Electronic Colloquium on Computational Complexity, Revision 2 of Report No. 163 (2013)

Fourier Concentration from Shrinkage

Russell Impagliazzo* Valentine Kabanets†

March 23, 2016

Abstract

For a class $\mathcal{F}$ of formulas (general de Morgan or read-once de Morgan), the *shrinkage exponent* $\Gamma_{\mathcal{F}}$ is the parameter measuring the reduction in size of a formula $F \in \mathcal{F}$ after F is hit with a random restriction. A Boolean function $f\colon \{0,1\}^n \to \{1,-1\}$ is *Fourier-concentrated* if, when viewed in the Fourier basis, f has most of its total mass on "low-degree" coefficients. We show a direct connection between the two notions by proving that *shrinkage implies Fourier concentration*: for a shrinkage exponent $\Gamma_{\mathcal{F}}$, a formula $F \in \mathcal{F}$ of size s will have most of its Fourier mass on the coefficients of degree up to about $s^{1/\Gamma_{\mathcal{F}}}$. More precisely, for a Boolean function $f\colon \{0,1\}^n \to \{1,-1\}$ computable by a formula of (large enough) size s and for any parameter $r > 0$,

$$\sum_{A \subseteq [n]\,:\,|A| \geqslant s^{1/\Gamma} \cdot r} \hat{f}(A)^2 \leqslant s \cdot \mathsf{polylog}(s) \cdot exp\left(-\frac{r^{\frac{\Gamma}{\Gamma-1}}}{s^{o(1)}}\right),$$

where Γ is the shrinkage exponent for the corresponding class of formulas: $\Gamma = 2$ for de Morgan formulas, and $\Gamma = 1/\log_2(\sqrt{5}-1) \approx 3.27$ for read-once de Morgan formulas. This Fourier concentration result is optimal, to within the $o(1)$ term in the exponent of s.

As a standard application of these Fourier concentration results, we get that subquadratic-size de Morgan formulas have negligible correlation with parity. We also show the tight $\Theta(s^{1/\Gamma})$ bound on the average sensitivity of read-once formulas of size s, which mirrors the known tight bound $\Theta(\sqrt{s})$ on the average sensitivity of general de Morgan s-size formulas.

Keywords: formula complexity, random restrictions, de Morgan formulas, read-once de Morgan formulas, shrinkage exponent, Fourier analysis of Boolean functions, Fourier concentration, average sensitivity

1 Introduction

Over the past thirty years, there have been a number of striking examples of interplay between complexity and algorithms. We know that computationally hard problems are useful for building secure cryptosystems [BM84, Yao82, HILL99], and derandomization [NW94, BFNW93, IW97, Uma03]. On the other hand, circuit lower bounds are implied by non-trivial algorithms for SAT [KL82, Kan82, Wil13, Wil14] or Polynomial Identity Testing [KI04]. It has also been observed that *techniques* used to prove existing circuit lower bounds are often useful for designing learning algorithms [LMN93], SAT algorithms [Zan98, San10, ST12, IMP12, BIS12, CKK^{+}15, CKS15, Tal15],

*Department of Computer Science, University of California, San Diego, La Jolla, CA 91097-0114, USA; russell@cs.ucsd.edu. Work supported by the Simons Foundation and NSF grant CCF-121351.

†School of Computing Science, Simon Fraser University, Burnaby, BC, Canada, V5A 1S6; kabanets@cs.sfu.ca

ISSN 1433-8092

and pseudorandom generators (PRGs) [Bra10, IMZ12, GMR$^+$12, TX13] for the same class of circuits. In particular, the method of *random restrictions*, useful for proving lower bounds against AC^0 circuits [FSS84, Yao85, Hås86] and de Morgan formulas [Sub61, And87, Hås98, San10, KR13, KRT13, Tal14], turns out to be also useful for designing such algorithms for the same circuit class.

We give another example of the connection between random restrictions and algorithms for small de Morgan formulas, by relating the *shrinkage exponent* to the *Fourier spectrum* for such formulas.

For a class $\mathcal{F}$ of formulas (general de Morgan or read-once de Morgan), the *shrinkage exponent* $\Gamma_{\mathcal{F}}$ is the parameter measuring the reduction in size of a formula $F \in \mathcal{F}$ after F is hit with a random restriction: if every variable of an s-size formula $F \in \mathcal{F}$ is kept alive with probability p, and set uniformly randomly to 0 or 1 otherwise, then the minimum formula size of the restricted function is expected to be at most about $p^{\Gamma_{\mathcal{F}}} \cdot s$. A Boolean function $f\colon \{0,1\}^n \to \{1,-1\}$ is *Fourier-concentrated* if, when viewed in the Fourier basis, f has most of its total mass on "low-degree" coefficients.

We show a direct connection between the two notions by proving that *shrinkage implies Fourier concentration*: for a shrinkage exponent $\Gamma_{\mathcal{F}}$, a formula $F \in \mathcal{F}$ of size s will have most of its Fourier mass on the coefficients of degree up to about $s^{1/\Gamma_{\mathcal{F}}}$. More precisely, we prove the following.

Theorem 1.1 (Main Result)**.** *For $\mathcal{F}$ either the class of general de Morgan formulas or the class of read-once de Morgan formulas, let $f\colon \{0,1\}^n \to \{1,-1\}$ be a Boolean function computable by a formula in $\mathcal{F}$ of size s. Then for any sufficiently large s and for any parameter $t > 0$, we have*

$$\sum_{A \subseteq [n]\,:\,|A| \geqslant t} \hat{f}(A)^2 \leqslant s \cdot \mathsf{polylog}(s) \cdot exp\left(-\left(\frac{t^{\Gamma}}{s^{1+o(1)}}\right)^{\frac{1}{\Gamma-1}}\right),$$

where $\Gamma = \Gamma_{\mathcal{F}}$ is the shrinkage exponent for the corresponding class $\mathcal{F}$ of formulas: $\Gamma_{\mathcal{F}} = 2$ for de Morgan formulas, and $\Gamma_{\mathcal{F}} = 1/\log_2(\sqrt{5}-1) \approx 3.27$ for read-once de Morgan formulas.

This Fourier concentration result is *optimal*, to within the $o(1)$ term in the exponent of s. (We get the version stated in the abstract earlier by using $t = s^{1/\Gamma} \cdot r$, for any $r > 0$.)

Rather than the standard shrinkage in expectation, we actually need *concentrated* shrinkage of de Morgan formulas under random restrictions, which means that a formula shrinks in size *with high probability* when hit by a random restriction. Such concentrated shrinkage is implicitly proved by [IMZ12] (which considered the case of certain *pseudorandom* restrictions), building upon the earlier "shrinkage in expectation" results by [HRY95, Hås98].

We establish these "shrinkage into Fourier concentration" implications for both general and read-once de Morgan formulas. A weak version of such Fourier concentration for de Morgan formulas follows from Khrapchenko's lower-bound technique for formulas [Khr71]. A stronger version of Fourier concentration can be deduced from known results in the "quantum computation" literature; see Section 1.2 below for more details. Our proof is a classical argument to establish an even stronger (almost tight) Fourier concentration result. The main novelty of our proof is that it exploits the discovered connection between shrinkage and Fourier concentration. Thanks to this connection, we also get the (almost tight) Fourier concentration result for read-once de Morgan formulas (which are not distinguished from general de Morgan formulas by the "quantum arguments").

These Fourier concentration results for small de Morgan formulas are similar to the Fourier concentration result for AC^0 circuits shown in the celebrated paper by Linial, Mansour, and Nisan [LMN93] (and our proof is inspired by the proof in [LMN93]). As an immediate consequence, we obtain, similarly to [LMN93], strong correlation lower bounds against parity, learning algorithms under the uniform distribution, and average sensitivity bounds for both general de Morgan formulas and read-once de Morgan formulas.

1.1 Other results

1.1.1 Correlation bounds

The Fourier transform of a Boolean function $f\colon \{0,1\}^n \to \{1,-1\}$ is a way to express f in the orthogonal basis of functions

$$\chi_S(x_1,\ldots,x_n) = (-1)^{\sum_{i\in S} x_i},$$

over all subsets $S \subseteq [n]$. Intuitively, the coefficient of f at the basis function χ_S, denoted $\hat{f}(S)$, measures the correlation between f and the parity function on the inputs x_i, for $i \in S$. Thus, one would expect that the classes of circuits for which the parity function is hard to compute would not have much weight on high-degree Fourier coefficients $\hat{f}(S)$ for large sets S, i.e., that such circuits would exhibit *concentration* of the Fourier spectrum over low-degree coefficients.

The first such connection between complexity of computing parity and Fourier concentration was shown by Linial, Mansour, and Nisan [LMN93], based on the strong average-case lower bounds for AC^0 circuits against the parity function [Hås86]. As mentioned earlier, using the results in quantum query complexity [FGG08, ACR$^+$07, RŠ08, Rei09, Rei11], one can also show a version of Fourier concentration for de Morgan formulas of sub-quadratic size.

We extend the approach of [LMN93] to the case of de Morgan formulas of sub-quadratic size. Such formulas cannot compute the parity function in the worst case [Khr71], or even on average (as follows from the work in the quantum setting [BBC$^+$01, Rei11]). As an immediate corollary of Theorem 1.1, we get that a size-s de Morgan formula on n inputs may compute the parity function with bias at most $exp(-n^2/s^{1+o(1)})$. This is tight up to the $o(1)$ term (see Lemma 5.3).

1.1.2 Average sensitivity

Informally, the average sensitivity of a Boolean function $f\colon \{0,1\}^n \to \{1,-1\}$, denoted $AS(f)$, measures the number of influential coordinates in a typical input $x \in \{0,1\}^n$, where a coordinate $i \in [n]$ is influential if flipping the ith bit in x flips the value $f(x)$; we give a more formal definition below. The Fourier concentration we show immediately yields the upper bound $s^{1/\Gamma+o(1)}$ on the average sensitivity of read-once de Morgan formulas of size s, where $\Gamma \approx 3.27$ is the shrinkage exponent for read-once formulas. However, we show (thanks to a personal communication by Nitin Saurabh) that the stronger upper bound $O(s^{1/\Gamma})$ can be obtained from [Bop89]. We then demonstrate the matching lower bound $\Omega(s^{1/\Gamma})$. As the average sensitivity of general de Morgan formulas is $O(\sqrt{s})$ by Khrapchenko's bound [Khr71] (as noted, e.g., in [BDS00, GKLR12]), we get the following tight connection between the shrinkage exponent and the average sensitivity for the class of (general and read-once) de Morgan formulas.

Theorem 1.2. *Let $f\colon \{0,1\}^n \to \{1,-1\}$ be a Boolean function computable by a de Morgan formula of size s. Then $AS(f) \leqslant O(s^{1/\Gamma})$, where Γ is the shrinkage exponent for the corresponding class of formulas: $\Gamma = 2$ for de Morgan formulas, and $\Gamma = 1/\log_2(\sqrt{5}-1) \approx 3.27$ for read-once de Morgan formulas. The average sensitivity $\Omega(s^{1/\Gamma})$ can be achieved with size s de Morgan formulas for $\Gamma = 2$, and read-once formulas for $\Gamma = 1/\log_2(\sqrt{5}-1) \approx 3.27$.*

1.1.3 Learning

As a consequence of our Fourier concentration result, we can also get, similarly to [LMN93], that the class of de Morgan formulas of size s is *learnable* to within error $\epsilon > 0$ in time about

$$n^{s^{1/\Gamma+o(1)}\cdot(\log 1/\epsilon)^{1-1/\Gamma}},$$

over the uniform distribution, where $\Gamma = 2$ for general de Morgan formulas, and $\Gamma \approx 3.27$ for read-once de Morgan formulas. We don't explicitly prove these results here since much better learning algorithms are already known for both general and read-once de Morgan formulas. For general de Morgan formulas, using the quantum-setting results on the sign degree [Lee09], one gets a PAC-learning algorithm for size s de Morgan formulas that runs in time $n^{O(\sqrt{s})}$. For read-once de Morgan formulas, Schapire [Sch94] gives a polynomial-time learning algorithm in the PAC model for any product distribution (hence also for the uniform distribution).

