Electronic Colloquium on Computational Complexity, Report No. 17 (2019)

Fourier bounds and pseudorandom generators for product tests

Chin Ho Lee[*]

February 5, 2019

Abstract

We study the Fourier spectrum of functions $f\colon \{0,1\}^{mk} \to \{-1,0,1\}$ which can be written as a product of k Boolean functions f_i on disjoint m-bit inputs. We prove that for every positive integer d,

$$\sum_{S\subseteq[mk]:|S|=d} |\hat{f_S}| = O(m)^d.$$

Our upper bound is tight up to a constant factor in the $O(\cdot)$. Our proof builds on a new "level-d inequality" that bounds above $\sum_{|S|=d} \hat{f_S}^2$ for any $[0,1]$-valued function f in terms of its expectation, which may be of independent interest.

As a result, we construct pseudorandom generators for such functions with seed length $\tilde{O}(m + \log(k/\varepsilon))$, which is optimal up to polynomial factors in $\log m$, $\log\log k$ and $\log\log(1/\varepsilon)$. Our generator in particular works for the well-studied class of combinatorial rectangles, where in addition we allow the bits to be read in any order. Even for this special case, previous generators have an extra $\tilde{O}(\log(1/\varepsilon))$ factor in their seed lengths.

Using Schur-convexity, we also extend our results to functions f_i whose range is $[-1,1]$.

[*] Northeastern University. Supported by NSF CCF award 1813930. Work done in part while visiting Amnon Ta-Shma at Tel Aviv University, with support from the Blavatnik family fund and ISF grant no. 952/18.

ISSN 1433-8092

1 Introduction

In this paper we study tests on n bits which can be written as a product of k bounded real-valued functions defined on disjoint inputs of m bits. We first define them formally.

Definition 1 (Product tests). *A function $f\colon \{0,1\}^n \to [-1,1]$ is a* product test *with k functions of input length m if there exist k disjoint subsets $I_1, I_2, \ldots, I_k \subseteq \{1,2,\ldots,n\}$ of size $\le m$ such that $f(x) = \prod_{i \le k} f_i(x_{I_i})$ for some functions f_i with range in $[-1,1]$. Here x_{I_i} are the $|I_i|$ bits of x indexed by I_i.*

More generally, the range of each function f_i can be $\mathbb{C}_{\le 1} := \{z \in \mathbb{C} : |z| = 1\}$, the complex unit disk [GKM15, HLV18], or the set of square matrices over a field [RSV13]. However, in this paper we only focus on the range $[-1,1]$. As we will soon explain, our results do not hold for the broader range of $\mathbb{C}_{\le 1}$.

The class of product tests was first introduced by Gopalan, Kane and Meka under the name of *Fourier shapes* [GKM15]. However, in their definition, the subsets I_i are fixed. Motivated by the recent constructions of pseudorandom generators against *unordered* tests, which are tests that read input bits in arbitrary order [BPW11, IMZ12, RSV13, SVW14], Haramaty, Lee and Viola [HLV18] considered the generalization in which the subsets I_i can be arbitrary as long as they are of bounded size and pairwise disjoint.

Product tests generalize several restricted classes of tests. For example, when the range of the functions f_i is $\{0,1\}$, product tests correspond to the AND of disjoint Boolean functions, also known as the well-studied class of *combinatorial rectangles* [AKS87, Nis92, NZ96, INW94, EGL+98, ASWZ96, Lu02, Vio14, GMR+12, GY14]. When the range of the f_i is $\{-1,1\}$, they correspond to the XOR of disjoint Boolean functions, also known as the class of *combinatorial checkerboards* [Wat13]. More importantly, product tests also capture *read-once space computation*. Specifically, Reingold, Steinke and Vadhan [RSV13] showed that the class of read-once width-w branching programs can be encoded as product tests with outputs $\{0,1\}^{w\times w}$, the set of $w \times w$ Boolean matrices.

In the past year, the study of product tests [HLV18, LV17] has found applications in constructing state-of-the-art pseudorandom generators (PRGs) for space-bounded algorithms. Using ideas in [GMR+12, GY14, LV17, CHRT18], Meka, Reingold and Tal [MRT18] constructed a pseudorandom generator for width-3 read-once branching programs (ROBPs) on n bits with seed length $\tilde{O}(\log n \log(1/\varepsilon))$, giving the first improvement of Nisan's generator [Nis92] in the 90s. Building on [RSV13, HLV18, CHRT18], Forbes and Kelley significantly simplified the analysis of [MRT18] and constructed a generator that fools *unordered* polynomial-width read-once branching programs. Thus, it is motivating to further study product tests, in the hope of gaining more insights into constructing better generators for space-bounded algorithms, and resolving the long-standing open problem of RL vs. L.

In this paper we are interested in understanding the Fourier spectrum of product tests. We first define the *Fourier weight* of a function. For a function $f\colon \{0,1\}^n \to \mathbb{R}$, consider its Fourier expansion $f = \sum_{S \subseteq [n]} \hat{f}_S \chi_S$.

