum_{i_1, i_2, \ldots, i_t \in [n]} a_{i_1,\ldots,i_t} \cdot x^{(1)}_{i_1} \cdot x^{(2)}_{i_2} \cdots x^{(t)}_{i_t} ,$$

where $a_{i_1,\ldots,i_t} \in \mathbb{F}_2$. In other words, f is linear over $\mathbb{F}_2$ as a function of each individual block $x^{(j)}$. We sometime permute the variables indices, which of course does not change the formula complexity of the function. Hence, we will be interested in functions on nt variables, such that there exists a partition of the variables to t sets of size n each and such that with respect to these blocks, the function is t-linear. We compare t-linear functions with degree-t polynomials over nt variables. It is easy to see that the former class is a special case of the latter class. A simple argument (see Lemmas 4.6 and 4.7) shows that the formula complexity of the hardest t-linear function and the hardest degree-t function on nt variables is equal up to $\mathrm{polylog}(n)$ factors, for any constant t. Thus, understanding both cases is essentially the same.

A simple counting argument shows that almost all t-linear functions (alternatively, degree-t functions) have formula complexity $\widetilde{\Omega}(n^t)$. On the other hand by expressing a t-linear function as the parity of at most n^t monomials one easily get that the formula size of each such function is at most $O(n^{2t} \cdot t)$. We believe the latter to be tight.

Conjecture 1 (There exists a n^{2t}-hard t-linear function)**.** *Let $t \in \mathbb{N}$ be some constant. There exists a t-linear function with formula complexity $\widetilde{\Omega}(n^{2t})$.*

The following lemma captures a useful trick introduced by Razborov and Rudich [RR97]. The lemma allows to show that Conjecture 1 implies that a random t-linear function has formula complexity $\widetilde{\Omega}(n^{2t})$ with probability at least $1/2$.

Lemma 4.3 (Most t-linear functions are hard)**.** *Let $t \in \mathbb{N}$ be some constant. Let f be the t-linear with highest formula complexity. Then, a random t-linear function has formula complexity at least $L(f)/4$ with probability at least $1/2$.*

Let f' be the degree-t function with highest formula complexity. Then, a random degree-t function has formula complexity at least $L(f')/4$ with probability at least $1/2$.

Proof. We show the first claim, as the second is proven similarly. Let f be the hardest t-linear function and let $s = L(f)$. We say that a function is easy if its formula complexity is smaller than $s/4$. Let h be a random t-linear function and let $g = f \oplus h$. Then, g is also a random t-linear function. Next, we show that if both h and g are easy, i.e., if both have formula size smaller than $s/4$, then we get a contradiction. Indeed, since $f = g \oplus h = (g \wedge \bar{h}) \vee (\bar{g} \wedge h)$ we have $L(f) \le L(g) + L(\bar{h}) + L(\bar{g}) + L(h) = 2 \cdot (L(g) + L(h)) < s$. Thus, the probability that a random function is easy is at most $1/2$, since otherwise by applying a union bound there are two easy functions g and h such that $g \oplus h = f$. □

Using Theorem 3.3, we are able to attain a non-trivial $\widetilde{\Omega}(n^{t+1})$ lower bound for most t-linear functions. We are unaware of any "classical" proof for this statement that does not use the quantum method (Theorem 1.2). We rely on the following result of Kaufman, Lovett and Porat [KLP12].

Theorem 4.4 ([KLP12, Theorem 3.1], Special case)**.** *Let $d \in \mathbb{N}$ be a constant. Let $\varepsilon \in (0, 1/2]$. Let $g : \{0,1\}^n \to \{0,1\}$. Let $\mathcal{C}$ be the Reed-Muller code with parameters n and d. Then, there are at most $(1/\varepsilon)^{C_d \cdot n^{d-1}}$ codewords of $\mathcal{C}$ of relative hamming distance at most $\frac{1}{2} \cdot (1 - \varepsilon)$ from g.*

Theorem 4.5. *Let $d \in \mathbb{N}$ be a constant. Then, almost all polynomials f with degree at most d over $\mathbb{F}_2$ have $L(f) = \Omega(n^{d+1}/\log^2 n)$.*

Proof. The proof follows the proof outline of Theorem 4.1. We consider the Reed-Muller code with parameters n and d. This is the set of all truth-tables of degree at most d polynomials. For any formula F of size $s \leq c \cdot n^d/\log n$, on variables $x_1, \ldots, x_n$, let $tt(F) \in \{0,1\}^{2^n}$ be the truth-table of F. There are at most $(9n)^s$ formulas of size s on $x_1, \ldots, x_n$ (see [Juk12, Theorem 1.23] for a proof).