1.2 Related work

As noted by Ganor, Komargodski, Lee, and Raz [GKLR12], the following Fourier concentration result is implied by Khrapchenko's bound [Khr71]: For f computable by size s de Morgan formula, and for any $0 < \epsilon < 1$,

$$\sum_{|A|>s^{1/2}/\epsilon} \hat{f}(A)^2 \leqslant O(\epsilon).$$

The results in quantum query complexity [FGG08, ACR+07, RŠ08, Rei09, Rei11] imply that every de Morgan formula F of size s can be approximated by a polynomial of degree $D \leqslant O(r \cdot s^{1/2})$ with point-wise error at most 2^{-r}, and hence also in the ℓ_2-norm with the same error 2^{-r}. This implies that the Fourier spectrum of F above the degree D is at most 2^{-r}. Hence, for a Boolean function f computed by a de Morgan formula of size s, and for any $t > 0$,

$$\sum_{|A|\geqslant t} \hat{f}(A)^2 \leqslant exp(-t/s^{1/2}). \tag{1}$$

Our Theorem 1.1 provides the stronger bound $exp(-t^2/s^{1+o(1)})$, which is tight to within the $o(1)$ term in the exponent of s (see Lemma 5.3).

As observed in [KRT13], the Fourier concentration in Eq. (1) implies that any de Morgan formula of size

$$s = o((n/\log(1/\epsilon))^2)$$

has correlation at most ϵ with the n-bit parity. The Fourier concentration bound in our Theorem 1.1 implies the correlation at most ϵ for formula size

$$s = (n^2/\log(1/\epsilon))^{1-o(1)}$$

(tight to within the $o(1)$ term).

Our proof of Theorem 1.1 exhibits a connection between the Fourier concentration parameters for a class of formulas and the shrinkage exponent for the same class of formulas. This connection also allows us to get Fourier concentration for the case of read-once formulas, whereas the aforementioned quantum results (based on point-wise polynomial approximations) do not distinguish between read-once and general de Morgan formulas[1].

[1]The $O(\sqrt{s})$ upper bound on the degree of point-wise polynomial approximations is in fact tight for read-once formulas (e.g., an n-variable OR function), and so quantum arguments (which automatically yield point-wise approximations) cannot possibly yield better ℓ_2 approximation bounds for read-once formulas.

For read-once formulas of size s, the upper bound $O(s^{1/\Gamma})$ on the average sensitivity, where Γ is the corresponding shrinkage exponent for read-once formulas, is implicit in the work of [Bop89]. This observation was made by Nitin Saurabh [personal communication, 2013], and we include his argument, with his permission, in Section 7.2.

1.3 Our techniques

Our starting point is the result from [LMN93] which relates the Fourier spectrum of a given Boolean function f for "large" Fourier coefficients to the *expected* Fourier spectrum of the corresponding "large" Fourier coefficients for a *random restriction* of the function f; here a random restriction is obtained by first deciding, with probability p for each variable, whether to restrict it, and then assigning randomly each selected variable either 0 or 1. If the function after a random restriction is likely to depend on fewer than t variables (for some parameter t), then all Fourier coefficients of degree at least t are zero (since a function that depends on fewer than t variables has zero correlation with the parity function of t variables). This is surely the case when the restricted formula is of size less than t. Thus, if we have a "high-probability" shrinkage result for a given class of formulas under random restrictions (showing that a random restriction is likely to shrink the size of a given formula), we immediately get a corresponding Fourier concentration result, where the error bound of the concentration result is the same as the error bound for the shrinkage result.

However, for the case of general de Morgan formulas, such a "high-probability" shrinkage result is simply not true. The problem is posed by the presence of "heavy" variables, the variables that occur too often in a given formula. The notion of a random restriction needs to be modified so that the heavy variables are always restricted, while each of the remaining light variables is chosen to be restricted with some probability p. We adapt the result of [LMN93] mentioned above to the setting of such modified restrictions.

Still, in order to get strong Fourier concentration, one needs the parameter p of a random restriction to be quite small (e.g., n^{ϵ}/n), while the known shrinkage result of [IMZ12] applies only to relatively large values of p (e.g., $p > n^{-1/8}$). The solution is to apply a number of restrictions recursively, each with a relatively large value of p_i, so that the product of the p_i's is as small as we want. Fortunately, the connection between the Fourier spectrum of the original function and of its appropriate random restriction fits in well with such a recursive argument.

A similar approach also works for the case of read-once de Morgan formulas, which are known to shrink with high probability under "pseudorandom" restrictions [IMZ12]. The analysis of [IMZ12] can be used also for the case of truly random restrictions, yielding an exponentially small error. In fact, the case of read-once formulas is slightly simpler as there are no heavy variables.

To prove the optimality of our Fourier concentration, we exhibit a family of small de Morgan formulas that have non-trivial correlation with the parity function. Roughly, the constructed formula computes the AND of parities of small disjoint subsets of the input variables (see Lemma 5.3). This is a standard construction; see, e.g., [Man95, Hås14] for some of the earlier uses.

For the *lower bound* on the average sensitivity of read-once formulas, we use an explicit family of read-once formulas (NAND trees) constructed by [PZ93] (building on [Val84b]), which are known to be shrinkage-resistant. We show that the same read-once formulas of size s have average sensitivity $\Omega(s^{1/\Gamma})$ (see Theorem 7.5).

Remainder of the paper. We state the basics in Section 2, and show how to adapt the approach of [LMN93] in Section 3. We prove the required concentrated shrinkage results for general and read-once de Morgan formulas in Section 4. We derive the Fourier concentration result for general de

Morgan formulas in Section 5, and for read-once formulas in Section 6. In Section 7 we give the application of the Fourier concentration result to correlation with parity, and show tight average sensitivity bounds for read-once de Morgan formulas. We make concluding remarks in Section 8. The appendix contains some proofs omitted from the main body of the paper.

2 Preliminaries

2.1 Notation

We denote by $[n]$ the set $\{1, 2, \ldots, n\}$. We use $exp(a)$ to denote the exponential function 2^a, where a is some numerical expression. All logarithms are base 2 unless explicitly stated otherwise.

2.2 Formulas

A *de Morgan formula* F on n variables $x_1, \ldots, x_n$ is a binary tree whose leaves are labeled by variables or their negations, and whose internal nodes are labeled by the logical operations AND or OR. The *size* of a formula F, denoted by $L(F)$, is the number of leaves in the tree.

A de Morgan formula is called *read-once* if every variable appears at most once in the tree. Note that the size of a read-once formula on n variables is at most n.

2.3 Fourier transform

We review some basics of Fourier analysis of Boolean functions (see, e.g., [Wol08] for a survey, or [O'D14] for a more comprehensive treatment). We think of an n-variate *Boolean* function as $\{-1, 1\}$-valued, i.e., as $f\colon \{0, 1\}^n \to \{-1, 1\}$. For a subset $A \subseteq [n]$, define $\chi_A\colon \{0, 1\}^n \to \{-1, 1\}$ to be

$$\chi_A(x_1, \ldots, x_n) := (-1)^{\sum_{i \in A} x_i}.$$

Let $f\colon \{0, 1\}^n \to \mathbb{R}$ be any function. The *Fourier coefficient* of f at A is defined as

$$\hat{f}(A) := \mathbf{Exp}_{x \in \{0,1\}^n}[f(x) \cdot \chi_A(x)].$$

Note that $\hat{f}(A)$ is exactly the *advantage*[2] of f at computing χ_A, the parity of the inputs from A.

The *Parseval identity* is

$$\sum_{A \subseteq [n]} \hat{f}(A)^2 = \mathbf{Exp}_{x \in \{0,1\}^n}\left[f(x)^2\right].$$

Note that for a Boolean function $f : \{0, 1\}^n \to \{-1, 1\}$, we get

$$\sum_{A \subseteq [n]} \hat{f}(A)^2 = 1.$$

2.4 Random restrictions

For $0 < p < 1$, we define a p-restriction ρ of the set of n variables $x_1, \ldots, x_n$ as follows: for each $i \in [n]$, with probability p assign x_i the value $*$ (i.e., leave x_i unrestricted), and otherwise assign x_i uniformly at random a value 0 or 1. We denote by R_p the distribution of p-restrictions. For a Boolean function $f(x_1, \ldots, x_n)$ and a random restriction ρ, f_ρ denotes the restricted function obtained from f using ρ; f_ρ is a function of the variables left unrestricted by ρ.

[2]Recall that, for functions g and h defined over the same domain $\mathcal{D}$, the advantage of g at computing h is $\mathbf{Pr}_{x \in \mathcal{D}}[g(x) = h(x)] - \mathbf{Pr}_{x \in \mathcal{D}}[g(x) \neq h(x)]$.

2.5 Chernoff-Hoeffding bound

We will use the following version of the Chernoff-Hoeffding bound [Che52, Hoe63].

Lemma 2.1 (Chernoff-Hoeffding)**.** *Let $X = \sum_{i=1}^{t} X_i$ be the sum of independent random variables such that each X_i is in the range $[0, s]$, and $\mathbf{Exp}[X] < E$, for $s, E \geqslant 1$. Then*

$$\mathbf{Pr}[X > 8 \cdot E] < 2^{-E/s}.$$

3 Fourier concentration via random restrictions

We use the following result of [LMN93]; for completeness, we include its proof in the appendix.

Theorem 3.1 ([LMN93])**.** *For arbitrary n-variate Boolean function f, integer $t > 0$ and a real number $0 < p < 1$ such that $pt \geqslant 8$,*

$$\sum_{|A| \geqslant t} \hat{f}(A)^2 \leqslant 2 \cdot \mathbf{Exp}_{\rho \in R_p} \left[\sum_{B\,:\,|B| \geqslant pt/2} \widehat{f_\rho}(B)^2 \right].$$

Imagine we had a "dream version" of the concentrated shrinkage result for de Morgan formulas: For any $0 < p < 1$, a given de Morgan formula F on n variables of size s will shrink to size $s' \leqslant p^2 s$ with probability $1 - \gamma$, for some "small" γ. Let us pick p so that $p^2 s < n$.

Note that a formula of size s' depends on at most s' variables, and hence, all its Fourier coefficients for the sets of size greater than s' are 0. In the notation of Theorem 3.1, every p-restriction ρ, such that the formula size of F_ρ is less than $pt/2$, contributes 0 to the overall expectation; every other restriction ρ (where the formula doesn't shrink) contributes at most 1 (by the Parseval equality). Equating $p^2 s$ and $pt/2$, we get for every $t \geqslant 2ps$,

$$\sum_{|A| > t} \hat{F}(A)^2 \leqslant 2\gamma. \tag{2}$$

For $s \leqslant n^{2-2\epsilon}$, we can achieve the bound of Eq. (2) by setting $p = n^\epsilon / n$ and $t = 8n/n^\epsilon$.

In reality, we don't have such concentrated shrinkage for very small values of γ because of "heavy" variables (those that appear too frequently in the formula)[3]. In order to achieve γ that is inverse-exponentially small in s, we will make sure that heavy variables are always restricted.

