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

Fourier Concentration from Shrinkage

Russell Impagliazzo* Valentine Kabanets†

November 27, 2013

Abstract

For Boolean functions computed by de Morgan formulas of subquadratic size or read-once de Morgan formulas, we prove a sharp concentration of the Fourier mass on "small-degree" coefficients. For a Boolean function $f : \{0,1\}^n \to \{1,-1\}$ computable by a de Morgan formula of size s, we show that

$$\sum_{A\subseteq[n]\,:\,|A|>s^{1/\Gamma+\epsilon}} \hat{f}(A)^2 \leqslant exp(-s^{\epsilon/3}),$$

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. We prove that this Fourier concentration is essentially optimal.

As an application, we get that subquadratic-size de Morgan formulas have negligible correlation with parity, and are learnable under the uniform distribution, and also lossily compressible, in subexponential time. We also prove that the average sensitivity of a read-once function f on n variables is at most $n^{1/\Gamma+o(1)}$, and is at least $\Omega(n^{1/\Gamma})$.

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, Wil10, Wil11] 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$^+$13, CKS13], and pseudorandom generators (PRGs) [Bra10, IMZ12, GMR$^+$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], 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. We show tight Fourier concentration for small de Morgan formulas, which is similar to the Fourier concentration for AC^0 circuits shown in the celebrated paper by Linial, Mansour, and Nisan [LMN93]. More precisely, we use concentrated shrinkage of de Morgan formulas under random restrictions to show that most of the Fourier mass of such formulas lies

*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

on low-weight coefficients. Here *concentrated* shrinkage 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 *pseudorandom* restrictions), building upon the earlier "shrinkage in expectation" results by [HRY95, Hås98].

As an immediate consequence of this Fourier concentration, 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 (with better parameters for read-once formulas). We provide more details next.

1.1 Our results

1.1.1 Fourier concentration and correlation bounds

The Fourier transform of a Boolean function $f : \{0,1\}^n \to \{1,-1\}$ is a way to express f in the orthogonal basis of functions $\chi_S(x_1,\dots,x_n) = (-1)^{\sum_{i\in S} x_i}$, 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]. We extend the approach of [LMN93] to the case of subquadratic-size de Morgan formulas, which 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]).

Our main result is the following.

Theorem 1.1. *Let $f : \{0,1\}^n \to \{1,-1\}$ be a Boolean function computable by a de Morgan formula of size s. Then, for any constant $0 < \epsilon < 1/2$, and any sufficiently large s,*

$$\sum_{A\subseteq[n]\,:\,|A|>s^{1/\Gamma+\epsilon}} \hat{f}(A)^2 \leqslant exp(-s^{\epsilon/3}),$$

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.

In words, this means that we have a very good approximation (in the ℓ_2 norm) to a de Morgan formula of size s with a polynomial of degree about $s^{1/\Gamma}$, where Γ is the shrinkage exponent for the class of formulas. We also show our Fourier concentration is essentially optimal (Lemma 5.3).

As an immediate corollary of Theorem 1.1, we get that the parity on n bits cannot be computed correctly on more than $1/2 + exp(-s^{\epsilon/3})$ fraction of inputs by any de Morgan formula of size $s < n^{2/(1+2\epsilon)} \leqslant n^{2(1-\epsilon)}$, for any constant $0 < \epsilon < 1/2$.

1.1.2 Learning and compression

As a consequence of such Fourier concentration, we get, similarly to [LMN93], that the class of de Morgan formulas of size s is *learnable* in time $n^{s^{1/\Gamma+\epsilon}}$ to within error $exp(-s^{\Omega(\epsilon)})$, over the uniform distribution, where $\Gamma = 2$ for general de Morgan formulas, and $\Gamma \approx 3.27$ for read-once de Morgan formulas (see Theorem 7.2).

This class of formulas is also *lossily compressible* in the sense of [CKK$^+$13]. That is, given the truth table of a Boolean function $f : \{0,1\}^n \to \{1,-1\}$ which is promised to be computable by an unknown de Morgan (read-once) formula of size s, we can compute in deterministic time $2^{O(n)}$, a Boolean circuit C of size about $n^{s^{1/\Gamma+\epsilon}}$, where Γ is the corresponding shrinkage exponent, such that C agrees with f on all but $exp(-n^{\Omega(1)})$ fraction of n-bit inputs.

1.1.3 Average sensitivity

Informally, the average sensitivity of a Boolean function $f : \{0,1\}^n \to \{1,-1\}$ 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. We also give an almost matching lower bound $\Omega(s^{1/\Gamma})$.

1.2 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 a random restriction is likely to have fewer than t variables (for some parameter t), then all Fourier coefficients of degree at least t are zero (since it's impossible to have a nonzero correlation with the parity function on t variables if your function depends on fewer than t variables). 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.

This approach works directly for the case of *read-once* de Morgan formulas, which are known to shrink with high probability under "pseudorandom" restrictions [IMZ12], and the same analysis of [IMZ12] can be used also for the case of truly random restrictions, yielding an exponentially small error, as shown in our Theorem 6.3 below.

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.

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).

The learning, compression, and average sensitivity results follow immediately from the Fourier concentration, using standard methods (cf. [LMN93]). For the *lower bound* on the average sensitivity of read-once formulas, we use an explicit family of read-once formulas 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).

1.3 Related work

For size s de Morgan formulas, Ganor, Komargodski, and Raz [GKR12] proved the upper bound $\sum_{|A|>s^{1/2}/\epsilon} \hat{f}(A)^2 \leqslant O(\epsilon)$, which is tight for constant $\epsilon > 0$. In contrast, our Theorem 1.1 shows an exponentially small upper bound for somewhat larger sets A, which is essentially tight (cf. Lemma 5.3).

Lee [Lee09] shows (based on a long line of work in the quantum setting [BBC$^+$01, FGG08, ACR$^+$07, RŠ08, Rei09]) that every de Morgan formula of size s can be computed as the $sign(p)$ for a multilinear polynomial p of degree $O(\sqrt{s})$. In particular, this completely resolves a conjecture by O'Donnell and Servedio made in the conference version of [OS10], which implies that Boolean functions computable by size s de Morgan formulas are PAC-learnable in time $O(n^{\sqrt{s}})$. For general de Morgan formulas, this is stronger than our uniform-distribution learning result as it holds for *any* distribution. On the other hand, our proof is completely classical, and we also get a better running time for the case of read-once formulas.