Take balls of relative-radius $\frac{1}{2} \cdot (1 - 2^{-\gamma n})$ around all formulas of size at most s, for a small enough constant $\gamma > 0$ to be determined later. By Theorem 4.4, any ball of such radius contains at most $2^{\gamma \cdot C_d \cdot n^d}$ codewords. For a small enough choice of the constants $\gamma > 0$ and $c > 0$, these balls contain at most $(9n)^s \cdot 2^{\gamma \cdot C_d \cdot n^d} \ll 2^{\binom{n}{d}}$ many codewords. Hence, almost all codewords avoid all such balls. For any such codeword $p(x)$, we know that $p(x)$ is a polynomial which is $\frac{1}{2} + 2^{-\Omega(n)}$ hard to compute over that uniform distribution by any de Morgan formula of size s. Hence,

$$L(p) \geq \Omega\left(\frac{sn}{\log n}\right) \geq \Omega\left(\frac{n^{d+1}}{\log^2 n}\right). \qquad \square$$

In the rest of the section, we show simple lemmas that allows us to deduce similar lower bounds for almost all t-linear functions. First, we show that any degree-t $\mathbb{F}_2$-polynomial over nt variables can be written as the sum of not too many t-linear functions, where the t-linearity of each function is with respect to a different partition of the nt variables into t blocks of size n each.

Lemma 4.6 (Decomposition of degree-t polynomials as the sum of t-linear functions)**.** *Any degree-t Boolean function on nt variables can be expressed as the sum of $m = O(\log n \cdot t \cdot e^t)$ functions which are t-linear.*

Proof. Let $f(x_1, \ldots, x_{nt}) = \sum_{S \subseteq [nt], |S| \leq t} a_S \cdot \prod_{i \in S} x_i$ be a degree-t function over nt variables. We choose a random balanced coloring of the nt variables in t colors, that is we choose $C_1 : [nt] \to [t]$ uniformly at random from the family of all colorings which have exactly n variables colored in each color. We consider only sets $S \subseteq [nt]$ of size at most t, for which at most one variable in S is colored with each color, and call such sets C_1-multicolored. We take $f_1(x_1, \ldots, x_{nt})$ to be the partial sum of $a_S \cdot \prod_{i \in S} x_i$ for all sets S that are C_1-multicolored. Note that f_1 is t-linear with respect to the partition induced by the coloring C_1. The t-linear function f_1 covers some of the monomials of f, but not necessarily all of them. We continue with another random balanced coloring $C_2 : [nt] \to [t]$ and take all sets S that are C_2-multicolored but not C_1-multicolored. We continue this way for $m = \mathrm{polylog}(n) \cdot e^t$ steps, where the exact value of m will be determined later. We show that with high probability every monomial was covered by one of the functions, i.e., was multicolored by one of the C_i's.

It is enough to show that every set S of size t, regardless of whether it appears in the expansion of f or not, is multicolored with high probability by one of the C_i's. This is enough since any monomial S' of f is a subset of one of these sets of size t, say S, and if S is multicolored by C_i, then so is S'. Take a monomial $S \subseteq [nt]$ of size t. The probability that S is multicolored by C_1 is

$$\frac{t! \cdot \binom{tn-t}{n-1, n-1, \ldots, n-1}}{\binom{tn}{n, n, \ldots, n}} \geq \frac{t!}{t^t} \geq e^{-t}.$$

Furthermore, conditioned on not being colored by $C_1, \ldots, C_{i-1}$, the probability that S is multicolored by C_t is at least e^{-t} as well. We get that

$$\mathbf{Pr}[S \text{ is not multicolored by } C_1, \ldots, C_m] \leq (1 - e^{-t})^m.$$

Picking $m = 2 \cdot e^t \cdot t \cdot \ln n$ this probability is at most n^{-2t}. We can now apply a union bound over all sets of size t, and get that with high probability all of them are multicolored by some coloring out of $C_1, \dots, C_m$.