Also, the best known concentrated shrinkage results of [IMZ12, KRT13] do not work for very small p. The way around it is to apply a number of random restrictions one after the other, for appropriately chosen $p_1, p_2, \ldots, p_k$, thereby simulating a single restriction with the parameter $p = \prod_{i=1}^{k} p_i$; such a workaround was already used in [IMZ12] and [KRT13].

The following lemma will handle heavy variables. Intuitively, it says that each variable restricted increases the effective degree of where the Fourier coefficients could be large by at most 1.

Lemma 3.2. *Let f be a Boolean function, and x a variable for f. Let f_0 be f with x set to 0, f_1 with x set to 1. For any $\delta \geqslant 0$, if*

$$\sum_{A\,:\,|A| \geq t} \widehat{f_0}(A)^2 \leq \delta \quad \textit{and} \quad \sum_{A\,:\,|A| \geq t} \widehat{f_1}(A)^2 \leq \delta,$$

[3] For example, consider $g(x_1, \ldots, x_n) = f(x_1, \ldots, x_k)$, where $k = O(\log n)$ and f requires formula size $s \approx n^2$; such a function f exists by a counting argument. For any $1/n < p < 1$, a p-restriction of g will leave all $x_1, \ldots, x_k$ unrestricted, and hence fail to shrink g at all, with probability $\gamma \geqslant p^k > 1/n^{O(\log n)}$.

then
$$\sum_{A\,:\,|A|\geq t+1} \hat{f}(A)^2 \leq \delta.$$

Proof. For $y := 1-2x$, we can write
$$\begin{aligned} f &= \frac{(1+y)f_0}{2} + \frac{(1-y)f_1}{2} \\ &= \frac{f_0+f_1}{2} + y \cdot \frac{f_0-f_1}{2}. \end{aligned}$$
Then, for any set A not containing x,
$$\begin{aligned} \hat{f}(A)^2 + \hat{f}(x \cup A)^2 &= \left(\frac{\widehat{f_0}(A)+\widehat{f_1}(A)}{2}\right)^2 + \left(\frac{\widehat{f_0}(A)-\widehat{f_1}(A)}{2}\right)^2 \\ &= \frac{\widehat{f_0}(A)^2}{2} + \frac{\widehat{f_1}(A)^2}{2}. \end{aligned}$$
Summing this over all sets A with $|A| \geq t$ yields at most δ by the assumptions for the restricted functions. Every set B with $|B| \geq t+1$ (containing x or not) is included in this sum. □

So to upper-bound the Fourier mass of the coefficients for sets A with $|A| \geqslant t$, the idea is to set all "heavy" variables (say, z of them), and upper-bound the Fourier mass for each restricted function over the coefficients for sets B with $|B| \geqslant t-z$. If we can bound the Fourier mass of each restricted function by some δ, then, by Lemma 3.2, we get the same upper bound for the Fourier mass of the original function over the sets of size greater than $(t-z)+z=t$, as required.

4 Concentrated shrinkage of de Morgan formulas

Here we prove the following shrinkage results for general and read-once de Morgan formulas, implicit in [IMZ12].

Theorem 4.1 (Shrinkage of general de Morgan formulas)**.** *There exists a constant $c>0$ such that, for every L and every de Morgan formula F with $L(F) \leq L$ on n variables that does not have any variable appearing more than h times, and for every $0<p<1$,*
$$\mathbf{Pr}_{\rho\in R_p}\left[L(F_\rho) \geqslant c \cdot p^2 \cdot \log^{3/2}(1/p) \cdot L\right] \leqslant L(F) \cdot exp\left(-p^6 \cdot L/h\right).$$

Theorem 4.2 (Shrinkage of read-once de Morgan formulas)**.** *There exist constants $d, d' > 0$ such that the following holds for any read-once de Morgan formula $F(x_1, \ldots, x_n)$ and $0<p<1$:*
$$\mathbf{Pr}_{\rho\in R_p}\left[L(F_\rho) \geqslant d \cdot p^{\Gamma} \cdot n\right] \leqslant exp\left(-d' \cdot p^{2\Gamma} \cdot n\right),$$
where $\Gamma = 1/\log(\sqrt{5}-1) \approx 3.27$.

Both of these results are proved using the well-known "shrinkage in expectation" results for the corresponding classes of formulas [Hås98, HRY95, DZ94]. The proof idea is to decompose a given formula into a few batches of independent subformulas (with some extra conditions) and apply "shrinkage in expectation" to each subformula. Since the subformulas in each batch are independent, we can use the Chernoff-Hoeffding inequality to argue that the shrinkage occurs with high probability in each batch, and hence, by the union bound, also for the entire original formula.

We provide more details below. First, in Section 4.1, we give arguments common for the proofs of both these results. Then we prove Theorem 4.1 in Section 4.2, and Theorem 4.2 in Section 4.3.

4.1 Preliminary arguments

We will be using the following "shrinkage in expectation" results. Håstad [Hås98] showed that the shrinkage exponent for de Morgan formulas is 2 (see also [Tal14] for a tighter proof[4]).

Theorem 4.3 ([Hås98]). *There exists a constant $c > 0$ such that, for every de Morgan formula F on n variables and for every $0 < p < 1$,*

$$\mathbf{Exp}_{\rho \in R_p}\left[L(F_\rho)\right] \leqslant c \cdot \left(p^2 \cdot \mu(p, L(F)) \cdot L(F) + p \cdot \sqrt{L(F)}\right),$$

where $\mu(p, L(F)) = 1 + \log^{3/2} \min\{1/p, L(F)\}$.

Håstad, Razborov, and Yao [HRY95] settled the shrinkage exponent for read-once formulas; their result was tightened by Dubiner and Zwick [DZ94].

Theorem 4.4 ([HRY95, DZ94]). *For every read-once formula $F(x_1, \ldots, x_n)$ and a parameter $0 < p < 1$,*

$$\mathbf{Exp}_{\rho \in R_p}[L(F_\rho)] \leqslant O\left(p^\Gamma \cdot n + p \cdot n^{1/\Gamma}\right),$$

where $\Gamma = 1/\log(\sqrt{5} - 1) \approx 3.27$.

Next, we decompose a given (general or read-once) de Morgan formula as follows.

Lemma 4.5 ([IMZ12]). *There is a constant $d_0 > 0$ such that, for every $s > 0$ and for every de Morgan formula F on the set X of variables with $L(F) \geqslant s$, there exist de Morgan formulas $G_1, \ldots, G_m$, for $m \leqslant d_0 \cdot (L(F)/s)$, satisfying the following conditions:*

1. *$L(G_i) \leqslant s$, for all $1 \leqslant i \leqslant m$,*
2. *for each $1 \leqslant i \leqslant m$, G_i has at most 2 occurrences of "special" variables outside of X (different variables for different G_i's), and*
3. *for any restriction ρ of the variables X, $L(F_\rho) \leqslant \sum_{i=1}^{m} L((G_i)_{\rho'})$, where $\rho'(x) = \rho(x)$ for $x \in X$ and $\rho'(x) = *$ otherwise.*

Moreover, if F is a read-once formula, then so is every formula G_i in the collection.

Proof sketch. Find a subformula of size between $s/2$ and s; a maximal subformula of size at most s has size at least $s/2$. Replace the subformula with a new variable, called a subtree variable. Repeatedly find either a subformula with exactly 2 subtree variables and of size less than s, or a subformula with at most 1 subtree variable and of size between $s/2$ and s; replace the found subformula with a new subtree variable. (To find a required subformula, take a minimal subformula of size between $s/2$ and s. If it has more than 2 subtree variables, take a minimal subformula with at least 2 such variables; since each of its child formulas has at most 1 subtree variable, it must have exactly 2.) Since each time, we either remove at least $s/2$ nodes and create 1 new subtree variable, or reduce the number of subtree variables by one, we get at most $d_0 \cdot (L(F)/s)$ subformulas, for some constant $d_0 > 0$, where each subformula is of size at most s and with at most 2 subtree variables. □

[4] In fact, starting from the tight Fourier concentration result for de Morgan formulas (obtained via quantum arguments, cf. Section 1.2), [Tal14] proves a tight version of Theorem 4.3 with $\mu(p, L(F)) = 1$. For our purposes, the original version of Theorem 4.3 (which is proved using classical arguments only) is sufficient.

The special variables correspond to the inputs which are outputs of some other subformulas. We want to analyze the effect of a random restriction on F by using the upper bound of item (3) of Lemma 4.5. To this end, we need to handle random restrictions that leave some specified variables (the "special" variables in our case) unrestricted.

The idea is to take each subformula G_i and construct a new subformula G_i' by replacing each special variable in G_i with a restriction-resistant formula (on new variables, different for different special variables); here we call a formula "restriction-resistant" if, with probability at least $3/4$ over the random restrictions, the resulting restricted formula remains a non-constant function. Then we upper-bound the expected size $\mathbf{Exp}_{\rho'}[L((G_i)_{\rho'})]$, for ρ' that leaves special variables unrestricted, by twice the expected size $\mathbf{Exp}_{\rho}[L((G_i')_{\rho})]$, for a standard random restriction ρ. The latter expectation can be upper-bounded using the above-mentioned "shrinkage in expectation" results.

For general de Morgan formulas, the parity function on k inputs is likely to stay a non-constant function, with high probability over the p-restrictions, where $pk \gg 1$; the size of such a de Morgan formula is $O(k^2)$. For read-once de Morgan formulas, the existence of restriction-resistant formulas follows from the work by Valiant [Val84a]. We state this result with its proof next.

Lemma 4.6 ([IMZ12]). *For every $0 < p < 1$, there exists a read-once de Morgan formula H of size $O(1/p^{\Gamma})$, for $\Gamma = 1/\log_2(\sqrt{5}-1) \approx 3.27$, such that, with probability at least $3/4$ over the p-restrictions ρ, we have*

$$H_\rho(\vec{0}) = 0 \quad \text{and} \quad H_\rho(\vec{1}) = 1, \tag{3}$$

where $\vec{0}$ and $\vec{1}$ denote the inputs of all 0's and all 1's, respectively.

The proof of Lemma 4.6 uses the following notion. For a Boolean function $f(x_1, \ldots, x_n)$ and a parameter $p \in [0,1]$, Boppana [Bop89] defined the *amplification function*

$$A_f(p) := \mathbf{Pr}_{x_1,\ldots,x_n}[f(x_1, \ldots, x_n) = 1],$$

where each x_i is chosen independently at random to be 1 with probability p and 0 otherwise. Boppana [Bop89] also observed that Valiant [Val84a] implicitly proved the following[5].

Theorem 4.7 ([Val84a]). *Let T_k be a complete binary tree of depth $2k$ whose root is labeled with OR, the next layer of nodes with AND, the next layer with OR, and so on in the alternating fashion for all layers but the leaves. Let F_k be the read-once formula computed by T_k on 2^{2k} variables. Then for $\psi = (\sqrt{5}-1)/2$ and any $p \in [0,1]$,*

$$A_{F_k}(\psi - (1-\psi)p) < 1/8 \quad \text{and} \quad A_{F_k}(\psi + (1-\psi)p) > 7/8,$$

for $2k = \log_{2\psi} \frac{\psi - 1/\sqrt{3}}{(1-\psi)p} + O(1) = \log_{2\psi}(1/p) + O(1)$. The size of F_k is $2^{2k} = O(1/p^{1/\log_2 2\psi}) = O(1/p^{\Gamma})$, for $\Gamma = 1/\log_2(\sqrt{5}-1) \approx 3.27$.

Proof of Lemma 4.6. We use Theorem 4.7 to argue the existence of the required read-once formula H. Consider the following distribution D_k on read-once formulas:

> Take T_k. Independently, assign each leaf of T_k the value 1 with probability $2\psi - 1$, and $*$ otherwise. Label the $*$ leaves with distinct variables x_i's. Output the resulting read-once formula in the variables x_i's.