The tight average sensitivity bound $\Theta(\sqrt{s})$ for general de Morgan formulas of size s follows from the work of [Shi00, Lee09] (using the quantum approach); an alternative (classical) proof is also given by [GKR12].

As observed in [KRT13], the quantum-setting work of [BBC$^+$01, Rei11] implies that any de Morgan formula of size $o((n/\log(1/\epsilon))^2)$ has correlation at most $1/2+\epsilon$ with the n-bit parity. This is comparable to the correlation bound implied by our Fourier concentration result (Corollary 7.1).

Remainder of the paper. We state the basics in Section 2, and show how to adapt the approach of [LMN93] in Section 3. We show the required concentrated shrinkage result for de Morgan formulas in Section 4, and use it to derive the Fourier concentration result for such formulas in Section 5. We prove Fourier concentration for read-once formulas in Section 6. In Section 7 we give the applications of the Fourier concentration result to correlation with parity, learning, compression, and average sensitivity for de Morgan formulas. We state some open questions 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). We think of an n-variate *Boolean* function as $\{-1,1\}$-valued, i.e., as $f : \{0,1\}^n \to \{-1,1\}$. For a subset $A \subseteq [n]$, we denote by χ_A the Boolean function mapping $x_1,\ldots,x_n \in \{0,1\}^n$ to the parity $(-1)^{\sum_{i\in A} x_i}$. Let $f : \{0,1\}^n \to \mathbb{R}$ 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* of f at computing χ_A, the parity of the inputs from A.

The *Parseval identity* is $\sum_{A\subseteq[n]} \hat{f}(A)^2 = \mathbf{Exp}_{x\in\{0,1\}^n}[f(x)^2]$. 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 class 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. For a Boolean function f, a subset S of variables, and a string $r \in \{0,1\}^{|S|}$, the notation $f_{S\leftarrow r}$ means 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 $[n]\setminus S$.

3 Fourier concentration via random restrictions

3.1 The starting point

We use the following result of [LMN93]; for completeness, we prove it in Section A of 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 $pt \geqslant 8$,*

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

3.2 Intuition for de Morgan formulas

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 $p^2 s < n$; this is possible for $s < n^2$.

Note that a formula of size s' has 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 $t = 2ps$, $\sum_{|A|>t} \hat{F}(A)^2 \leqslant 2\cdot\gamma$. To be nontrivial, we need $t < n$.

If $s \leqslant n^{2-2\epsilon}$ and $p = n^{\epsilon}/n$, then $t = n^{1-\epsilon}$. (Note that since we need $pt \geqslant 8$, we get $p \geqslant \Omega(1/\sqrt{s})$, and thus, $t \geqslant \Omega(\sqrt{s})$.)

In reality, we don't have such concentrated shrinkage. First, it is not true because a formula may have "heavy" variables (those that appear too frequently in the formula), and if such a heavy variable is missed (assigned $*$) by a p-random restriction, no substantial shrinkage of the formula size will occur. Thus we need to ensure that the heavy variables are always restricted.

Secondly, 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 $p_1, p_2, \dots, p_k$, 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 both $\sum_{A,|A|\geq t} \hat{f}_0(A)^2 \leq \delta$ and $\sum_{A,|A|\geq t} \hat{f}_1(A)^2 \leq \delta$, then $\sum_{A,|A|\geq t+1} \hat{f}(A)^2 \leq \delta$.*

Proof. For $y := 1-2x$, we can write $f = 1/2(1+y)f_0 + 1/2(1-y)f_1 = 1/2(f_0+f_1)+(1/2)(f_0-f_1)y$. Then, for any set A not containing x,

$$\begin{aligned}\hat{f}(A)^2 + \hat{f}(x \cup A)^2 &= (1/2(\hat{f}_0(A) + \hat{f}_1(A)))^2 + (1/2(\hat{f}_0(A) - \hat{f}_1(A)))^2 \\ &= 1/2 \cdot \hat{f}_0(A)^2 + 1/2 \cdot \hat{f}_1(A)^2.\end{aligned}$$

Summing this over all A with $|A| \geq t$ yields at most δ by the assumptions for the restricted functions. Every B of size $\geq t+1$ (containing x or not) is included in this sum. □

So to upperbound the Fourier mass of the coefficients for sets A with $|A| > t$, the idea is to set all "heavy" variables (say, z of them), and upperbound the Fourier mass for each restricted function over the coefficients for sets B with $|B| > t - z$. 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

Håstad [Hås98] showed that the shrinkage exponent for de Morgan formulas is 2:

Lemma 4.1 ([Hås98])**.** *There exists a $c > 0$ 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 $\mu(p, L(F)) = 1 + \log^{3/2} \min\{1/p, L(F)\}$.

Building on Lemma 4.1, Impagliazzo et al. [IMZ12] (see also [KRT13]) proved shrinkage occurs with high probability, not just in expectation. We prove the following version of this result for completeness and with parameters optimized for our application; see Section B in the Appendix.

Lemma 4.2 (implicit in [IMZ12])**.** *There exists a $c > 0$ 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 $0 < p < 1$,*

$$\mathbf{Pr}_{\rho \in R_p}\left[L(F_\rho) \geqslant c \cdot p^2 \cdot \log^{3/2}(1/p) \cdot L\right] \leqslant L(F) \cdot exp\left(-p^6 \cdot L/h\right).$$

5 Fourier concentration of de Morgan formulas

The main result of this section is the following.

Theorem 5.1. *Let $f : \{0,1\}^n \to \{-1,1\}$ be a Boolean function computed by a de Morgan formula $F(x_1,\ldots,x_n)$ of size at most s. Then, for any constant $0 < \epsilon < 1/2$, and any sufficiently large s,*

$$\sum_{|A|>s^{1/2+\epsilon}} \hat{f}(A)^2 \leqslant exp\left(-s^{\epsilon/3}\right).$$

5.1 Proof of the main result

Theorem 5.1 is implied by the following quantitative version (by equating ϵ of Theorem 5.1 with $\epsilon_k/2$ of Theorem 5.2).

Theorem 5.2. *For each integer $k \geqslant 0$ there is a constant b so that, for any Boolean function f computed by a de Morgan formula of size at most s, and for $\epsilon_k = (11/12)^k$ and $t = s^{(1+\epsilon_k)/2}$,*

$$\sum_{|A|>t} \hat{f}(A)^2 \le ks \cdot exp\left(-\Omega\left(\frac{s^{\epsilon_k/4}}{\log^b s}\right)\right),$$

for s sufficiently large.