Overall, we showed that with high probability the t-linear functions $f_1, \dots, f_m$ cover all monomials of f, which implies that $f = \bigoplus_{i=1}^m f_i$. □

Lemma 4.7 (degree-t polynomials vs. t-linear functions)**.** *Let t be some constant. Then,*

1. *If there exists a degree-t polynomial over nt variables whose formula size is at least s, then there exists a t-linear function whose formula size is at least $\Omega\left(\frac{s}{(t \cdot e^t \cdot \log n)^2}\right)$.*
2. *If there exists a t-linear function whose formula size is at least s, then there exists a degree-t function whose formula size is at least s.*

Proof. 1. We shall prove the contrapositive: if all t-linear functions have formula size at most $s' = o\left(\frac{s}{(t \cdot e^t \cdot \log n)^2}\right)$, then any degree-$t$ function have formula size smaller than s. We use Lemma 4.6 to write $f = \oplus_{i=1}^m f_i$ where each f_i is t-linear and $m = O(\log n \cdot t \cdot e^t)$. To construct a formula for f we take a formula of size $O(m^2)$ for $\oplus_m(x_1, \dots, x_m)$ and replace each leaf marked by x_i with the formula for f_i. This results in a formula of size at most $O(m^2 \cdot s') \le s$ computing f.

2. This holds trivially as any t-linear function is in particular a degree-t function. □

Equipped with Lemmas 4.3, 4.5 and 4.7, we prove that most t-linear functions have formula size $\widetilde{\Omega}(n^{t+1})$.

Theorem 4.8. *Let $t \in \mathbb{N}$ be a constant. Then, most t-linear functions $f : (\{0,1\}^n)^t \to \{0,1\}$, have $L(f) = \Omega(n^{t+1}/\log^4 n)$.*

Proof. By Theorem 4.5 there exists a degree-t function f' with $L(f') \ge \Omega(n^{t+1}/\log^2 n)$. By Lemma 4.7 there exists a t-linear function $f : (\{0,1\}^n)^t \to \{0,1\}$ with $L(f) \ge \Omega(n^{t+1}/\log^4 n)$. Then, by Lemma 4.3, most t-linear functions have formula size at least $\Omega(n^{t+1}/(4\log^4 n))$. □

4.2.1 Towards Breaking the n^3 Barrier in Formula Lower Bounds

Definition 4.9. *Let $n \in \mathbb{N}$, $t < n$. We define the* Andreev-Reed-Muller function *with parameter t, denoted by $\mathrm{ARM}_t : \{0,1\}^{2n} \to \{0,1\}$. Let $m = n^{1/t}$. The function ARM_t is defined on $n + m^t = 2n$ inputs. The function ARM_t, on inputs $x \in \{0,1\}^n$ and $y \in \{0,1\}^{m^t}$, interprets y as the coefficient of a t-linear function on t blocks of m variables each. In other words, y describes a function $h_y(z) = \sum_{i_1,\dots,i_t \in [m]} \left(y_{i_1,\dots,i_t} \cdot z_{i_1}^{(1)} \cdots z_{i_t}^{(t)}\right)$. The input x is viewed of as an (mt)-by-(n/mt) matrix of bits. We take $\{z_i^{(j)}\}_{j \in [t], i \in [m]}$ to be the mt parities of the rows of this matrix. The output of the function $\mathrm{ARM}_t(x, y)$ is $h_y(z)$.*

We note that ARM_t is an explicit $(t+1)$-linear function.

Lemma 4.10. *Assuming Conjecture 1, $L(\mathrm{ARM}_t) = \widetilde{\Omega}(n^{4-2/t})$.*

Proof. Recall that the function ARM_t is defined over $x \in \{0,1\}^n$ and $y \in \{0,1\}^{m^t}$ and y describes a function $h_y(z) = \sum_{i_1,\dots,i_t \in [m]} \left(y_{i_1,\dots,y_t} \cdot z_{i_1}^{(1)} \cdots z_{i_m}^{(m)}\right)$. Let y_0 be the coefficients of the hardest t-linear function. Conjecture 1 implies that $L(h_{y_0}) = \widetilde{\Omega}(m^{2t}) = \widetilde{\Omega}(n^2)$. We get

$$L(f) \ge L(h_{y_0} \circ \oplus_{n/(tm)}) = \Omega\left(L(h_{y_0}) \cdot \left(\frac{n}{tm}\right)^2\right) \ge \widetilde{\Omega}\left(n^2 \cdot \left(\frac{n}{tm}\right)^2\right) = \widetilde{\Omega}\left(n^{4-2/t}\right) ,$$