[5] See also the lecture notes by Uri Zwick, `www.cs.tau.ac.il/~zwick/circ-comp-new/six.ps`, for an explicit proof.

Let F be a random read-once formula sampled according to D_k. Let ρ be a random p-restriction on the variables of F. Consider $F_\rho(\vec{1})$. This restricted formula on the all-one input string induces the probability distribution on the leaves of T_k where each leaf, independently, gets value 1 with probability

$$2\psi - 1 + 2(1-\psi)p + 2(1-\psi)(1-p)/2 = \psi + (1-\psi)p.$$

Using Theorem 4.7, we get

$$\begin{aligned}\mathbf{Pr}_{F\in D_k,\rho\in R_p}[F_\rho(\vec{1}) = 1] &= A_{F_k}(\psi + (1-\psi)p)\\ &> 7/8.\end{aligned}$$

Now consider $F_\rho(\vec{0})$. It induces the probability distribution on the leaves of T_k where each leaf, independently, is 1 with probability

$$2\psi - 1 + 2(1-\psi)(1-p)/2 = \psi - (1-\psi)p,$$

and 0 otherwise. Using Theorem 4.7, we get

$$\begin{aligned}\mathbf{Pr}_{F\in D_k,\rho\in R_p}[F_\rho(\vec{0}) = 1] &= A_{F_k}(\psi - (1-\psi)p)\\ &< 1/8.\end{aligned}$$

We get by the union bound that

$$\begin{aligned}\mathbf{Pr}_{F\in D_k,\rho\in R_p}[F_\rho(\vec{1}) = 0 \text{ or } F_\rho(\vec{0}) = 1] &< 1/8 + 1/8\\ &= 1/4.\end{aligned}$$

Finally, by averaging, there exists a particular read-once formula $H \in D_k$ such that, with probability at least $3/4$ over the random p-restrictions ρ, we have $H_\rho(\vec{0}) = 0$ and $H_\rho(\vec{1}) = 1$. The size of this formula H is at most that of F_k, which is $O(1/p^\Gamma)$. $\square$

Now we can analyze the expected shrinkage of de Morgan formulas under p-restrictions that leave some specified variables unrestricted. Let G_i be any formula in the decomposition of Lemma 4.5, with at most two occurrences of special variables. Let H be a shrinkage-resistant formula in the sense that, with probability at most $1/4$ over p-restrictions σ, the restricted formula H_σ is not a constant function. Let G_i' be obtained from G_i by replacing the special variables in G_i by independent copies of the formula H on new, disjoint sets of variables. Let ρ' be a p-restriction on the variables of G_i such that the special variables are assigned *. Let ρ be a p-restriction on all variables of G_i' which agrees with ρ' on all variables of G_i.

We have the following.

Claim 4.8. $\mathbf{Exp}_{\rho'}\left[L((G_i)_{\rho'})\right] \leqslant 2\cdot \mathbf{Exp}_\rho\left[L((G_i')_\rho)\right]$.

Proof of Claim 4.8. Let A be the event that a random p-restriction on the variables of two copies of H leaves both these formulas non-constant. By the union bound, the probability of A is at least $1/2$. Conditioned on A, we have

$$L\left((G_i)_{\rho'}\right) \leqslant L\left((G_i')_\rho\right),$$

since $(G_i')_\rho$ contains $(G_i)_{\rho'}$ as a subfunction. Thus, for a fixed ρ', and for a random ρ extending ρ', we get

$$\mathbf{Exp}_\rho[L((G_i')_\rho)] \geqslant (1/2)\cdot L((G_i)_{\rho'}).$$

Taking the expectation over ρ' on both sides of this inequality yields the desired claim. $\square$

Now we are ready to prove our concentrated shrinkage results.

4.2 Proof of Theorem 4.1

Let $s = c_0 p^{-2}$ for some constant c_0. Using Lemma 4.5, decompose a given formula F into $O(L(F)/s)$ subformulas G_i's.

Let H be a de Morgan formula on $2/p$ fresh variables that computes the parity function. Each such de Morgan formula for parity on $2/p$ variables has size $O(1/p^2)$. The probability that each of $2/p$ variables is assigned (0 or 1) by a random p-restriction is

$$\begin{aligned}(1-p)^{2/p} &\leqslant e^{-2}\\ &\leqslant 1/4.\end{aligned}$$

Thus H is shrinkage-resistant.

Form G'_i by replacing special variables in G_i by independent copies of the formula H. Since each G_i is of size at most $s = c_0/p^2$ and the size of H is $O(1/p^2)$, we get that each G'_i has size c'_0/p^2, for some constant c'_0. By Claim 4.8 and Håstad's Theorem 4.3, we get, for each G_i,

$$\mathbf{Exp}[L((G_i)_{\rho'})] \leqslant 2 \cdot \mathbf{Exp}_\rho[L((G'_i)_\rho)] \leqslant c_1 \cdot \log^{3/2} s, \tag{4}$$

for some constant c_1, where ρ' is a p-restriction on the variables of G_i excluding the special variables, and ρ is a p-restriction extending ρ' to all variables of G'_i.

Thus, we have a collection of $O(L(F)/s)$ formulas G_i, each of size at most s, such that no variable appears in more than h of the G_j's, and such that

$$L(F_\rho) \leqslant \sum L((G_i)_{\rho'}).$$

So our lemma reduces to showing concentration for the latter sum of random variables whose expectations are upper-bounded by Eq. (4).

Since each G_i shares any variables with at most sh other G_j's, we can partition G_i's into $O(sh)$ batches, each of at most $O(L(F)/(s^2h))$ formulas, so that the formulas in each batch are totally independent, having no variables in common. By Eq. (4), the expected total formula size within each batch is

$$O(L(F) \cdot (\log^{3/2} s)/(s^2h)).$$

As a random variable, this is the sum of independent random variables in the range $[0, s]$. By the Chernoff-Hoeffding bound of Lemma 2.1, the probability that the sum of the formula sizes in any batch is larger than

$$c_3 \cdot L(F) \cdot (\log^{3/2} s)/(s^2h)$$

is less than

$$2^{-\Omega(L(F)\cdot(\log^{3/2} s)/s^3h)}.$$

There are strictly less than $L(F) \leqslant L$ batches, so the union bound yields that all batches are of size

$$O(L(F) \cdot (\log^{3/2} s)/(s^2h)),$$

except with probability at most

$$L \cdot \exp(-\Omega(L(F)/(s^3h))) = L \cdot \exp(-\Omega(p^6 \cdot L(F)/h)).$$

If they are, then summing up over the at most $O(sh)$ batches, we get

$$\begin{aligned}L(F_\rho) &\leqslant O(L(F) \cdot (\log^{3/2} s)/s)\\ &= O(p^2 \cdot L(F) \cdot \log^{3/2}(1/p)).\end{aligned}$$

4.3 Proof of Theorem 4.2

Set $s = c/p^\Gamma$, for a constant c to be determined. Using Lemma 4.5, partition a given formula F (of size n) into $O(n/s)$ subformulas $G_1, \ldots, G_m$ of size at most s each.

Let H be a shrinkage-resistant read-once formula from Lemma 4.6 of size $O(1/p^\Gamma)$. Define G_i' to be G_i with special variables in G_i replaced by independent copies of H. Note that

$$L(G_i') \leqslant L(G_i) + O(1/p^\Gamma),$$

which can be made at most $2 \cdot L(G_i)$, by choosing the constant c to be sufficiently large. By Claim 4.8 and Theorem 4.4, we get for each G_i that

$$\mathbf{Exp}_{\rho'}[L((G_i)_{\rho'})] \leqslant c' \cdot p^\Gamma \cdot s, \tag{5}$$

for some constant c', where ρ' is a p-restriction over the variables of G_i excluding the special variables.

By Lemma 4.5, we have

$$L(F_\rho) \leqslant \sum_i L((G_i)_{\rho'}).$$

Note that the latter is the sum of independent random variables, as different G_i's have no variables in common (due to F being read-once). Each of these random variables is in the range $[0, s]$, with expectation upper-bounded by Eq. (5). Hence, the expectation of the sum of these random variables is at most $c''np^\Gamma$, for some constant c''. By the Chernoff-Hoeffding bound of Lemma 2.1, the probability that $L(F_\rho)$ is greater than $8c''np^\Gamma$ is less than

$$exp\left(-c'' \cdot n \cdot p^\Gamma/s\right) \leqslant exp\left(-d' \cdot p^{2\Gamma} \cdot n\right),$$

for some constant $d' > 0$.

5 Fourier concentration of de Morgan formulas

5.1 Concentration

For parameters s and t, denote by $\mathcal{F}(s,t)$ the sum $\sum_{|A| \geqslant t} \hat{f}(A)^2$, where the formula size of f is at most s. The main result of this section is the following.

Theorem 5.1.

$$\mathcal{F}(s,t) \leq s \cdot \mathsf{polylog}(s) \cdot exp\left(-\frac{t^2}{s^{1+\delta(s)}}\right),$$

where $\delta(s) = O((\log\log s)^2/\log s) = o(1)$.

Proof. Starting with an initial formula f of size s and the parameter t, we will apply a sequence of restrictions from R_{p_i} to f, for a sequence of probabilities p_i (to be determined). After stage i, we get a restricted formula f_{i+1} from the previous formula f_i, and the new parameter t_{i+1} from t_i. We then use Theorem 3.1 to reduce the task of upper-bounding $\mathcal{F}(s_i, t_i)$ to that of $\mathcal{F}(s_{i+1}, t_{i+1})$. For our choice of p_i's, the sequence of s_i's will decrease rapidly until at some stage $\ell = O(\log\log s)$ we get $s_\ell < t_\ell$, at which point the recursion stops as we get $\mathcal{F}(s_\ell, t_\ell) = 0$. The bound on $\mathcal{F}(s,t)$ will be essentially the sum of the probabilities, for $0 \leqslant i \leqslant \ell$, that a random restriction $\rho \in R_{p_i}$ fails to shrink the function f_i to the size guaranteed by Theorem 4.1. We provide the details next.

For a parameter $h \in \mathbb{N}$, a variable of f is called *h-heavy* if this variable has more than h occurrences in a minimal formula for f. Let n_h denote the total number of h-heavy variables of f.

Let f_i be a function with formula size at most s_i, and let t_i be the parameter t at stage i. Set $h_i = (2s_i)/t_i$. Let n_{h_i} denote the number of h_i-heavy variables in the formula for f_i. We get that $n_{h_i} \leqslant s_i/h_i = t_i/2$. Let f' be any restriction of f_i assigning values to all h_i-heavy variables. Let $t_i' = t_i/2$. Since $t_i' + n_{h_i} \leqslant t_i$, we get by Lemma 3.2 that it suffices to show, for each f', an upper bound on $\sum_{|A| \geq t_i'} \widehat{f'}(A)^2$. By Theorem 3.1, the latter is at most

$$2 \cdot \mathbf{Exp}_{\rho \in R_{p_i}} \left[\sum_{B \,:\, |B| \geq t_{i+1}} \widehat{f'_\rho}(B)^2 \right],$$

where $t_{i+1} = p_i t_i'/2 = p_i t_i/4$.