Proof. The proof is by induction on k. The base case $k = 0$ is trivial, since the formula depends on at most s variables, and the sum is over sets of size larger than s.

Assume the theorem holds for $k \ge 0$. We will prove it for $k+1$. Let $t = s^{(1+\epsilon_{k+1})/2}$. Let $h = \sqrt{s}$. There are at most $\sqrt{s}$ variables that are more than h-heavy in a minimal formula for f. Let f' be any restriction of f assigning values to the heavy variables. We will show that each f' has

$$\sum_{|A|\ge t'} \hat{f'}(A)^2 \le ks \cdot exp\left(-s^{\Omega(\epsilon_k)}\right),$$

where $t' = t/2$. The claim will then follow from Lemma 3.2, since $t > t' + \sqrt{s}$ when $s > 4^{(12/11)^k}$.

For each such f', consider a random restriction $\rho \in R_p$ for $p = s^{-1/24}/(c(\log s)^a)$, for constants a and c to be chosen later. By Theorem 3.1, we get

$$\sum_{A\,:\,|A|\ge t'} \hat{f'}(A)^2 \le 2 \cdot \mathbf{Exp}_\rho\left[\sum_{B\,:\,|B|\ge pt'/2} \hat{f}'_\rho(B)^2\right].$$

By Lemma 4.2, except with probability $s \cdot exp(-p^6 s/h) = s \cdot exp(-\Omega(s^{1/4}/\log^{6a} s))$, the function f'_ρ has formula size at most $s'' = c_1 p^2 s \log^{3/2} s = (c_1/c^2)s^{11/12}(\log s)^{3/2-2a} = s^{11/12}(\log s)^{3/2-2a}$, where we set $c := \sqrt{c_1}$.

Let t'' be the inductive Fourier coefficient size for the theorem with k and the formula size s'', i.e.,

$$\begin{aligned} t'' &= (s'')^{(1+\epsilon_k)/2} \\ &= s^{(11/24)(1+\epsilon_k)} \cdot (\log s)^{(3/4-a)(1+\epsilon_k)} \\ &= s^{11/24+\epsilon_{k+1}/2} \cdot (\log s)^{(3/4-a)(1+\epsilon_k)}. \end{aligned}$$

We claim we can pick a so that $pt'/2$ (the crucial value in Theorem 3.1) is greater than t''. First,

$$\begin{aligned} pt'/2 &= s^{(1+\epsilon_{k+1})/2} \cdot s^{-1/24}/(4c(\log s)^a) \\ &= s^{11/24+\epsilon_{k+1}/2}/(4c(\log s)^a) \\ &= (t''/4c)(\log s)^{a\epsilon_k-(3/4)(1+\epsilon_k)}. \end{aligned}$$

For any $a > (3/4)(1+\epsilon_k)/\epsilon_k$, the exponent of $\log s$ will be greater than zero, so the whole expression will be larger than t'' for sufficiently large s.

Thus, except for the $s \cdot exp(-s^{1/4}/\log^{6a} s)$ fraction of ρ's where shrinkage fails, we have

$$\begin{aligned} \sum_{B\,:\,|B|\geq pt'/2} \hat{f}'_\rho(B)^2 &\leq ks'' \cdot exp\left(-\Omega\left(\frac{(s'')^{\epsilon_k/4}}{\mathsf{poly}\log s''}\right)\right) \\ &\leq ks \cdot exp\left(-\Omega\left(\frac{s^{\epsilon_{k+1}/4}}{\mathsf{poly}\log s}\right)\right). \end{aligned}$$

Since the sum of squares of Fourier coefficients is bounded by 1 (by Parseval's identity), we can add the chance of failure, getting the bound $(k+1)s \cdot exp\left(-\Omega\left(s^{\epsilon_{k+1}/4}/\mathsf{poly}\log s\right)\right)$, as required. □

5.2 Optimality of the Fourier concentration for de Morgan formulas

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 $\sum_{|A|>t} \hat{f}(A)^2$, we need to set $t > \sqrt{s}$.

In fact, as we shall argue, in order to get the upper bound 2^{-n^γ}, for some $\gamma > 0$, we need to set $t > \sqrt{s} \cdot n^{\gamma/2}$. This shows that our Fourier concentration for sub-quadratic size de Morgan formulas, Theorem 5.1, is tight, up to a constant factor in front of the parameter ϵ.

Lemma 5.3. *For any $\gamma > 0$ and $t \leqslant n$, there is a de Morgan formula on n inputs of size $O(t^2/n^\gamma)$ that computes the parity on t bits with advantage 2^{-n^γ}.*

Proof. Consider the following formula $F(x_1, \ldots, x_n)$. Set $m = \lfloor n^\gamma \rfloor$. 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 $O((t^2/m^2) \cdot m) = O(t^2/m)$.

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. □

By Lemma 5.3, a function f computed by a de Morgan formula of size s may have $\hat{f}(A) \geqslant 2^{-n^\gamma}$ for a set A of size $|A| \leqslant t$ for t satisfying $O(t^2/n^\gamma) = s$, i.e., for $t = O(\sqrt{s} \cdot n^{\gamma/2})$. It follows that in order to achieve $\sum_{|A|>t} \hat{f}(A)^2 < 2^{-n^{\epsilon/6}}$, one needs to set $t > \sqrt{s} \cdot n^{\epsilon/12}$.

6 Fourier concentration of read-once de Morgan formulas

Define $\Gamma := 1/\log(\sqrt{5}-1) \approx 3.27$, the optimal shrinkage exponent for read-once de Morgan formulas [PZ93, HRY95]. The main result of this section is the following.

Theorem 6.1. *Let $F(x_1, \ldots, x_n)$ be any read-once de Morgan formula computing the Boolean function $f : \{0,1\}^n \to \{1,-1\}$. Then for every $0 < \epsilon < 1$ such that $\epsilon > \Omega(\sqrt{\log\log n/\log n})$,*

$$\sum_{|A|>n^{1/\Gamma+\epsilon}} \hat{f}(A)^2 \leqslant exp(-n^{\epsilon/3}).$$

The bound in Theorem 6.1 is close to optimal (see Remark 7.6 below). We prove Theorem 6.1 in Section 6.2, after we argue concentrated shrinkage for read-once formulas.

6.1 Concentrated shrinkage for read-once de Morgan formulas

Here we prove the concentrated shrinkage result for read-once formulas, using the approach of [IMZ12]. First we prove a version useful for relatively large parameters p.