by using Theorem 2.6, which completes the proof. □

It is not hard to see that if we replace Conjecture 1 with the guarantee of Theorem 4.8, then we get that the explicit $(t+1)$-linear function ARM_t has formula size $\widetilde{\Omega}(n^{3-1/t})$. One important and long-standing goal is to break the n^3 formula-size barrier, a barrier that has been standing since Håstad's work [Hås98]. By Theorem 4.8, most 3-linear functions require formula size $\widetilde{\Omega}(n^4)$. However, we want to point to an explicit such function. Using Andreev's idea, as done in Definition 4.9, allows us to get an explicit $(t+1)$-linear function with $L(f) \geq \widetilde{\Omega}(n^{3-1/t})$, still shy of the desired n^3 barrier.

Conjecture 1 and Lemma 4.10 allows us to break the n^3 barrier for any $t \geq 3$. What can be done if we just assume Conjecture 1 for the case of bilinear functions (i.e., $t = 2$)? In this case, it is not clear how to get an explicit function with better than n^3 formula size lower bound, as Lemma 4.10 implies only an $\widetilde{\Omega}(n^3)$ lower bound. In fact, this bound cannot be improved, since $L(\mathrm{ARM}_2) \leq O(n^3)$ by the most straightforward formula for ARM_2. [2]

This raises the question: which candidates among bilinear functions, other than totally random bilinear functions, do we have to break the n^3 barrier?

One barrier that must be attained before proving $n^{3+\varepsilon}$ lower bounds for a bilinear function is some sort of rigidity bounds on the matrix associated with the bilinear function, which were not known until recently. Recall that a matrix A is said to have rigidity s for rank r if A differs from any matrix of rank at most r in at least s entries. As shown in the next lemma, $n^{1.5+\varepsilon/2}$ rigidity for rank $n^{0.5+\varepsilon/2}$ is required if one wishes to prove $n^{3+\varepsilon}$ lower bounds for bilinear function.

Lemma 4.11. *Let $A \in \mathbb{F}_2^{n\times n}$ be a matrix and let $f_A(x,y) = \sum_{i,j} A_{i,j}x_iy_j$ be a bilinear function associated with A. If $A = S + L$, where S has at most s ones and L has rank at most r, then $L(f_A) = O(s^2 + n^2r^2)$. In particular, if $L(f_A) = \omega(n^{3+\varepsilon})$ then A has rigidity $n^{1.5+\varepsilon/2}$ for rank $n^{0.5+\varepsilon/2}$.*

Proof. First, we observe that $f_A(x,y) = f_S(x,y) \oplus f_L(x,y)$ where $f_S(x,y) = \sum_{i,j} S_{i,j}x_iy_j$ and similarly $f_L(x,y) = \sum_{i,j} L_{i,j}x_iy_j$. We bound $L(f_S)$ and $L(f_L)$ separately. $L(f_S)$ is at most $O(s^2)$ since we can implement a parity of the s non-zero monomials in the expansion of f_S. $L(f_L)$ is at most $O(r^2n^2)$ since we can write

$$f_L(x,y) = \sum_{i=1}^{r} \ell_i(x)\ell_i'(y) \mod 2 \tag{1}$$

where ℓ_i and ℓ_i' are linear (i.e., parity) functions of x and y respectively. Equation (1) and the formula complexity of parity functions immediately implies that $L(f_L) = O(r^2 \cdot n^2)$.

Combining the two upper bounds and using $f_A = f_S \oplus f_L = (f_S \wedge \bar{f}_L) \vee (\bar{f}_S \wedge f_L)$ gives $L(f) \leq 2L(f_S) + 2L(f_L) = O(s^2 + r^2n^2)$. □

The challenge of proving $n^{1.5+\Omega(1)}$ rigidity for rank $n^{0.5+\Omega(1)}$ was met in [GT16] for two families of semi-random[3] bilinear functions: random Hankel matrices and matrices sampled from a 2^{-n}-biased sample space. Both semi-random constructions were converted in [GT16] to explicit 3-linear and 4-linear functions which are at least as hard in any reasonable model of computation (in particular, in the model of de Morgan formulas).

First, we define random Hankel matrices and 2^{-n}-biased matrices:

[2] similarly, $L(\mathrm{ARM}_t) \leq O(n^{4-2/t})$ for any constant t.

[3] by semi-random, we mean random constructions that use $O(n)$-bits of randomness.