Definition 2 (*d*th level Fourier weight in L_q-norm)**.** *Let $f\colon \{0,1\}^n \to \mathbb{C}_{\le 1}$ be any function. The* dth level Fourier weight of f in L_q-norm *is*

$$W_{q,d}[f] := \sum_{|S|=d} |\hat{f}_S|^q.$$

We denote by $W_{q,\le d}[f]$ the sum $\sum_{\ell=0}^{d} W_{q,\ell}[f]$.

Several papers have studied the Fourier spectrum of different classes of tests. This includes constant-depth circuits [Man95, Tal17], read-once branching programs [RSV13, SVW14, CHRT18], and low-sensitivity functions [GSW16]. More specifically, these papers showed that they have *bounded L_1 Fourier tail*, that is, there exists a positive number b such that for every test f in the class and every positive integer d, we have

$$W_{1,d}[f] \le b^d.$$

One technical contribution of this paper is giving tight upper and lower bounds on the L_1 Fourier tail of product tests.

Theorem 3. *Let $f\colon \{0,1\}^n \to [-1,1]$ be a product test of k functions $f_1,\ldots,f_k$ with input length m. Suppose there is a constant $c > 0$ such that $|\mathbb{E}[f_i]| \le 1 - 2^{-cm}$ for every f_i. For every positive integer d, we have*

$$W_{1,d}[f] \le \bigl(72(\sqrt{c}\cdot m)\bigr)^d.$$

Theorem 3 applies to Boolean functions f_i with outputs $\{0,1\}$ or $\{-1,1\}$. Moreover, the parity function on mk bits can be written as a product test with outputs $\{-1,1\}$, which has $\hat{f}_{[mk]} = 1$. So product tests do not have non-trivial L_2 Fourier tail. (See [Tal17] for a definition.)

We also obtain a different upper bound when the f_i are arbitrary $[-1,1]$-valued functions.

Theorem 4. *Let $f\colon \{0,1\}^n \to [-1,1]$ be a product test of k functions $f_1,\ldots,f_k$ with input length m. Let d be a positive integer. We have*

$$W_{1,d}[f] \le \bigl(85\sqrt{m\ln(4ek)}\bigr)^d.$$

We note that Theorems 3 and 4 are incomparable, as one can take $m = 1$ and $k = n$, or $m = n$ and $k = 1$.

Claim 5. *For all positive integers m and d, there exists a product test $f\colon \{0,1\}^{mk} \to \{0,1\}$ with $k = d\cdot 2^m$ functions of input length m such that*

$$W_{1,d}[f] \ge (m/e^{3/2})^d.$$

This matches the upper bound $W_{1,d}[f] = O(m)^d$ in Theorem 3 up to the constant in the $O(\cdot)$. Moreover, applying Theorem 4 to the product test f in Claim 5 gives $W_{1,d}[f] = O(\sqrt{m\log(2k)})^d = O(m + \sqrt{m\log d})^d$. Therefore, for all integers m and $d \le 2^{O(m)}$, there exists an integer k and a product test f such that the upper bound $W_{1,d}[f] = O(\sqrt{m\log(2k)})^d$ is tight up to the constant in the $O(\cdot)$.

We now discuss some applications of Theorems 3 and 4 in pseudorandomness.

Pseudorandom generators. In recent years, researchers have developed new frameworks to construct pseudorandom generators against different classes of tests. Gopalan, Meka, Reingold, Trevisan and Vadhan [GMR+12] refined a framework introduced by Ajtai and Wigderson [AW89] to construct better generators for the classes of combinatorial rectangles and read-once DNFs. Since then, this framework has been used extensively to construct new PRGs against different classes of tests [TX13, GKM15, GY14, RSV13, SVW14, CSV15, HLV18, HT18, ST18, LV17, CHRT18, FK18, MRT18, DHH18]. Recently, a beautiful work by Chattopadhyay, Hatami, Hosseini and Lovett [CHHL18] developed a new framework of constructing PRGs against any classes of functions that are closed under restriction and have bounded L_1 Fourier tail. Thus, applying their result to Theorems 3 and 4, we can immediately obtain a non-trivial PRG for product tests. However, using the recent result of Forbes and Kelley [FK18] and exploiting the structure of product tests, we use the Ajtai–Wigderson framework to construct PRGs with much better seed length than using [CHHL18] as a blackbox.

Theorem 6. *There exists an explicit generator $G\colon \{0,1\}^\ell \to \{0,1\}^n$ that fools the XOR of any k Boolean functions on disjoint inputs of length $\leq m$ with error ε and seed length $O(m + \log(n/\varepsilon))(\log m + \log\log(n/\varepsilon))^2 = \tilde{O}(m + \log(n/\varepsilon))$.*

Here $\tilde{O}(1)$ hides polynomial factors in $\log m$, $\log\log k$, $\log\log n$ and $\log\log(1/\varepsilon)$. When $mk = n$ or $\varepsilon = n^{-\Omega(1)}$, the generator in Theorem 6 has seed length $\tilde{O}(m + \log(k/\varepsilon))$, which is optimal up to $\tilde{O}(1)$ factors.