Theorem 6.2. *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 $0 < p < 1$:*

$$\mathbf{Pr}_{\rho\in R_p}\left[L(F_\rho) \geqslant d \cdot p^\Gamma \cdot (\log 1/p)^{\Gamma-1} \cdot n\right] \leqslant exp(-d' \cdot p^{12} \cdot n).$$

We give the proof of Theorem 6.2 in Section 6.3. Assuming this theorem, we now derive a shrinkage result for smaller values of p.

Theorem 6.3. *For any read-once de Morgan formula $F(x_1, \ldots, x_n)$ and $p = (n^\epsilon/n)^{1/\Gamma}$, for some $\epsilon \in [0,1]$, we have*

$$\mathbf{Pr}_{\rho\in R_p}\Big[L(F_\rho) > (\log n)^b \cdot p^\Gamma \cdot n\Big] \leqslant exp(-\Omega(n^{\epsilon/2})),$$

for some $b = O(1/\epsilon)$.

Proof. Set $q := n^{-\epsilon/24}$, and $k := 24(1-\epsilon)/(\Gamma\cdot\epsilon)$. We will apply k random q-restrictions to our original formula F. Let F_i be the formula F after i restrictions are applied to F, and let s_i be the size of F_i; we have $F_0 = F$ and $s_0 = n$.

Consider stage i, for $1 \leqslant i \leqslant k$. If $s_{i-1} \leqslant p^\Gamma n = n^\epsilon$, then s_i will also be less than n^ϵ with probability 1. Assuming $s_{i-1} \geqslant n^\epsilon$, we get by Theorem 6.2 that $\mathbf{Pr}_{\rho\in R_q}[s_i \geqslant d\cdot q^\Gamma(\log 1/q)^{\Gamma-1}\cdot s_{i-1}] \leqslant exp(-d'\cdot q^{12}\cdot n^\epsilon) \leqslant exp(-d'n^{\epsilon/2})$. It follows that with probability at least $1-k\cdot exp(-d'n^{\epsilon/2})$, $s_k \leqslant (d\cdot(\log^3 n))^k q^{k\Gamma} n = (\log n)^b \cdot p^\Gamma \cdot n = (\log n)^b \cdot n^\epsilon$, for some $b = O(1/\epsilon)$. □

6.2 Proof of Theorem 6.1

For $p = (n^\epsilon/n)^{1/\Gamma}$, we get by Theorem 6.3 that for all but $\gamma := exp(-\Omega(n^{\epsilon/2}))$ fraction of p-restrictions shrink the formula F to size at most $(\log n)^b \cdot n^\epsilon$. Set t so that $(\log n)^b \cdot n^\epsilon = pt/2$. We get $t = n^{1/\Gamma+\epsilon(1-1/\Gamma)} \cdot (\log n)^{O(1/\epsilon)} \leqslant n^{1/\Gamma+\epsilon}$, for ϵ such that $(\log n)^{O(1/\epsilon)} \leqslant n^{\epsilon/3}$, which holds for $\epsilon > \Omega(\sqrt{\log\log n/\log n})$. By Theorem 3.1, we conclude that $\sum_{|A|>t}\hat{f}(A)^2 \leqslant 2\gamma \leqslant exp(-\Omega(n^{\epsilon/2})) \leqslant exp(-n^{\epsilon/3})$, as required.

6.3 Proof of Theorem 6.2

For the proof, we need the following shrinkage result of [HRY95].

Theorem 6.4 ([HRY95])**.** *For every read-once formula $F(x_1, \ldots, x_n)$ and a parameter $0 < p < 1$,*

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

The proof idea for Theorem 6.2 is to decompose a given formula into independent subformulas (with some extra conditions) and apply Theorem 6.4 to each subformula. Since the subformulas are independent, we can use the Chernoff-Hoeffding inequality to argue that the shrinkage occurs with high probability.

Lemma 6.5 ([IMZ12])**.** *For every positive ℓ and any read-once de Morgan formula F on the set X of variables with $L(F) \geqslant \ell$, there exist m read-once de Morgan formulas $G_1, \ldots, G_m$ for $L(F)/\ell \leqslant m \leqslant 6L(F)/\ell$, such that*

1. *$L(G_i) \leqslant \ell$, for all $1 \leqslant i \leqslant m$,*
2. *for each $1 \leqslant i \leqslant m$, G_i depends on at most 2 "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.*

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 6.5. 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 replace each special variable with a restriction-resistant read-once formula (on new variables, different for different special variables), and then use a standard random restriction on the resulting, slightly larger read-once formula. The following lemma (building on the work by Valiant [Val84a]) shows the existence of requisite read-once formulas; see Section C in the appendix for the proof.

Lemma 6.6 ([IMZ12])**.** *For every $0 < p < 1$, there exists a read-once de Morgan formula H of size $O(1/p^4)$ such that, for all but at most $1/4$ of p-restrictions ρ, we have*

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

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

Now we can analyze the expected shrinkage of read-once de Morgan formulas under p-restrictions that leave some specified variables unrestricted.

Lemma 6.7. *[IMZ12] There exist constants c and c' such that the following holds. Let G be a read-once de Morgan formula with at most 2 special variables, and let ρ' be a p-restriction such that the special variables in G remain unrestricted. If $p \geqslant (c/L(G))^{1/4}$, then*

$$\mathbf{Exp}_{\rho'}\left[L(G_\rho)\right] \leqslant c' \cdot p^\Gamma \cdot (\log 1/p)^{\Gamma-1} \cdot L(G).$$

Proof. Let G' be a read-once formula obtained from G by replacing each special variable of G with a fresh copy of the read-once formula H given by Lemma 6.6. Let A be the event that a random p-restriction ρ on the input variables of these copies of H satisfies Eq. (1). By Lemma 6.6, $\mathbf{Pr}[A] \geqslant 1 - 2 \cdot (1/4) = 1/2$, since G has at most two special variables.