By Theorem 4.1, except with probability

$$s_i \cdot exp\left(-p_i^6 \cdot \frac{s_i}{h_i}\right) = s_i \cdot exp\left(-p_i^6 \cdot \frac{t_i}{2}\right) \tag{6}$$

over the random restrictions $\rho \in R_{p_i}$, the function f'_ρ has formula size at most

$$s_{i+1} = p_i^2 \cdot s_i \cdot \Delta,$$

where $\Delta = c \log^{3/2} s$, for the constant c as in Theorem 4.1. We will choose p_i's so that the ratio s_i/t_i becomes less than 1 within few iterations. To that end, we chose p_i so that

$$\frac{s_{i+1}}{t_{i+1}} \leqslant \left(\frac{s_i}{t_i}\right)^{\frac{5}{6}} \cdot \frac{1}{2}. \tag{7}$$

By the definitions of s_{i+1} and t_{i+1}, we have

$$\frac{s_{i+1}}{t_{i+1}} \leqslant \frac{s_i}{t_i} \cdot p_i \cdot 4\Delta,$$

and so we can satisfy Eq. (7) by setting

$$p_i = \left(\frac{t_i}{s_i}\right)^{\frac{1}{6}} \cdot \frac{1}{8\Delta}.$$

For this choice of p_i, the error probability in Eq. (6) becomes at most $s_i \cdot \epsilon_i$ for

$$\epsilon_i = exp\left(-\frac{t_i^2}{s_i} \cdot \frac{1}{2(8\Delta)^6}\right). \tag{8}$$

Using the Parseval identity (to bound by 1 the contribution of those restrictions that do not shrink the formula), we get from the above that

$$\mathbf{Exp}_{\rho \in R_{p_i}} \left[\sum_{B \,:\, |B| \geq t_{i+1}} \widehat{f'_\rho}(B)^2 \right] \leqslant s_i \cdot \epsilon_i + \mathcal{F}(s_{i+1}, t_{i+1}).$$

Hence, overall, we have

$$\mathcal{F}(s_i, t_i) \leqslant 2 \cdot (s_i \cdot \epsilon_i + \mathcal{F}(s_{i+1}, t_{i+1})). \tag{9}$$

Let ℓ be the smallest integer such that $s_\ell < t_\ell$. We will argue below that $\ell = O(\log\log s)$.

Claim 5.2. *For some $\ell = O(\log\log s)$, we get $s_\ell < t_\ell$.*

Proof. By Eq. (7), we have

$$\frac{s_{i+1}}{t_{i+1}} < \left(\frac{s_i}{t_i}\right)^{\frac{5}{6}} \cdot \frac{1}{2}.$$

Unwinding the recurrence for $i+1$ iterations, we get

$$\frac{s_{i+1}}{t_{i+1}} < \left(\frac{s}{t}\right)^{\left(\frac{5}{6}\right)^{i+1}} \cdot \frac{1}{2},$$

which is less than 1 if $i + 1 > \log_{6/5} \log_2(s/t)$. □

For the ℓ as in Claim 5.2, we get $\mathcal{F}(s_\ell, t_\ell) = 0$ (since a formula g depending on fewer than t_ℓ variables has $\hat{g}(B) = 0$ for every set B of size at least t_ℓ). Thus the recurrence in Eq. (9), when started at $i = 0$, will terminate after at most ℓ steps. It follows that $\mathcal{F}(s_0, t_0)$ is at most

$$2s_0\epsilon_0 + 2^2 s_1\epsilon_1 + \cdots + 2^{\ell+1} s_\ell \epsilon_\ell \leqslant 2^{\ell+2} \cdot s \cdot \epsilon^\star, \tag{10}$$

where $\epsilon^\star = \max_{0 \leqslant i \leqslant \ell}\{\epsilon_i\}$. Let $0 \leqslant m \leqslant \ell$ be such that $\epsilon^\star = \epsilon_m$. By unwinding the recurrence in Eq. (8) for ϵ_m, we get

$$\begin{aligned}
\epsilon_m &= exp\left(-\frac{t_m^2}{s_m} \cdot \frac{1}{2(8\Delta)^6}\right) \\
&\leqslant exp\left(-\frac{t_{m-1}^2}{s_{m-1}} \cdot \frac{1}{2(8\Delta)^6} \cdot \frac{1}{16\Delta}\right) \\
&\leqslant exp\left(-\frac{t^2}{s} \cdot \frac{1}{2(8\Delta)^6 \cdot (16\Delta)^m}\right) \\
&\leqslant exp\left(-\frac{t^2}{s} \cdot \frac{1}{2(8\Delta)^6 \cdot (16\Delta)^\ell}\right).
\end{aligned}$$

Plugging in this upper bound on $\epsilon^\star = \epsilon_m$ into Eq. (10), we conclude that

$$\mathcal{F}(s,t) \leqslant s \cdot \mathsf{polylog}(s) \cdot exp\left(-\frac{t^2}{s \cdot (\log s)^{O(\log\log s)}}\right),$$

which completes the proof. □

5.2 Optimality

Let $f : \{0,1\}^n \to \{1,-1\}$ be a Boolean function computed by a de Morgan formula of size s. Since the parity of m bits can be computed by a size $O(m^2)$ de Morgan formula, we have that $\hat{f}(A) = 1$ for a set $A \subseteq [n]$ of size $|A| = O(\sqrt{s})$. Thus, in order to get a non-trivial upper-bound on the Fourier spectrum $\sum_{|A| \geqslant t} \hat{f}(A)^2$, we need to set $t > \sqrt{s}$. We will show something a bit stronger.

Lemma 5.3. *For any $t \leqslant n$, there is a de Morgan formula of size s on n inputs that computes the parity on t bits with advantage $2^{-O(t^2/s)}$.*

Proof. Consider the following formula $F(x_1, \ldots, x_n)$. Set $m = \lfloor ct^2/s \rfloor$, for some constant $c > 0$ to be determined. Without loss of generality assume that m is odd; otherwise take $m - 1$. Divide $x_1, \ldots, x_t$ into m disjoint blocks of size t/m each. Compute the parity of each block, using a de Morgan formula of size $O(t^2/m^2)$, and output the AND of the results over all blocks. The overall formula size of F is $O((t^2/m^2) \cdot m) = O(t^2/m) = O(s/c)$, which can be made at most s, for a sufficiently large constant c.

Next we argue that F has advantage 2^{-m} in computing the parity of $x_1, \ldots, x_t$. Note that F is correct when all m blocks have odd parity, which happens with probability 2^{-m}. If not all blocks have odd parity, our formula always outputs 0, which is correct for exactly $1/2$ of the inputs. $\square$

By Lemma 5.3, a function f computed by a de Morgan formula of size s may have $\hat{f}(A) \geqslant 2^{-O(t^2/s)}$ for a set A of size t. Hence, we get that

$$\mathcal{F}(s,t) \geqslant exp(-O(t^2/s)),$$

implying that our Fourier concentration result for de Morgan formulas, Theorem 5.1, is tight, up to the $o(1)$ term in the exponent of s.

6 Fourier concentration of read-once de Morgan formulas

6.1 Concentration

Here we let $\mathcal{F}(n,t)$ denote the sum $\sum_{|A| \geqslant t} \hat{f}(A)^2$, where f has the *read-once* formula size at most n. The main result of this section is the following.

Theorem 6.1.

$$\mathcal{F}(n,t) \leqslant O(\log n) \cdot exp\left(-\left(\frac{t^{\Gamma}}{n^{1+\delta(n)}}\right)^{\frac{1}{\Gamma-1}}\right),$$

where $\delta(n) = O((\log\log n)/\log n) = o(1)$ *and* $\Gamma = 1/\log(\sqrt{5}-1) \approx 3.27$.

Proof. Our proof strategy is similar to that in Theorem 5.1. We define a sequence of p_i's, and apply restrictions from R_{p_i} to an initial read-once formula f for ℓ steps, each time getting a new read-once formula f_{i+1} of size at most n_{i+1} and a new parameter t_{i+1}. We argue that within $\ell = O(\log\log n)$, we get $n_\ell < t_\ell$, and hence our recursion will stop. The original sum $\mathcal{F}(n,t)$ will be upper-bounded by the sum of error probabilities from Theorem 4.2 that a function from iteration i failed to shrink. We provide the details next.

Let f_i be a function computable by a read-once formula of size at most n_i, and let t_i be the parameter t at stage i. Set $t_{i+1} = p_i t_i/2$. By Theorem 3.1, we have

$$\mathcal{F}(n_i,t_i) \leqslant 2 \cdot \mathbf{Exp}_{\rho\in R_{p_1}}\left[\sum_{B\,:\,|B|\geqslant t_{i+1}} \widehat{(f_i)_\rho}(B)^2\right]. \tag{11}$$

By Theorem 4.2, except with probability at most

$$\epsilon_i = exp(-d' \cdot p_i^{2\Gamma} \cdot n_i) \tag{12}$$

over $\rho \in R_{p_i}$, the function $f_{i+1} = (f_i)_\rho$ has read-once formula size at most

$$n_{i+1} = p_i^{\Gamma} \cdot n_i \cdot d,$$

for some constants $d, d' > 0$. With foresight, set

$$p_i = \left(\left(\frac{t_i}{n_i} \right)^{\frac{1}{2}} \cdot \frac{1}{4d} \right)^{\frac{1}{\Gamma-1}}.$$

We have

$$\begin{aligned} \frac{n_{i+1}}{t_{i+1}} &\leqslant \frac{n_i}{t_i} \cdot (2d) \cdot p_i^{\Gamma-1} \\ &= \left(\frac{n_i}{t_i} \right)^{\frac{1}{2}} \cdot \frac{1}{2}. \end{aligned}$$

It is easy to see (cf. the proof of Claim 5.2) that, for some $\ell \leqslant \log\log n + 1$, we get $n_\ell < t_\ell$, at which point we have $\mathcal{F}(n_\ell, t_\ell) = 0$.

By Eq. (11), we have

$$\mathcal{F}(n_i, t_i) \leqslant 2(\epsilon_i + \mathcal{F}(n_{i+1}, t_{i+1})).$$

Starting at $i = 0$ and unwinding this recurrence for ℓ steps, we get

$$\begin{aligned} \mathcal{F}(n, t) &\leqslant 2 \cdot \sum_{i=0}^{\ell} 2^i \cdot \epsilon_i \\ &\leqslant 2^{\ell+2} \cdot \epsilon_m \end{aligned}$$

where $0 \leqslant m \leqslant \ell$ is such that $\epsilon_m = \max_{0 \leqslant i \leqslant \ell}\{\epsilon_i\}$. As $\ell \leqslant \log\log n + 1$, we get

$$\mathcal{F}(n, t) \leqslant O(\log n) \cdot \epsilon_m. \tag{13}$$

Using our choice of p_i in Eq. (12), we have

$$\begin{aligned} \epsilon_i &= exp\left(-d' \cdot \left(\left(\frac{t_i}{n_i} \right)^{\frac{1}{2}} \cdot \frac{1}{4d} \right)^{\frac{2\Gamma}{\Gamma-1}} \cdot n_i \right) \\ &= exp\left(-n_i \cdot \left(\frac{t_i}{n_i} \right)^{\frac{\Gamma}{\Gamma-1}} \cdot \frac{d'}{(4d)^{\frac{2\Gamma}{\Gamma-1}}} \right) \\ &= exp\left(-\left(\frac{t_i^\Gamma}{n_i \cdot (4d)^{2\Gamma}} \right)^{\frac{1}{\Gamma-1}} \cdot d' \right). \end{aligned}$$

Unwinding this recurrence for m steps, we get

$$\begin{aligned} \epsilon_m &= exp\left(-\left(\frac{t_m^\Gamma}{n_m \cdot (4d)^{2\Gamma}} \right)^{\frac{1}{\Gamma-1}} \cdot d' \right) \\ &\leqslant exp\left(-\left(\frac{t_{m-1}^\Gamma}{n_{m-1} \cdot (4d)^{2\Gamma} \cdot d2^\Gamma} \right)^{\frac{1}{\Gamma-1}} \cdot d' \right) \\ &\leqslant exp\left(-\left(\frac{t^\Gamma}{n \cdot (4d)^{2\Gamma} \cdot (d2^\Gamma)^m} \right)^{\frac{1}{\Gamma-1}} \cdot d' \right), \end{aligned}$$

which is at most

$$exp\left(-\left(\frac{t^{\Gamma}}{n\cdot(\log n)^{O(1)}}\right)^{\frac{1}{\Gamma-1}}\right),$$

since $m \leqslant \ell \leqslant \log\log n + 1$. Using this upper bound on ϵ_m in Eq. (13) completes the proof. □

6.2 Optimality

For every n and $t \geqslant n^{1/\Gamma}$, we give an example of a function $f\colon \{0,1\}^n \to \{-1,1\}$ that matches the upper bound of Theorem 6.1, to within the $o(1)$ term in the exponent of n.