1. An n-by-n Hankel matrix over $\mathbb{F}_2$ is a matrix $(A)_{i,j\in[n]}$ such that there exist values $a_2,\ldots,a_{2n}\in\mathbb{F}_2$ for which $A_{i,j}=a_{i+j}$ for all $(i,j)\in[n]^2$. A random Hankel matrix is a uniform choice of such a matrix. Alternatively, we can sample a random Hankel matrix by drawing $2n-1$ independent random bits $a_2,\ldots,a_{2n}\in\mathbb{F}_2$ and letting $A_{i,j}=a_{i+j}$ for all $(i,j)\in[n]^2$.

2. A distribution D over n-by-n Boolean matrices is called ε-biased if for any subset of the entries $S\subseteq[n]\times[n]$ it holds that

$$\mathop{\mathbf{Pr}}_{A\sim D}\left[\bigoplus_{(i,j)\in S} A_{i,j}=1\right]\in\left[\frac{1-\varepsilon}{2},\frac{1+\varepsilon}{2}\right].$$

   We call a random matrix A an ε-biased matrix if the distribution of A is ε-biased. For $\varepsilon=2^{-n}$ we take an explicit construction by Mossel, Shpilka and Trevisan [MST06] of a 2^{-n}-biased distribution. In their construction, only $O(n)$ random bits are used to sample a matrix, and the entries of the matrix are bilinear functions of those random bits.

Both constructions use $O(n)$-random bits, which should be compared to the n^2 random bits needed to define a totally random matrix over $\mathbb{F}_2$.

Theorem 4.12 ([GT16]). *The following holds:*
(1) a 2^{-n}-biased matrix has $\widetilde{\Omega}(n^{1.8})$ rigidity for rank $\widetilde{\Omega}(n^{0.6})$ with high probability.
(2) a random Hankel matrix has $\widetilde{\Omega}(n^{1.8})$ rigidity for rank $\widetilde{\Omega}(n^{0.6})$ with high probability.

This indicates that both constructions require $n^{3+\Omega(1)}$ formula size, with high probability. We make the following bolder conjecture.

Conjecture 2. *Given a matrix A, let $f_A(x,y)\triangleq\sum_{i,j}A_{i,j}x_iy_j$. Then,*
(1) a 2^{-n}-biased matrix has $L(f_A)=\widetilde{\Omega}(n^4)$ with high probability.
(2) a random Hankel matrix A has $L(f_A)=\widetilde{\Omega}(n^4)$ with high probability.

Note that one has to first prove Conjecture 1 for $t=2$, before proving Conjecture 2. Nevertheless, in Lemma 4.13, we prove unconditional super-quadratic lower bounds for these two cases by reducing them to the totally random matrix case and then using Theorem 4.8.

Lemma 4.13. *Given a matrix A, let $f_A(x,y)\triangleq\sum_{i,j}A_{i,j}x_iy_j$. Then,*
(1) a 2^{-n}-biased matrix has $\mathbf{E}_A[L(f_A)]\geq\widetilde{\Omega}(n^{2.5})$.
(2) a random Hankel matrix A has $\mathbf{E}_A[L(f_A)]\geq\widetilde{\Omega}(n^{2.4})$.

Proof. We reduce the semi-random cases to the totally random case on a smaller matrix. In both (1) and (2), we define a parameter $m\leq n$ and consider $m\times m$ submatrices of A (we set $m=\sqrt{n}$ for the first case and $m=n^{0.4}/\log n$ for the latter). The submatrices are induced by a set of rows $I\subseteq[n]$ and a set of columns $J\subseteq[n]$, where I contains exactly one row out of each block of n/m rows and similarly J contains exactly one row out of each block of n/m columns.

1. Let $m=\sqrt{n}$ and fix a choice of I,J as above. Let A be a random 2^{-n}-biased matrix. Let $A|_{I\times J}$ the submatrix induced by rows I and columns J. By Vazirani's XOR Lemma, the statistical distance of $A|_{I\times J}$ from a totally uniform matrix is $2^{-n}\cdot2^{n/2}$. Thus, by Theorem 4.8 with high probability $f_A|_{I\times J}$ has formula size at least $\widetilde{\Omega}(m^3)$. Thus, for any fixed I,J we have that $\mathbf{E}_A[L(f_A|_{I\times J})]\geq\widetilde{\Omega}(m^3)$. By averaging over all possible I,J, we get

$$\mathop{\mathbf{E}}_{A,I,J}[L(f_A|_{I\times J})]\geq\widetilde{\Omega}(m^3).$$

By Theorem 2.7, we have that

$$\mathop{\mathbf{E}}_{A}[L(f_A)] \geq \mathop{\mathbf{E}}_{A}\left[\Omega((n/m)^2)\cdot \mathop{\mathbf{E}}_{I,J}[L(f_A|_{I\times J})]\right] \geq \widetilde{\Omega}((n/m)^2\cdot m^3) = \widetilde{\Omega}(n^{2.5}).$$