Let ρ be a p-restriction over the variables of G' (which include the original non-special variables of G and the new variables of the copies of H), and let ρ' be a restriction over the variables of G that agrees with ρ on all non-special variables of G and leaves the special variables of G unrestricted. Conditioned on A, we have $L(G'_\rho) \geqslant L(G_{\rho'})$. Hence, we get

$$\begin{aligned}\mathbf{Exp}_\rho[L(G'_\rho)] &\geqslant \mathbf{Pr}[A] \cdot \mathbf{Exp}_\rho[L(G'_\rho) \mid A] \\ &\geqslant (1/2) \cdot \mathbf{Exp}_{\rho'}[L(G_{\rho'} \mid A] \\ &= (1/2) \cdot \mathbf{Exp}_{\rho'}[L(G_{\rho'})],\end{aligned}$$

which implies that $\mathbf{Exp}_{\rho'}[L(G_{\rho'})] \leqslant 2 \cdot \mathbf{Exp}_{\rho}[L(G'_{\rho})]$. Finally, applying Theorem 6.4 to the formula G', we get for some constant $c'' > 0$ that

$$\mathbf{Exp}_{\rho'}[L(G_{\rho'})] \leqslant c'' \cdot p^{\Gamma} \cdot (\log 1/p)^{\Gamma-1} \cdot L(G').$$

Note that $L(G') \leqslant L(G)+O(1/p^4)$, which can be made at most $2 \cdot L(G)$ by choosing $p \geqslant (c/L(G))^{1/4}$ for a sufficiently large constant $c > 0$. The lemma follows for the constant $c' = 2 \cdot c''$. □

Finally, we are ready to prove Theorem 6.2.

Proof of Theorem 6.2. Set $\ell := c/p^4$, for c as in Lemma 6.7. Using Lemma 6.5, partition a given formula F (of size n) into $n/\ell \leqslant m \leqslant 6n/\ell$ subformulas $G_1, \dots, G_m$ of size at most ℓ each. Using Lemma 6.7, we get for each G_i that

$$\mathbf{Exp}_{\rho'}[L((G_i)_{\rho'})] \leqslant c' \cdot p^{\Gamma} \cdot (\log 1/p)^{\Gamma-1} \cdot \ell,$$

where ρ' is a p-restriction leaving the special variables unrestricted.

Define random variables $X_i := L((G_i)_{\rho'})$ for each $1 \leqslant i \leqslant m$. Note that, by Lemma 6.5, we have $L(F_\rho) \leqslant \sum_{i=1}^{m} X_i$, where ρ is a random p-restriction over $x_1, \dots, x_n$ which agrees with ρ' on each x_i, $1 \leqslant i \leqslant n$.

Since F is a read-once formula, we get that all G_i's have disjoint sets of variables, and hence, the random variables $X_1, \dots, X_m$ are independent. Also, we have for each $1 \leqslant i \leqslant m$, $\mathbf{Exp}_{\rho'}[X_i] \leqslant \mu$, where $\mu = c' \cdot p^{\Gamma} \cdot (\log 1/p)^{\Gamma-1} \cdot \ell$. Using the Hoeffding inequality, we get

$$\begin{aligned}
\mathbf{Pr}_{\rho'}\left[\sum_{i=1}^{m} X_i \geqslant 2 \cdot m\mu\right] &\leqslant \mathbf{Pr}_{\rho'}\left[\sum_{i=1}^{m} X_i - \mathbf{Exp}\left[\sum_{i=1}^{m} X_i\right] \geqslant m \cdot \mu\right] \\
&\leqslant exp\left(\frac{-2(m \cdot \mu)^2}{m \cdot \ell^2}\right) \\
&\leqslant exp(-2\mu^2 n/\ell^3) \\
&= exp(-\Omega(p^{12} \cdot n)),
\end{aligned}$$

where we used the facts that $0 \leqslant X_i \leqslant \ell$, $m \geqslant n/\ell$, $\mu \geqslant 1$, and $\ell = c/p^4$.

So, with high probability, $L(F_\rho) \leqslant \sum_{i=1}^{m} X_i \leqslant 2m\mu \leqslant 2c'p^{\Gamma}(\log 1/p)^{\Gamma-1}\ell(6n/\ell)$, as required. □

7 Applications

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 $s = n^{2-2\epsilon}$, for some $0 < \epsilon < 1/2$, agrees with the parity function on n bits on at most $1/2 + exp(-s^{\epsilon/3})$ 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 essentially optimal.

7.2 Learning

As in [LMN93], the Fourier concentration result yields a learning algorithm for Boolean functions f computable by small de Morgan formulas, where the learner is given labeled examples $(x, f(x))$ for the uniform distribution over inputs x. The learning algorithm produces a function g such that g agrees with f on almost all inputs x. The error of the learning algorithm (i.e., the fraction of inputs where g and f disagree) and its running time depend on the parameters of the Fourier concentration result for the corresponding model of computation.

Theorem 7.2. *Under the uniform distribution, one can learn, to within error $exp(-n^{\Omega(\epsilon)})$ for any $0 < \epsilon < 1/2$, Boolean functions $f : \{0,1\}^n \to \{1,-1\}$ computable by formulas of size s in time $exp(s^{1/\Gamma+\epsilon})$, where $\Gamma = 2$ for general de Morgan formulas and $\Gamma \approx 3.27$ for read-once formulas.*

Proof sketch. The proof mimics the analogous result in [LMN93] for AC^0 circuits. Namely, for the case of a de Morgan formula of size s on n inputs that computes a Boolean function f, we approximate f with the Fourier expansion truncated at the degree $d := s^{1/2+\epsilon}$, denoted $\tilde{f}$. The normalized squared ℓ_2-norm of the difference $\|f - \tilde{f}\|^2/2^n$ is by Theorem 5.1 at most $\gamma := exp(-s^{\epsilon/3})$. It is easy to see that the function $sign(\tilde{f})$, the sign of $\tilde{f}$, agrees with f on all but at most γ fraction of inputs in $\{0,1\}^n$. We can learn $\tilde{f}$ by estimating all Fourier coefficients $\hat{f}(A)$, for $|A| \leqslant d$, through random sampling (assuming the uniform distribution on the samples $(x, f(x))$). It follows that one can learn a function $sign(p)$, for some degree d multilinear polynomial p, that agrees with f in all but γ fraction of inputs, where the learning algorithm takes time polynomial in the number of Fourier coefficients of degree at most d, i.e., at most $poly(n^d) = exp(s^{1/2+\epsilon} \cdot \log n)$.

For read-once de Morgan formulas, we proceed in the same way as above, but using Theorem 6.1. This gives us degree $d = n^{1/\Gamma+\epsilon}$, and hence, the running time of the learning algorithm $poly(n^d) = exp(n^{1/\Gamma+\epsilon} \cdot \log n)$, as required. □

As mentioned earlier, using the quantum-setting results on the sign degree of de Morgan formulas [Lee09], one gets a PAC-learning algorithm for size s de Morgan formulas that runs in time $n^{O(\sqrt{s})}$. This is better than our uniform-distribution learning algorithm of Theorem 7.2; however, our result is proved in the classical setting. On the other hand, for read-once de Morgan formulas, our learning algorithm appears to be the fastest known. Previously, a comparable running time $\approx exp(n^{1/3})$ was known for (read-once) DNFs, albeit in the stronger setting of PAC learning [KS04].