Lemma 6.2. *For every n and $t \geqslant n^{1/\Gamma}$, there exist a Boolean function $f\colon \{0,1\}^n \to \{-1,1\}$ computable by a read-once de Morgan formula, and a constant $d > 0$ such that*

$$\sum_{|A|\geqslant t} \hat{f}(A)^2 \geqslant exp\left(-d\cdot\left(\frac{t^{\Gamma}}{n}\right)^{\frac{1}{\Gamma-1}}\right).$$

Proof. For a parameter $\ell \geqslant 1$ to be determined, partition the variables $x_1,\dots,x_n$ into ℓ disjoint sets $X_1,\dots,X_\ell$ of size n/ℓ each, and define f to be the Boolean function computed by the formula

$$F(x_1,\dots,x_n) = \wedge_{i=1}^{\ell} H(X_i),$$

where H is the shrinkage-resistant formula of size n/ℓ from Lemma 4.6. To show the required lower bound on the Fourier mass of f above level t, we proceed in two steps: (1) show a lower bound on the expected Fourier mass for the restriction f_ρ of f to a family of subsets of total size above $\Omega(tp)$, for an appropriately chosen parameter $0 < p < 1$, and (2) use the known connections between the Fourier spectra of a function and its random restriction to argue that essentially the same lower bound as in step (1) applies also to the Fourier mass of f above level t.

For step (1), we prove the following.

Claim 6.3. *For $p = \Theta((\ell/n)^{1/\Gamma})$ and some constant $C > 0$,*

$$\mathbf{Exp}_{\rho\in R_p}\left[\sum_{\varnothing\neq A_1\subseteq X_1,\dots,\varnothing\neq A_\ell\subseteq X_\ell} \widehat{f_\rho}(A_1\cup\cdots\cup A_\ell)^2\right] \geqslant 2^{-C\cdot\ell}.$$

Proof of Claim 6.3. For the proof, we shall need the following simple facts.

Fact 6.4. *For each non-constant Boolean function g on at most c variables, there exists a subset $\varnothing \neq S \subseteq [c]$ such that $|\hat{g}(S)| \geqslant 2^{-c}$.*

Proof of Fact 6.4. Since g is non-constant, $\hat{g}(S) \neq 0$ for some $\varnothing \neq S \subseteq [c]$. As each Fourier coefficient of a c-variate Boolean function is of the form $k/2^c$ for an integer k, the claim follows. □

Fact 6.5. *For $G(x_1,\dots,x_{2c}) = G_1(x_1,\dots,x_c) \wedge G_2(x_{c+1},\dots,x_{2c})$, let $g_1, g_2\colon \{0,1\}^c \to \{-1,1\}$ and $g\colon \{0,1\}^{2c} \to \{-1,1\}$ be the Boolean functions computed by the formulas G_1, G_2, and G, respectively. Then for any non-empty subsets $S_1 \subseteq \{1,\dots,c\}$ and $S_2 \subseteq \{c+1,\dots,2c\}$, we have*

$$\hat{g}(S_1\cup S_2) = -\frac{1}{2}\cdot\widehat{g_1}(S_1)\cdot\widehat{g_2}(S_2).$$

Proof of Fact 6.5. Observe that $g = \frac{1}{2}\cdot(1+g_1+g_2-g_1\cdot g_2)$, with the first three terms on the right-hand side having no Fourier mass on $S_1\cup S_2$. □

Now we continue with the proof of the claim. Each copy of the formula H is of size $n' = n/\ell$. By Lemma 4.6, we have for $p = \Theta((n')^{-1/\Gamma})$ that, with probability at least 3/4 over random restrictions $\rho \in R_p$, the function computed by H_ρ is non-constant. On the other hand, by Theorem 4.4 and Markov's inequality, the restriction of H under $\rho \in R_p$ has size at most c, for some constant $c > 0$, with probability at least 3/4. It follows that, with probability at least 1/2 over random restrictions $\rho \in R_p$, both conditions hold for H, i.e., the function computed by H_ρ is a non-constant function on at most c variables, for some constant $c > 0$.

Since the ℓ copies of H depend on disjoint sets of variables $X_1, \ldots, X_\ell$, we conclude that, with probability at least $2^{-\ell}$ over $\rho \in R_p$, each restricted formula $H_\rho(X_i)$, for $1 \leqslant i \leqslant \ell$, computes a non-constant Boolean function on at most c variables. For such a restriction ρ, we get by Fact 6.4 that there exist non-empty sets $S_1, \ldots, S_\ell$, where each $S_i \subseteq X_i$, such that, for each $1 \leqslant i \leqslant \ell$, $|\widehat{g_i}(S_i)| \geqslant 2^{-c}$, where g_i is the Boolean function computed by the restricted formula $H_\rho(X_i)$. Applying Fact 6.5 inductively to the formula $F_\rho = \wedge_{i=1}^{\ell} H_\rho(X_i)$, we get that $\left|\widehat{f_\rho}(S_1 \cup \ldots S_\ell)\right| \geqslant 2^{1-\ell} \cdot 2^{-c\ell} \geqslant 2^{-(c+1)\ell}$. It follows that

$$\mathbf{Exp}_{\rho\in R_p}\left[\sum_{\varnothing\neq A_1\subseteq X_1,\ldots,\varnothing\neq A_\ell\subseteq X_\ell} \widehat{f_\rho}(A_1 \cup \cdots \cup A_\ell)^2\right] \geqslant 2^{-\ell} \cdot 2^{-2(c+1)\ell},$$

which is at least $2^{-C\cdot\ell}$, for $C = 2c + 3$. □

Then, for step (2), we use the well-known fact (see, e.g., [O'D14, Proposition 4.17]) that, for any Boolean function $g(x_1, \ldots, x_n)$ and any subset $S \subseteq [n]$,

$$\mathbf{Exp}_{\rho\in R_p}[\widehat{g_\rho}(S)^2] = \sum_{A\subseteq[n]} \hat{g}(A)^2 \cdot \mathbf{Pr}_{\rho\in R_p}[A_\rho = S],$$

where A_ρ denotes the subset of elements of A that were left unrestricted by the random p-restriction ρ (where each element of A is left unrestricted, independently, with probability p). Applying this to our function f, we get that

$$\mathbf{Exp}_\rho\left[\sum_{\varnothing\neq A_1\subseteq X_1,\ldots,\varnothing\neq A_\ell\subseteq X_\ell} \widehat{f_\rho}(A_1 \cup \cdots \cup A_\ell)^2\right] = \sum_{A\subseteq[n]} \hat{f}(A)^2 \cdot \mathbf{Pr}_\rho\left[\forall i \in [\ell], A_\rho \cap X_i \neq \varnothing\right],$$

and hence, by Claim 6.3,

$$2^{-C\cdot\ell} \leqslant \sum_{A\subseteq[n]} \hat{f}(A)^2 \cdot \mathbf{Pr}_\rho\left[\forall i \in [\ell], A_\rho \cap X_i \neq \varnothing\right]. \tag{14}$$

We shall need the following.

Claim 6.6. *For any constant $D > 0$, let $A \subseteq [n]$ be any set such that $|A| \leqslant \frac{\ell}{D\cdot p}$. Then*

$$\mathbf{Pr}_{\rho\in R_p}\left[\forall i \in [\ell], A_\rho \cap X_i \neq \varnothing\right] \leqslant \left(\frac{2}{D}\right)^{\ell/2}.$$

Proof of Claim 6.6. By averaging, for at least $\ell/2$ blocks X_i's, we have $|A \cap X_i| \leqslant \frac{2}{Dp}$. For each such block X_i, we have by the union bound that $\mathbf{Pr}_{\rho\in R_p}[A_\rho \cap X_i \neq \varnothing] \leqslant \frac{2}{D}$. The claim follows. □

Claim 6.6 and Parseval's identity imply that, for any constant $D > 0$, we have

$$\sum_{A\subseteq[n]} \hat{f}(A)^2 \cdot \mathbf{Pr}_{\rho\in R_p}\left[\forall i \in [\ell], A_\rho \cap X_i \neq \varnothing\right] \leqslant \left(\sum_{|A|\geqslant \frac{\ell}{Dp}} \hat{f}(A)^2\right) + \left(\frac{2}{D}\right)^{\ell/2}.$$

By Eq. (14), we conclude that

$$\sum_{|A|\geqslant \frac{\ell}{Dp}} \hat{f}(A)^2 \geqslant 2^{-C\cdot\ell} - (2/D)^{\ell/2}.$$

For $D = 2^{2C+3}$, we get

$$\sum_{|A|\geqslant \frac{\ell}{Dp}} \hat{f}(A)^2 \geqslant \frac{1}{2} \cdot 2^{-C\cdot\ell}. \tag{15}$$

Finally, set ℓ so that $t = \ell/(Dp)$. As $p = \Theta((\ell/n)^{1/\Gamma})$, we get $\ell = \Theta(t(\ell/n)^{1/\Gamma})$, which yields $\ell = \Theta\left(\left(t^\Gamma/n\right)^{\frac{1}{\Gamma-1}}\right)$. By Eq. (15), the lemma follows. □

7 Other results

7.1 Correlation with Parity

Subquadratic-size de Morgan formula have exponentially small correlation with the parity function.

Corollary 7.1. *Every de Morgan formula of size at most $s = n^{2-\epsilon}$, for some $0 < \epsilon \leqslant 1$, agrees with the parity function on n bits on at most*

$$1/2 + exp(-n^{\epsilon - o(1)})$$

fraction of inputs.

Proof. Recall that the Fourier coefficient $\hat{f}(S)$ for a subset $S \subseteq [n]$ measures the correlation of f with the parity function on the positions in S. The result follows immediately from Theorem 5.1. □

By Lemma 5.3, this correlation bound is tight, up to the $o(1)$ term.

7.2 Average sensitivity

Recall that for a Boolean function $f : \{0,1\}^n \to \{1,-1\}$ and a string $w \in \{0,1\}^n$, the *sensitivity* of f at w is the number of Hamming neighbors w' of w such that $f(w) \neq f(w')$. The *average sensitivity* of f, denoted by $AS(f)$, is the average over all $w \in \{0,1\}^n$ of the sensitivity of f at w. It is shown by [KKL88] that

$$AS(f) = \sum_{A\subseteq[n]} |A| \cdot \hat{f}(A)^2. \tag{16}$$

The parity function on m bits has average sensitivity m. Since a de Morgan formula of size s can compute the parity on $\Omega(\sqrt{s})$ bits, we get a lower bound $\Omega(\sqrt{s})$ on the average sensitivity of de Morgan formulas of size s. The matching $O(\sqrt{s})$ upper bound on the average sensitivity of size s de Morgan formulas follows from Khrapchenko's result [Khr71] (as noted in [BDS00, GKLR12]).