2. Let $m = n^{0.4}/\log(n)$, and as before, let I and J be sets of size m picking exactly one row/column out of any block of $\frac{n}{m}$ rows/columns, respectively.

   We analyze how many colliding quadruples $\{(i,j,i',j') \in I\times J : (i,j) \neq (i',j'), i+j = i'+j'\}$ exist. For each of the m^3 choices for (i,j,i') the probability that the unique $j' = i+j-i'$ is chosen to J is at most m/n, thus the expected number of colliding quadruples is at most m^4/n. By Markov's inequality, at least $1/2$ of the choices of I,J have less than $2m^4/n$ colliding quadruples, and call such a choice for (I,J) "good".

   Fix a good choice for I,J, and consider the $I\times J$ submatrix of a random Hankel matrix. This is a matrix whose entries are marked with random bits $\{a_{i+j}\}_{i\in I, j\in J}$, where all but $O(m^4/n)$ of these bits appear exactly once in the submatrix. We relate the complexity of computing the submatrix $A|_{I\times J}$ and the complexity of computing a totally random $m\times m$ matrix. We couple the distributions of the two random matrices. We define independent random bits $\{b_{i,j}\}_{i\in I, j\in J}$. If (i,j) does not participate in any colliding quadruple (i,j,i',j') where $i+j = i'+j'$ then we take $b_{i,j} = a_{i+j}$. Otherwise we take $b_{i,j}$ to be a "fresh" random bit independent of all other choices. It is easy to convince oneself that $\{b_{i,j}\}$ are indeed m^2 independent random bits, and that we needed to introduce at most $O(m^4/n)$ new random bits that were not defined in by a_{i+j}.

   For a fixed choice of the random bits defining $A' := A|_{I\times J}$ and B, we relate the formula complexity of $f_{A'}$ and $f_{B'}$. Any formula for $f_{A'}$ can be transformed into a formula for f_B by considering the difference of the two matrices. Note that the matrices A' and B differ in at most $O(m^4/n)$ entries, which are the fresh bits that were introduced to define B. By $B = A' + (B - A')$, we have $L(f_B) \leq 2\cdot(L(f_{A'}) + L(f_{(B-A')}))$. Furthermore, we can compute $f_{(B-A')}$ using a naive formula of size quadratic in the sparsity of $(B-A')$, i.e. $O(m^8/n^2)$. Hence, we get

   $$\mathop{\mathbf{E}}_{A'}[2L(f_{A'})] + O(m^8/n^2) \geq \mathop{\mathbf{E}}_{B}[L(f_B)] \geq \Omega(m^3/\log^4 n),$$

   where we used Theorem 4.8 in the last inequality. By the choice $m = n^{0.4}/\log n$ we have $m^3/\log^4 n \gg m^8/n^2$, which gives $\mathbf{E}_{A'}[L(f_{A'})] \geq \Omega(m^3/\log^4 n)$. Since (I,J) is good for at least half of the choices of I,J, we have $\mathbf{E}_{A,I,J}[L(f_A|_{I\times J})] \geq \Omega(m^3/\log^4 n)$. Last, by Theorem 2.7, we have that

   $$\mathop{\mathbf{E}}_{A}[L(f_A)] \geq \mathop{\mathbf{E}}_{A}\left[\Omega((n/m)^2)\cdot \mathop{\mathbf{E}}_{I,J}[L(f_A|_{I\times J})]\right] \geq \widetilde{\Omega}((n/m)^2\cdot m^3) = \widetilde{\Omega}(n^{2.4}). \qquad \square$$

Remark. A small calculation shows that under Conjecture 1, the above proof gives $\widetilde{\Omega}(n^3)$ lower bounds for both semi-random cases. Indeed, in the first case we get $\widetilde{\Omega}((n/m)^2\cdot m^4) = \widetilde{\Omega}(n^3)$ for $m = \sqrt{n}$, and in the second case we can pick $m = \sqrt{n}/\mathrm{polylog}(n)$ since we only need to satisfy $m^4/\mathrm{polylog}(m) \gg m^8/n^2$.