7.3 Compression

Given the truth table of a function f computable by a de Morgan formula of size s on n inputs, we can compute in time $poly(2^n)$ all Fourier coefficients of f, and then define the approximation $\tilde{f}$ that is the truncated version of the Fourier expansion of f of degree at most $d = s^{1/2+\epsilon}$. As above, the function $g := sign(\tilde{f})$ is the Boolean function that agrees with f on all but at most $exp(-n^{\Omega(\epsilon)})$ fraction of inputs. The size of the circuit computing g is at most $poly(n^d)$, which is less than $2^n/n$ for $s < n^{2-2\epsilon}$, for $0 < \epsilon < 1/2$.

In the language of [KK13, CKK$^+$13], this means that we have a deterministic *lossy-compression* algorithm for the class of de Morgan formulas of sub-quadratic size.[1] Similarly, we get, for read-once de Morgan formulas on n inputs, a lossy-compression algorithm producing a circuit of size at most $exp(n^{1/3+\epsilon})$ which agrees with the formula on all but at most $exp(-n^{\Omega(\epsilon)})$ fraction of inputs.

[1] Using the quantum results for de Morgan formulas [Lee09], one gets a *lossless* compression algorithm for size s de Morgan formulas that produces a circuit of size $exp(\sqrt{s} \log n)$, agreeing with the de Morgan formula on all inputs.

7.4 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

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

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. Combining the result of [Lee09] on the approximate degree of size s de Morgan formulas being $O(\sqrt{s})$, with the result of Shi [Shi00] that approximate degree upperbounds the average sensitivity, we immediately get the matching $O(\sqrt{s})$ upper bound on the average sensitivity of size s de Morgan formulas. Ganor et al. [GKR12] give an alternative proof of this upper bound, using completely classical (non-quantum) arguments.

We show a stronger upper bound on the average sensitivity for *read-once* formulas, and also argue that it is almost tight. Recall that $\Gamma = 1/\log_2(\sqrt{5}-1) \approx 3.27$ is the shrinkage exponent for read-once formulas.

Theorem 7.3. *Let $f : \{0,1\}^n \to \{1,-1\}$ be a Boolean function computed by a read-once de Morgan formula. Then, for all sufficiently large n, $AS(f) \leqslant n^{1/\Gamma + o(1)}$.*

Proof. By Eq. (2) and Theorem 6.1, we get that, for every $\epsilon > \Omega(\sqrt{\log\log n/\log n})$,

$$\begin{aligned} AS(f) &= \sum_{|A| \leqslant n^{1/\Gamma+\epsilon}} |A| \cdot \hat{f}(A)^2 + \sum_{|A| > n^{1/\Gamma+\epsilon}} |A| \cdot \hat{f}(A)^2 \\ &\leqslant n^{1/\Gamma+\epsilon} + n \cdot \sum_{|A| > n^{1/\Gamma+\epsilon}} \hat{f}(A)^2 \\ &\leqslant n^{1/\Gamma+\epsilon} + n \cdot exp(-n^{\epsilon/3}), \end{aligned}$$

where we used the Parseval identity. Choose $\epsilon := c\sqrt{\log\log n/\log n}$, for a sufficiently large constant $c > 0$. Then $n \cdot exp(-n^{\epsilon/3}) \leqslant 1$, for sufficiently large n. We get $AS(f) \leqslant n^{1/\Gamma+\epsilon} + 1$, as required. □

Next we argue that our average sensitivity bound for read-once formulas is tight, up to the factor $n^{o(1)}$. We will need a variant of Valiant's read-once formula defined in [PZ93] (where it was used to argue that the the shrinkage exponent for read-once formulas can be at most Γ). The formula is defined as follows. Let $\{r_n\}$ be a sequence of bits. Set $F_{1,0} = x_1 \wedge x_2$, and $F_{1,1} = x_1 \vee x_2$. For $n > 1$, define $F_{n,b} = F_{n-1,1-b}$ NAND F_{n-1,r_n}, for $b \in \{0,1\}$. Here we use disjoint sets of variables for all of our subformulas.

Let $p_{n,b}$ denote the probability that $F_{n,b} = 1$ on a uniformly random input.

Lemma 7.4 ([PZ93])**.** *There exists a sequence $\{r_n\}$ such that*

1. $p_{n,0}, p_{n,1} \to \psi := \frac{\sqrt{5}-1}{2} \approx 0.62$, *as $n \to \infty$,*
2. $p_{n,0} < \psi < p_{n,1}$ *for every $n \geqslant 1$, and*
3. *there exists a constant $c > 0$ such that for every $n \geqslant 1$, $p_{n,1} - p_{n,0} < c \cdot \psi^n$.*

For $N = 2^n$, we define $F_N(x_1, \dots, x_N) := F_{n,0}(x_1, \dots, x_N)$, for $F_{n,0}$ given by Lemma 7.4. We have the following.

Theorem 7.5. *For all large enough $N = 2^n$, the read-once formula $F_N(x_1, \dots, x_N)$ is such that $AS(f_N) \geqslant \Omega(N^{1/\Gamma})$, where f_N is the $\{1, -1\}$-valued Boolean function computed by the formula F_N.*

Proof. For an N-variate Boolean function f, $AS(f) = \mathbf{Exp}_{w \in \{0,1\}^N} \left[\sum_{i=1}^{N} \chi(i, w) \right]$, where $\chi(i, w) = 1$ if $f(w_1, \dots, w_{i-1}, x_i, w_{i+1}, \dots, w_N)$ depends on x_i, and 0 otherwise. By the linearity of expectation, we have

$$AS(f) = \sum_{i=1}^{N} \mathbf{Pr}_{w \in \{0,1\}^N}[f(w_1, \dots, w_{i-1}, x_i, w_{i+1}, \dots, w_N) \text{ depends on } x_i],$$

where the ith probability expression in the summation above is also known as the influence of coordinate i on f, denoted $\mathbf{Inf}_i[f]$. We will show for our read-once formulas F_N computing the Boolean functions f_N that, for each $1 \leqslant i \leqslant N$, $\mathbf{Inf}_i[f_N] \geqslant \Omega(N^{1/\Gamma - 1})$, concluding the proof.