For *read-once* formulas of size s, Eq. (16) and Theorem 6.1 readily imply the upper bound $s^{1/\Gamma+o(1)}$ on average sensitivity, where $\Gamma = 1/\log_2(\sqrt{5}-1) \approx 3.27$ is the shrinkage exponent for read-once formulas. However, a stronger upper bound can be shown. As was pointed out to us by Nitin Saurabh (personal communication), the following bound is implicitly proved by Boppana [Bop89].

Theorem 7.2 (implicit in [Bop89]). *Let $f : \{0,1\}^n \to \{1,-1\}$ be a Boolean function computed by a read-once de Morgan formula. Then $AS(f) \leqslant n^{1/\Gamma}$.*

We will prove the theorem for $\{0,1\}$-valued Boolean functions; clearly this does not affect the average sensitivity. We again use Boppana's amplification function, A_f, mentioned earlier. Here we use a slightly more general definition of A_f: for a Boolean function $f : \{0,1\}^n \to \{0,1\}$ and parameters $p_1,\dots,p_n \in [0,1]$, define the *amplification function*

$$A_f(p_1,\dots,p_n) := \mathbf{Pr}_{x_1,\dots,x_n}[f(x_1,\dots,x_n)=1],$$

where each x_i is chosen independently at random to be 1 with probability p_i, and 0 with probability $1-p_i$. For $p \in [0,1]$, define

$$A_f(p) := A_f(p,\dots,p).$$

Boppana [Bop89, Theorem 2.1] proved the following upper bound on the *derivative* of A_f.

Theorem 7.3 ([Bop89]). *For any read-once formula f of size n and any $0 < p < 1$,*

$$A'_f(p) \leqslant n^{1/\Gamma} \cdot \frac{H(A_f(p))}{H(p)},$$

where $H(p) := -p\log_2 p - (1-p)\log_2(1-p)$ is the binary entropy function, and $\Gamma = 1/\log_2(\sqrt{5}-1)$.

Lemma 7.4 (N. Saurabh, personal communication). *For every monotone n-variate Boolean function f, we have $AS(f) = A'_f(1/2)$.*

Proof. Observe that

$$A'_f(1/2) = \sum_{i=1}^{n} \left.\frac{\partial A_f(p_1,\dots,p_n)}{\partial p_i}\right|_{(1/2,\dots,1/2)}.$$

On the other hand, using monotonicity of f, we will show that each ith summand on the right-hand side of the above formula is exactly equal to $\mathbf{Inf}_i[f]$, the influence of coordinate i on f. Since

$$AS(f) = \sum_{i=1}^{n} \mathbf{Inf}_i[f],$$

the lemma will follow.

We have

$$\mathbf{Inf}_i[f] = \sum_{x\in\{0,1\}^n \,:\, (f(x)=1)\wedge(f(x^i)=0)} \frac{1}{2^{n-1}},$$

where x^i denotes x with the ith coordinate flipped. Write

$$A_f(p_1,\dots,p_n) = \sum_{x\in\{0,1\}^n \,:\, f(x)=1} P_x,$$

where for $x = (x_1, \dots, x_n)$,

$$P_x := \prod_{i=1}^{n} p_i^{x_i}(1 - p_i)^{1-x_i}$$

is the probability mass contributed by the point x. Observe that, for points x and x^i, the partial derivatives of P_x and P_{x^i} with respect to p_i cancel each other. Thus, the points x and x^i such that $f(x) = f(x^i) = 1$ contribute 0 to the partial derivative of A_f with respect to p_i. Each x such that $f(x) = 1$ but $f(x^i) = 0$ must have its ith coordinate $x_i = 1$ by the monotonicity of f. Hence, each such x will contribute

$$(1/p_i) \cdot \prod_{j=1}^{n} p_j^{x_j}(1 - p_j)^{1-x_j}$$

to the partial derivative of A_f with respect to p_i. When all $p_j = 1/2$, this contribution is exactly $1/2^{n-1}$. □

We can now finish the proof of Theorem 7.2.

Proof of Theorem 7.2. Without loss of generality, a given read-once Boolean function f can be assumed monotone: we can always remove negations from any negative literals in the read-once formula f, without changing $AS(f)$. By Theorem 7.3 and Lemma 7.4, we get

$$\begin{aligned} AS(f) &\leqslant n^{1/\Gamma} \cdot H(A_f(1/2)) \\ &\leqslant n^{1/\Gamma}, \end{aligned}$$

as required. □

Next we show that the average sensitivity bound for read-once formulas in Theorem 7.2 is tight.

Theorem 7.5. *For all large enough n, there is an n-variate Boolean function f computable by a read-once formula of size n such that*

$$AS(f) \geqslant \Omega(n^{1/\Gamma}).$$

Proof. For every n-variate Boolean function f and for every $0 \leqslant t \leqslant n$, we get by Eq. (16) that

$$\begin{aligned} AS(f) &= \sum_{A \subseteq [n]} |A| \cdot \hat{f}(A)^2 \\ &\geqslant \sum_{|A| \geqslant t} |A| \cdot \hat{f}(A)^2 \\ &\geqslant t \cdot \sum_{|A| \geqslant t} \hat{f}(A)^2. \end{aligned}$$

On the other hand, for the read-once n-variate Boolean function f from Lemma 6.2, we have

$$\sum_{|A| \geqslant t} \hat{f}(A)^2 \geqslant \Omega(1),$$

for $t = n^{1/\Gamma}$. For this f, we conclude by the above that $AS(f) \geqslant \Omega(n^{1/\Gamma})$, as required. □

8 Concluding remarks

We argued that shrinkage implies Fourier concentration for de Morgan formulas. Tal [Tal14] has recently proved that, in some sense, the reverse is also true: starting with the known tight Fourier concentration result for de Morgan formulas (proved via quantum arguments), he shows a tight shrinkage result for de Morgan formulas, improving upon the parameters of [Hås98]. So there appears to be a certain equivalence between shrinkage and Fourier concentration for de Morgan formulas, which raises the issue of proving such connection more generally. For example, one could consider classes of formulas over different bases (say, monotone formulas).

Can one further improve the parameters of Theorem 1.1 (getting rid of the $o(1)$ term there)? Does k-wise independence ϵ-fool read-once formulas of size n for

$$k = O((\log 1/\epsilon) \cdot n^{1/\Gamma})$$

where Γ is the shrinkage exponent for read-once formulas? For general de Morgan formulas of size n, the corresponding statement follows from the quantum results on the approximate degree $O(\sqrt{s})$ [Rei11]. On the other hand, the approximate degree for read-once formulas of size n must be at least $n^{1/2}$ (the same as that for general de Morgan formulas of size n), and so one needs a different argument for showing such a k-wise independence result for read-once formulas.

Acknowledgments We thank Nitin Saurabh for telling us about his observation (Theorem 7.2), and allowing us to include his proof argument in our paper. We also thank Ilan Komargodski and Avishay Tal for their comments on an early version of the paper; special thanks to Avishay for pointing out an error (in the proof of Theorem 4.1) in an early version of this paper, and for sending us his paper [Tal14]. We thank the anonymous referees of CCC'14 for their comments, in particular, for pointing out to us that the known quantum results imply Fourier concentration for general de Morgan formulas (as outlined in Section 1.2). We thank the anonymous referees of the journal version for their thorough comments and suggestions that significantly improved the presentation. We are particularly indebted to one of the referees who suggested how to improve the parameters of our Fourier concentration result (Theorem 1.1) to make them almost tight, and how to simplify the proof of our Theorem 7.5. Thanks to this referee, the present paper has become considerably stronger than it was before!

References

[ACR+07] A. Ambainis, A.M. Childs, B. Reichardt, R. Špalek, and S. Zhang. Any And-Or formula of size n can be evaluated in time $n^{1/2+o(1)}$ on a quantum computer. In *Proceedings of the Forty-Eighth Annual IEEE Symposium on Foundations of Computer Science*, pages 363–372, 2007.

[And87] A.E. Andreev. On a method of obtaining more than quadratic effective lower bounds for the complexity of π-schemes. *Vestnik Moskovskogo Universiteta. Matematika*, 42(1):70–73, 1987. English translation in *Moscow University Mathematics Bulletin.*

[BBC+01] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. Quantum lower bounds by polynomials. *Journal of the Association for Computing Machinery*, 48(4):778–797, 2001.

[BDS00] A. Bernasconi, C. Damm, and I. Shparlinski. The average sensitivity of square-freeness. *Computational Complexity*, 9(1):39–51, 2000.

[BFNW93] L. Babai, L. Fortnow, N. Nisan, and A. Wigderson. BPP has subexponential time simulations unless EXPTIME has publishable proofs. *Computational Complexity*, 3:307–318, 1993.

[BIS12] P. Beame, R. Impagliazzo, and S. Srinivasan. Approximating AC^0 by small height decision trees and a deterministic algorithm for $\#\mathsf{AC}^0$SAT. In *Proceedings of the Twenty-Seventh Annual IEEE Conference on Computational Complexity*, pages 117–125, 2012.

[BM84] M. Blum and S. Micali. How to generate cryptographically strong sequences of pseudorandom bits. *SIAM Journal on Computing*, 13:850–864, 1984.

[Bop89] R. Boppana. Amplification of probabilistic Boolean formulas. In S. Micali, editor, *Randomness and Computation*, volume 5 of *Advances in Computer Research*, pages 27–45. JAI Press, Greenwich, CT, 1989. (preliminary version in FOCS'85).

[Bra10] M. Braverman. Polylogarithmic independence fools AC^0 circuits. *Journal of the Association for Computing Machinery*, 57(5):28:1–28:10, 2010.

[Che52] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, 23:493–509, 1952.

[CKK+15] R. Chen, V. Kabanets, A. Kolokolova, R. Shaltiel, and D. Zuckerman. Mining circuit lower bound proofs for meta-algorithms. *Computational Complexity*, 24(2):333–392, 2015.

[CKS15] R. Chen, V. Kabanets, and N. Saurabh. An improved deterministic #SAT algorithm for small de morgan formulas. *Algorithmica*, pages 1–20, 2015.

[DZ94] M. Dubiner and U. Zwick. How do read-once formulae shrink? *Combinatorics, Probability & Computing*, 3:455–469, 1994.

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

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

[GKLR12] A. Ganor, I. Komargodski, T. Lee, and R. Raz. On the noise stability of small de Morgan formulas. *Electronic Colloquium on Computational Complexity*, TR12-174, 2012.

[GMR+12] P. Gopalan, R. Meka, O. Reingold, L. Trevisan, and S. Vadhan. Better pseudorandom generators via milder pseudorandom restrictions. In *Proceedings of the Fifty-Third Annual IEEE Symposium on Foundations of Computer Science*, pages 120–129, 2012.

[Hås86] J. Håstad. Almost optimal lower bounds for small depth circuits. In *Proceedings of the Eighteenth Annual ACM Symposium on Theory of Computing*, pages 6–20, 1986.

[Hås98] J. Håstad. The shrinkage exponent of de Morgan formulae is 2. *SIAM Journal on Computing*, 27:48–64, 1998.

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

[HILL99] J. Håstad, R. Impagliazzo, L. Levin, and M. Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28:1364–1396, 1999.

[Hoe63] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[HRY95] J. Håstad, A.A. Razborov, and A.C. Yao. On the shrinkage exponent for read-once formulae. *Theoretical Computer Science*, 141(1&2):269–282, 1995.

[IMP12] R. Impagliazzo, W. Matthews, and R. Paturi. A satisfiability algorithm for AC^0. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 961–972, 2012.

[IMZ12] R. Impagliazzo, R. Meka, and D. Zuckerman. Pseudorandomness from shrinkage. In *Proceedings of the Fifty-Third Annual IEEE Symposium on Foundations of Computer Science*, pages 111–119, 2012.