4.3 On The KRW Conjecture, Direct Sums and Products

Conjecture 3 (Strong Direct Sum for Formulas)**.** *Let f be a function such that any formula of size s agrees with f on at most 0.51 fraction of the inputs. Then, $\oplus_k \circ f$ is a function such that any formula of size $o(sk)$ agrees with $\oplus_k \circ f$ on at most $1/2 + 2^{-\Omega(k)}$ fraction of the inputs.*

A natural step towards proving Conjecture 3 would be to first prove a direct product theorem.

Conjecture 4 (Strong Direct Product for Formulas)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be a function such that any formula of size s agrees with f on at most 0.51 fraction of the inputs. Then, $f^{\times k} : (\{0,1\}^n)^k \to \{0,1\}^k$, defined by*

$$f^{\times k}(x^{(1)}, x^{(2)}, \dots, x^{(k)}) = \left(f(x^{(1)}), f(x^{(2)}), \dots, f(x^{(k)})\right)$$

is a function such that any formula of size $o(sk)$ agrees with $f^{\times k}$ on at most $2^{-\Omega(k)}$ fraction of the inputs.

Conjecture 3 and Theorem 3.3 implies that $L(\oplus_k \circ f) \geq sk^2/\mathrm{polylog}(k)$, as conjectured by [KRW95] and [GMWW14].

Acknowledgements

I am indebted to Irit Dinur for asking me a crucial question that initiated this research project. During my Ph.D. thesis defense, while presenting the work with Ilan Komargodski and Ran Raz [KRT13], I raised as an open question whether some better tradeoffs between the size and hardness parameters could be attained? Irit asked if this is just a nice-to-have question, or does it imply something more interesting. This left me baffled at the time, as the only answer I had is that it is a natural question, however I have no clue what to do with it. Thinking more about this question made me realize that better tradeoffs would imply breaking the n^3 barrier using Corollary 3.4.

The work in Section 4.2 was highly motivated from conversations and e-mail exchange with Oded Goldreich. I wish to thank him for suggesting the title of this work, and for allowing me to include Conjecture 1 and the proof for Lemma 4.13 in this work.

I thank Dima Gavinsky, Johan Håstad, Or Meir, Ran Raz and Avi Wigderson for helpful discussions.

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 $\mathrm{n}^{1/2+\mathrm{o}(1)}$ 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.

[DM16] I. Dinur and O. Meir. Toward the KRW composition conjecture: Cubic formula lower bounds via communication complexity. In *CCC*, pages 3:1–3:51, 2016.

[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 *STOC*, pages 213–222, 2014.

[GT16] O. Goldreich and A. Tal. Matrix rigidity of random toeplitz matrices. *computational complexity*, pages 1–46, 2016.

[GW13] O. Goldreich and A. Wigderson. On the size of depth-three boolean circuits for computing multilinear functions. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:43, 2013.

[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 *STOC*, pages 526–535, 2007.

[HMP$^+$93] A. Hajnal, W. Maass, P. Pudlák, M. Szegedy, and G. Turán. Threshold circuits of bounded depth. *J. of Computer and System Sciences*, 46(2):129–154, April 1993.

[IK13] R. Impagliazzo and V. Kabanets. Fourier concentration from shrinkage. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:163, 2013. To appear in CCC 2014.

[IK16] R. Impagliazzo and V. Kabanets. Fourier concentration from shrinkage. *computational complexity*, pages 1–47, 2016.

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

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

[KLP12] T. Kaufman, S. Lovett, and E. Porat. Weight distribution and list-decoding size of reed-muller codes. *IEEE Trans. Information Theory*, 58(5):2689–2696, 2012.

[KR13] I. Komargodski and R. Raz. Average-case lower bounds for formula size. In *STOC*, pages 171–180, 2013.

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

[MST06] E. Mossel, A. Shpilka, and L. Trevisan. On epsilon-biased generators in NC^0. *Random Structures and Algorithms*, 29(1):56–81, 2006.

[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, 2009.

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

[RR97] A. A. Razborov and S. Rudich. Natural proofs. *J. Comput. Syst. Sci.*, 55(1):24–35, 1997.

[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, 2010.

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

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

ECCC ISSN 1433-8092

http://eccc.hpi-web.de