Consider an arbitrary leaf x_i, $1 \leqslant i \leqslant N$, of F_N. For the path in the formula from the leaf x_i to the root, all non-leaf nodes $v_1, \dots, v_n$ are NAND gates, except for the gate v_1 closest to the leaf x_i which may be either AND or OR. For each node v_j, let G_{v_j} be the subformula of F_N corresponding to the node u feeding into the gate v_j such that u is not on the path $v_1, \dots, v_n$.

For a given assignment w to the $N-1$ variables $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_N$, the restricted function is non-constant (i.e., depends on x_i) iff each formula G_{v_j}, for $2 \leqslant j \leqslant n$, evaluates to 1 under the assignment w, while G_{v_1} evaluates to 1 if v_1 is an AND gate, or to 0 if v_1 is an OR gate. Since these formulas G_{v_j}'s depend on disjoint sets of variables, we get that the probability over w that the variable x_j survives w is $(1/2) \cdot p_2 \cdot \dots \cdot p_n$, where, for $2 \leqslant j \leqslant n$, $p_j = \mathbf{Pr}_z[G_{v_j}(z) = 1]$, for z is a uniformly random assignment to the variables of G_{v_j}.

For $2 \leqslant j \leqslant n$, we have by construction that $G_{v_j} = F_{j-1,b_j}$ for some $b_j \in \{0, 1\}$. Hence, for each $2 \leqslant j \leqslant n$, we get by Lemma 7.4 that $p_j \geqslant p_{j-1,0}$. Thus, x_i survives a random assignment w with probability at least $(1/2) \prod_{j=1}^{n-1} p_{j,0}$.

Let $d \in \mathbb{N}$ be a smallest constant such that $c \cdot \psi^d \leqslant 1/2$, where c and ψ are as in Lemma 7.4. Note that, for each $j > d$, we have by Lemma 7.4 that $p_{j,0} > \psi - c\psi^j = \psi(1 - c\psi^{j-1})$. Thus, we get for some constant $\alpha > 0$ that

$$\begin{aligned} (1/2) \prod_{j=1}^{n-1} p_{j,0} &= \alpha \cdot \prod_{j=d+1}^{n-1} p_{j,0} \\ &\geqslant \alpha \cdot \prod_{j=d+1}^{n-1} \psi(1 - c \cdot \psi^{j-1}) \\ &= (\alpha/\psi^{d+1}) \cdot \psi^n \cdot \prod_{j=d+1}^{n-1} (1 - c \cdot \psi^{j-1}). \end{aligned}$$

Using the inequality $1 - y \geqslant e^{-2y}$ valid for all $0 \leqslant y \leqslant 1/2$, we get

$$\prod_{j=d+1}^{n-1} (1 - c \cdot \psi^{j-1}) \geqslant e^{-2c \sum_{j=d}^{n-2} \psi^j} \geqslant e^{-2c\psi^d/(1-\psi)},$$

which is some positive constant. Hence, the probability that x_i survives a random assignment w to the other variables of F_N is at least $\beta \cdot \psi^n$ for some constant $\beta > 0$. Finally, we have $\psi^n = 2^{n \log_2 \psi} = N^{\log_2(2\psi) - 1} = N^{1/\Gamma - 1}$, which concludes the proof. □

Remark 7.6. As a consequence of Theorem 7.5, to achieve $\sum_{|A|>t} \hat{f}(A)^2 \leqslant 1/n$, for a Boolean function $f : \{0,1\}^n \to \{1,-1\}$ computable by some read-once de Morgan formula, one needs $t \geqslant \Omega(n^{1/\Gamma})$. Thus, our Fourier concentration bound in Theorem 6.1 is close to optimal.

8 Open questions

We believe the average sensitivity for read-once formulas of size s is $\Theta(s^{1/\Gamma})$, where Γ is the shrinkage exponent for read-once formulas. Can one remove the extra $n^{o(1)}$ factor from Theorem 7.3?

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? Note that for general de Morgan formulas of size n, the corresponding statement follows from the quantum results on the approximate degree $O(\sqrt{s})$ [Lee09]. Observe that 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.

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.

[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.

[CKK$^+$13] R. Chen, V. Kabanets, A. Kolokolova, R. Shaltiel, and D. Zuckerman. Mining circuit lower bound proofs for meta-algorithms. *Electronic Colloquium on Computational Complexity*, 20(57), 2013.

[CKS13] R. Chen, V. Kabanets, and N. Saurabh. An improved deterministic #SAT algorithm for small de Morgan formulas. *Electronic Colloquium on Computational Complexity (ECCC)*, 20(150), 2013.

[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.

[GKR12] A. Ganor, I. Komargodski, and R. Raz. The spectrum 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.

[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.

[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.

[KK13] V. Kabanets and A. Kolokolova. Compression of Boolean functions. *Electronic Colloquium on Computational Complexity*, 20(24), 2013.

[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.

[KS04] A.R. Klivans and R.A. Servedio. Learning DNF in time $2^{\tilde{O}(n^{1/3})}$. *Journal of Computer and System Sciences*, 68(2):303–318, 2004.

[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.

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

[OS10] R. O'Donnell and R. Servedio. New degree bounds for polynomial threshold functions. *Combinatorica*, 30(3):327 – 358, 2010.

[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.

[Shi00] Y. Shi. Lower bounds of quantum black-box complexity and degree of approximating polynomials by influence of boolean variables. *Information Processing Letters*, 75(12):79 – 83, 2000.

[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*.

[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. (preliminary version in STOC'02).

[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.

[Wil10] R. Williams. Improving exhaustive search implies superpolynomial lower bounds. In *Proceedings of the Forty-Second Annual ACM Symposium on Theory of Computing*, pages 231–240, 2010.

[Wil11] R. Williams. Non-uniform ACC circuit lower bounds. In *Proceedings of the Twenty-Sixth Annual IEEE Conference on Computational Complexity*, pages 115–125, 2011.

[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

We first re-state Theorem 3.1.

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

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

Proof. We have

$$\sum_{|A|>t} \hat{f}(A)^2 \leqslant 2 \cdot \mathbf{Exp}_S\left[\sum_{A\,:\,|A\cap S|>pt/2} \hat{f}(A)^2\right] \tag{3}$$

$$= 2 \cdot \mathbf{Exp}_{S,r\in\{0,1\}^{|S^c|}}\left[\sum_{B\,:\,|B|>pt/2} \hat{f}_{S^c\leftarrow r}(B)^2\right] \tag{4}$$

$$= 2 \cdot \mathbf{Exp}_{\rho\in R_p}\left[\sum_{B\,:\,|B|>pt/2} \hat{f}_\rho(B)^2\right], \tag{5}$$

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).