[IW97] R. Impagliazzo and A. Wigderson. P=BPP if E requires exponential circuits: Derandomizing the XOR Lemma. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, pages 220–229, 1997.

[Kan82] R. Kannan. Circuit-size lower bounds and non-reducibility to sparse sets. *Information and Control*, 55:40–56, 1982.

[Khr71] V.M. Khrapchenko. A method of determining lower bounds for the complexity of π-schemes. *Matematicheskie Zametki*, 10(1):83–92, 1971. English translation in *Mathematical Notes of the Academy of Sciences of the USSR*.

[KI04] V. Kabanets and R. Impagliazzo. Derandomizing polynomial identity tests means proving circuit lower bounds. *Computational Complexity*, 13(1–2):1–46, 2004.

[KKL88] J. Kahn, G. Kalai, and N. Linial. The influence of variables on Boolean functions (extended abstract). In *Proceedings of the Twenty-Ninth Annual IEEE Symposium on Foundations of Computer Science*, pages 68–80, 1988.

[KL82] R.M. Karp and R.J. Lipton. Turing machines that take advice. *L'Enseignement Mathématique*, 28(3-4):191–209, 1982.

[KR13] I. Komargodski and R. Raz. Average-case lower bounds for formula size. In *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*, pages 171–180, 2013.

[KRT13] I. Komargodski, R. Raz, and A. Tal. Improved average-case lower bounds for DeMorgan formula size. *Electronic Colloquium on Computational Complexity*, 20(58), 2013.

[Lee09] T. Lee. A note on the sign degree of formulas. *CoRR*, abs/0909.4607, 2009.

[LMN93] N. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, Fourier transform and learnability. *Journal of the Association for Computing Machinery*, 40(3):607–620, 1993.

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

[NW94] N. Nisan and A. Wigderson. Hardness vs. randomness. *Journal of Computer and System Sciences*, 49:149–167, 1994.

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

[PZ93] M. Paterson and U. Zwick. Shrinkage of de Morgan formulae under restriction. *Random Structures and 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 *Proceedings of the Fiftieth Annual IEEE Symposium on Foundations of Computer Science*, pages 544–551, 2009.

[Rei11] B. Reichardt. Reflections for quantum query algorithms. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '11, pages 560–569, 2011.

[RŠ08] B. Reichardt and R. Špalek. Span-program-based quantum algorithms for evaluating formulas. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, pages 103–112, 2008.

[San10] R. Santhanam. Fighting perebor: New and improved algorithms for formula and QBF satisfiability. In *Proceedings of the Fifty-First Annual IEEE Symposium on Foundations of Computer Science*, pages 183–192, 2010.

[Sch94] R.E. Schapire. Learning probabilistic read-once formulas on product distributions. *Machine Learning*, 14(1):47–81, 1994.

[ST12] K. Seto and S. Tamaki. A satisfiability algorithm and average-case hardness for formulas over the full binary basis. In *Proceedings of the Twenty-Seventh Annual IEEE Conference on Computational Complexity*, pages 107–116, 2012.

[Sub61] B.A. Subbotovskaya. Realizations of linear function by formulas using $\vee$, &, $^{-}$. *Doklady Akademii Nauk SSSR*, 136(3):553–555, 1961. English translation in *Soviet Mathematics Doklady*.

[Tal14] A. Tal. Shrinkage of de Morgan formulae by spectral techniques. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*, pages 551–560, 2014.

[Tal15] A. Tal. #SAT Algorithms from Shrinkage. *Electronic Colloquium on Computational Complexity (ECCC)*, 22:114, 2015.

[TX13] L. Trevisan and T. Xue. A derandomized switching lemma and an improved derandomization of AC^0. In *Proceedings of the Twenty-Eighth Annual IEEE Conference on Computational Complexity*, pages 242–247, 2013.

[Uma03] C. Umans. Pseudo-random generators for all hardnesses. *Journal of Computer and System Sciences*, 67(2):419–440, 2003.

[Val84a] L.G. Valiant. Short monotone formulae for the majority function. *Journal of Algorithms*, 5(3):363–366, 1984.

[Val84b] L.G. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

[Wil13] R. Williams. Improving exhaustive search implies superpolynomial lower bounds. *SIAM J. Comput.*, 42(3):1218–1244, 2013.

[Wil14] R. Williams. Nonuniform ACC circuit lower bounds. *J. ACM*, 61(1):2:1–2:32, 2014.

[Wol08] R. de Wolf. *A Brief Introduction to Fourier Analysis on the Boolean Cube.* Number 1 in Graduate Surveys. Theory of Computing Library, 2008.

[Yao82] A.C. Yao. Theory and applications of trapdoor functions. In *Proceedings of the Twenty-Third Annual IEEE Symposium on Foundations of Computer Science*, pages 80–91, 1982.

[Yao85] A.C. Yao. Separating the polynomial-time hierarchy by oracles. In *Proceedings of the Twenty-Sixth Annual IEEE Symposium on Foundations of Computer Science*, pages 1–10, 1985.

[Zan98] F. Zane. *Circuits, CNFs, and Satisfiability.* PhD thesis, UCSD, 1998.

A Proof of Theorem 3.1

For a Boolean function f, a subset S of variables, and a string $r \in \{0,1\}^{|S|}$, denote by $f_{S\leftarrow r}$ the restriction of f where the variables in S are assigned the values given in r. We can combine different restrictions. For example, $f_{S\leftarrow r,\rho}$ means the restriction of f where we assign the values r to the variables in S, and then apply a restriction ρ to the resulting function in variables $[n] \setminus S$.

Now we give the proof of Theorem 3.1, which we re-state first.

Theorem A.1 ([LMN93]). *For arbitrary n-variate Boolean function f, integer $t > 0$ and a real number $0 < p < 1$ such that $pt \geqslant 8$,*

$$\sum_{|A|\geqslant t} \hat{f}(A)^2 \leqslant 2 \cdot \mathbf{Exp}_{\rho\in R_p}\left[\sum_{B\,:\,|B|\geqslant pt/2} \widehat{f_\rho}(B)^2\right].$$

Proof. We have

$$\sum_{|A|\geqslant t} \hat{f}(A)^2 \leqslant 2 \cdot \mathbf{Exp}_S\left[\sum_{A\,:\,|A\cap S|\geqslant pt/2} \hat{f}(A)^2\right] \tag{17}$$

$$= 2 \cdot \mathbf{Exp}_{S,r\in\{0,1\}^{|S^c|}}\left[\sum_{B\,:\,|B|\geqslant pt/2} \widehat{f_{S^c\leftarrow r}}(B)^2\right] \tag{18}$$

$$= 2 \cdot \mathbf{Exp}_{\rho\in R_p}\left[\sum_{B\,:\,|B|\geqslant pt/2} \widehat{f_\rho}(B)^2\right], \tag{19}$$

where the first expectation is over random sets S obtained by choosing each item $i \in [n]$, independently, with probability p; the second expectation is over S as before, and over uniformly random assignment r (for the variables outside of S).

Eq. (19) is by definition. Eq. (18) is proved in Lemma A.2 below. We show Eq. (17) next.

Consider any set A of size at least t. It will contribute $\hat{f}(A)^2$ to the expectation over S for every random set S that intersects A in at least $pt/2$ locations. The expected intersection size between S and A (where each element $i \in [n]$ is put into S with probability p) is $p|A| \geqslant pt$. By Chernoff, almost all sets S will intersect the set A in at least half the expected number of places; by requiring that $pt \geqslant 8$, we get that this holds for at least half of all random sets S. Multiplying this expectation by 2 ensures that each $\hat{f}(A)^2$ is counted at least once. □

Lemma A.2 ([LMN93]). *For a Boolean function f on n variables, an arbitrary subset $S \subseteq [n]$, and an integer k, we have*

$$\sum_{A\,:\,|A\cap S|\geqslant k} \hat{f}(A)^2 = \mathbf{Exp}_{r\in\{0,1\}^{|S^c|}}\left[\sum_{|B|\geqslant k} \widehat{f_{S^c\leftarrow r}}(B)^2\right]. \tag{20}$$

Proof. We start by re-writing the left-hand side of Eq. (20):

$$\sum_{A\,:\,|A\cap S|\geqslant k} \hat{f}(A)^2 = \sum_{B\subseteq S\,:\,|B|\geqslant k}\ \sum_{D\subseteq S^c} \hat{f}(B\cup D)^2. \tag{21}$$

For all sets $B \subseteq S$ and $D \subseteq S^c$, we have

$$\begin{aligned}
\hat{f}(B\cup D) &= \mathbf{Exp}_{x\in\{0,1\}^n}\left[f(x)\cdot\chi_{B\cup D}(x)\right]\\
&= \mathbf{Exp}_{r\in\{0,1\}^{|S^c|},r'\in\{0,1\}^{|S|}}\left[f_{S^c\leftarrow r}(r')\cdot\chi_{(B\cup D)\cap S}(r')\cdot\chi_{(B\cup D)\cap S^c}(r)\right]\\
&= \mathbf{Exp}_{r\in\{0,1\}^{|S^c|}}\left[\chi_D(r)\cdot\mathbf{Exp}_{r'\in\{0,1\}^{|S|}}\left[f_{S^c\leftarrow r}(r')\cdot\chi_B(r')\right]\right]\\
&= \mathbf{Exp}_{r\in\{0,1\}^{|S^c|}}\left[\chi_D(r)\cdot\widehat{f_{S^c\leftarrow r}}(B)\right].
\end{aligned}$$

Therefore, for every fixed $B \subseteq S$, we get

$$\begin{aligned}
\sum_{D\subseteq S^c}\hat{f}(B\cup D)^2 &= \sum_D\left(2^{-|S^c|}\cdot\sum_{r\in\{0,1\}^{|S^c|}}\chi_D(r)\cdot\widehat{f_{S^c\leftarrow r}}(B)\right)^2\\
&= 2^{-2|S^c|}\cdot\sum_{r_1,r_2\in\{0,1\}^{|S^c|}}\widehat{f_{S^c\leftarrow r_1}}(B)\cdot\widehat{f_{S^c\leftarrow r_2}}(B)\cdot\sum_D\chi_D(r_1\oplus r_2),
\end{aligned}$$

where $r_1 \oplus r_2$ denotes the bit-wise XOR of the two strings. Observing that

$$\sum_{D\subseteq S^c}\chi_D(r) = \begin{cases} 2^{|S^c|} & \text{if } r \text{ is an all-zero string}\\ 0 & \text{otherwise}\end{cases},$$

we can continue the above sequence of equalities, getting the following:

$$\begin{aligned}
\sum_{D\subseteq S^c}\hat{f}(B\cup D)^2 &= 2^{-|S^c|}\cdot\sum_{r\in\{0,1\}^{|S^c|}}\widehat{f_{S^c\leftarrow r}}(B)^2\\
&= \mathbf{Exp}_{r\in\{0,1\}^{|S^c|}}\left[\widehat{f_{S^c\leftarrow r}}(B)^2\right].
\end{aligned}$$

Finally, plugging in the last expression into the right-hand side of Eq. (21), we conclude

$$\sum_{A\,:\,|A\cap S|\geqslant k} \hat{f}(A)^2 = \sum_{B\subseteq S\,:\,|B|\geqslant k} \mathbf{Exp}_{r\in\{0,1\}^{|S^c|}}\left[\widehat{f_{S^c\leftarrow r}}(B)^2\right]$$

$$= \mathbf{Exp}_{r\in\{0,1\}^{|S^c|}}\left[\sum_{|B|\geqslant k} \widehat{f_{S^c\leftarrow r}}(B)^2\right],$$

as required. □

ECCC ISSN 1433-8092
http://eccc.hpi-web.de