The last equality, Eq. (5), is by definition. The second equality, Eq. (4), is proved in Lemma A.2 below. It remains to argue the first inequality, Eq. (3).

Consider any set A of size greater than t. It will contribute $\hat{f}(A)^2$ to the expectation over S for every random set S that intersects A in more than $pt/2$ locations. The expected intersection size between S and A (where each element $i \in [n]$ is put into S with probability p) is $p|A| > pt$. 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*

$$\sum_{A\,:\,|A\cap S|>k} \hat{f}(A)^2 = \mathbf{Exp}_{r\in\{0,1\}^{|S^c|}}\left[\sum_{|B|>k} \hat{f}_{S^c\leftarrow r}(B)^2\right]. \tag{6}$$

Proof. We start by re-writing the left-hand side of Eq. (6):

$$\sum_{A\,:\,|A\cap S|>k} \hat{f}(A)^2 = \sum_{B\subseteq S\,:\,|B|>k}\ \sum_{D\subseteq S^c} \hat{f}(B\cup D)^2. \tag{7}$$

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

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

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

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

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:

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

Finally, plugging in the last expression into the right-hand side of Eq. (7), we conclude

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

as required. □

B Proof of Lemma 4.2

We re-state Lemma 4.2 first.

Lemma B.1. *There exists a $c > 0$ 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 $0 < p < 1$,*

$$\mathbf{Pr}_{\rho \in R_p}[L(F_\rho) \geqslant c \cdot p^2 \cdot \log^{3/2}(1/p) \cdot L] \leqslant L(F) \cdot exp(-p^6 \cdot L/h).$$

Proof. Our proof is based on that from [IMZ12]. Let $s = c_0 p^{-2}$ for some constant c_0 . First, we partition F into $O(L(F)/s)$ subformulas, using Lemma 6.5 (which works also for general de Morgan formulas, not just read-once formulas). For completeness, we sketch this argument below.

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. (Take a minimal subformula with size greater than $s/2$. 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 s nodes and create at most 1 new subtree variable, or reduce the number of subtree variables by one, we get a partition of the formula into $O(L(F)/s)$ subformulas each of size at most s and with at most 2 subtree variables. Note that each subtree variable only occurs in one of the other subformulas in the partition, and only once within that one.

Next, consider any subformula in our partition F^j, with subtree variables t, T. Since each of t and T occurs only once in F^j, F^j is monotone in t or its negation, and in T or its negation. Let F_l represent the function F restricted by setting literal l to true. Assume without loss of generality that F^j is monotone in the positive direction for both variables. We can then write F^j as

$$F^j_{\neg t,\neg T} \vee \left(t \wedge \left(F^j_{t,\neg T} \vee \left(T \wedge F^j_{t,T}\right)\right)\right). \tag{8}$$

(A similar decomposition holds, and is simpler, if the subformula has one subtree variable; and if it has none, we leave it alone.) The same is true after restriction. Note that each of the restricted functions in Eq. (8) has formula size at most s and does not depend on subtree variables. In addition, each variable occurs in at most $3h$ of them.

Restrict all of the subformulas in the collection by a random restriction ρ. We can then create formulas for each subtree in our partition by replacing the subtree variables in the decompositions given by Eq. (8) with the corresponding recursively constructed formulas. Since each subtree variable appears at most once, the reconstructed formula has size at most the sum of the sizes of all the restricted sub-formulas $F^j_{l_1,l_2,\rho}$, for the various subtree literals l_1, l_2.

Thus, we have a collection of $O(L(F)/s)$ formulas G^j each of size at most s so that no variable appears in more than $3h$ of the G^j's and so that $L(F_\rho) \leq \sum L(G^j_\rho)$. So our lemma reduces to showing concentration for the latter sum.

Since each G^j shares any variables with at most $3sh$ other G^k, we can partition the G^j'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, sharing no variables in common.

By Håstad's Lemma 4.1, for each G^j, $\mathbf{Exp}[L(G^j_\rho)] \leq c_1 \cdot \log^{3/2} s$, for some constant c_1. So the expected total formula size within each of the batches is $O(L(f)\log^{3/2} s/(s^2h))$. As a random variable, this is the sum of independent random variables in the range 0 to s. We use the following concentration lemma: If $X = \sum X_i$ is the sum of independent variables in the range $0..s$ and $\mathbf{Exp}[X] < E$, then the probability that $X > 8E$ is at most $2 \cdot 2^{-E/s}$. In our case, this gives an upper bound on the probability that the sum of the formula sizes in any batch is larger than $c_3 L \log^{3/2} s/(s^2h)$ of $2^{-\Omega(L\log^{3/2} s/s^3h)}$. There are strictly less than $L(F) \leq L$ batches, so a union bound gives the probability that all batches are of size $O(L\log^{3/2} s/(s^2h))$ except with probability $L \cdot \exp(-\Omega(L(F)/(s^3h))) = L \cdot \exp(-\Omega(p^6 L(F)/h))$. If they are, then summing up over the at most $O(sh)$ batches, $L(F_\rho) \leq O(L\log^{3/2} s/s) = O(p^2 \cdot L \cdot \log^{3/2}(1/p))$. □

C Proof of Lemma 6.6

We re-state Lemma 6.6.

Lemma C.1 ([IMZ12]). *For every $0 < p < 1$, there exists a read-once de Morgan formula H of size $O(1/p^4)$ such that, for all but at most $1/4$ of p-restrictions ρ, we have*

$$H_\rho(\vec{0}) = 0 \qquad \text{and} \qquad H_\rho(\vec{1}) = 1,$$

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

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.[2]

Theorem C.2 ([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 $p \in [0,1]$,*

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

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 C.1. We use Theorem C.2 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.

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 C.2, we get

$$\mathbf{Pr}_{F \in D_k, \rho \in R_p}[F_\rho(\vec{1}) = 1] = A_{F_k}(\psi + (1-\psi)p) > 7/8. \tag{9}$$

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 C.2, we get

$$\mathbf{Pr}_{F \in D_k, \rho \in R_p}[F_\rho(\vec{0}) = 1] = A_{F_k}(\psi - (1-\psi)p) < 1/8. \tag{10}$$

Using Eqs. (9) and (10), we get by the union bound that

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

Finally, by averaging, there exists a particular read-once formula $H \in D_k$ such that, for all but less than 1/4 of 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 $O(1/p^{\Gamma}) \leqslant O(1/p^4)$. □

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