We now compare Theorem 6 with previous works. Using a completely different analysis, Lee and Viola [LV17] obtained a generator with seed length $\tilde{O}((m + \log k))\log(1/\varepsilon)$. When $m = O(\log n)$ and $k = 1/\varepsilon = n^{\Omega(1)}$, this is $\tilde{O}(\log^2 n)$, whereas the generator in Theorem 6 has seed length $\tilde{O}(\log n)$. When each function f_i is computable by a read-once width-w branching program on m bits, Meka, Reingold and Tal [MRT18] obtained a PRG with seed length $O(\log(n/\varepsilon))(\log m + \log\log(n/\varepsilon))^{2w+2}$. When $m = O(\log(n/\varepsilon))$, Theorem 6 improves on their generator on the lower order terms. As a result, we obtain a PRG for *read-once $\mathbb{F}_2$-polynomials*, which are a sum of monomials on disjoint variables over $\mathbb{F}_2$, with seed length $O(\log n/\varepsilon)(\log\log(n/\varepsilon))^2$. This also improves on the seed length of their PRG for read-once polynomials in the lower order terms by a factor of $(\log\log(n/\varepsilon))^4$.

Our generator in Theorem 6 also works for the AND of the functions f_i, corresponding to the class of *unordered* combinatorial rectangles. In fact, we have the following more general corollary.

Corollary 7. *There exists an explicit pseudorandom generator $G\colon \{0,1\}^\ell \to \{0,1\}^n$ with seed length $\tilde{O}(m+\log(n/\varepsilon))$ such that the following holds. Let $f_1, \ldots, f_k\colon \{0,1\}^{I_i} \to \{0,1\}$ be k Boolean functions where the subsets $I_i \subseteq [n]$ are pairwise disjoint and have size at most m. Let $g\colon \{0,1\}^k \to \mathbb{C}_{\leq 1}$ be any function and write g in its Fourier expansion $g = \sum_{S \subseteq [k]} \hat{g}_S \chi_S$. Then G fools $g(f_1, \ldots, f_k)$ with error $L_1[g] \cdot \varepsilon$, where $L_1[g] := \sum_{S \neq \emptyset} |\hat{g}_S|$.*

Proof. Let G be the generator in Theorem 6. Note that $\chi_S(f_1(x_{I_1}), \ldots, f_k(x_{I_k}))$ is a product test with outputs $\{-1,1\}$. So by Theorem 6 we have

$$\begin{aligned}
&\left|\mathbb{E}[g(f_1(U_{I_1}),\ldots,f_k(U_{I_k})) - \mathbb{E}[g(f_1(G_{I_1}),\ldots,f_k(G_{I_k})]\right| \\
&\leq \sum_S |\hat{g}_S| \left|\mathbb{E}[\chi_S(f_1(U_{I_1}),\ldots,f_k(U_{I_k}))] - \mathbb{E}[\chi_S(f_1(G_{I_1}),\ldots,f_k(G_{I_k})]\right| \\
&\leq L_1[g]\cdot\varepsilon. \qquad \square
\end{aligned}$$

Note that the AND function has $L_1[\mathrm{AND}] \leq 1$, and so the generator in Corollary 7 fools unordered combinatorial rectangles. Previous generators for unordered combinatorial rectangles use almost-bounded independence or small-bias distributions, and have seed length $O(\log(n/\varepsilon))(1/\varepsilon)$ [CRS00, DETT10].

When the functions f_i in the product tests have outputs $[-1,1]$, we also obtain the following generator.

Theorem 8. *There exists an explicit generator $G\colon \{0,1\}^\ell \to \{0,1\}^n$ that fools any product test with k functions of input length m with error ε and seed length $O(m+\log(k/\varepsilon))\log(k/\varepsilon)$ $(\log m + \log\log n) = \tilde{O}(m+\log(k/\varepsilon))\log(k/\varepsilon)$.*

When $m = o(\log n)$ and $k = 1/\varepsilon = 2^{o(\sqrt{\log n})}$, Theorem 8 gives a better seed length than Theorem 6. Thus the generator in Theorem 8 remains interesting for $f_i \in \{-1,1\}$ when a product test f depends on very few variables and the error ε is not so small.

Previous best generator [LV17] has an extra $\tilde{O}(\log(1/\varepsilon))$ in the seed length. However, the generator in [LV17] works even when the f_i have range $\mathbb{C}_{\leq 1}$, which implies generators for several variants of product tests such as generalized halfspaces and combinatorial shapes. (See [GKM15] for the reductions.)

Finally, when the subsets I_i of a product test are fixed and known in advanced, Gopalan, Kane and Meka [GKM15] constructed a PRG of the same seed length as Theorem 6, but again their PRG works more generally for the range of $\mathbb{C}_{\leq 1}$ instead of $\{-1,1\}$.

$\mathbb{F}_2$-polynomials. Chattopadhyay, Hatami, Lovett and Tal [CHLT19] recently constructed a pseudorandom generator for any class of functions that are closed under restriction, provided there is an upper bound on the second level Fourier weight of the functions in L_1-norm. They conjectured that every n-variate $\mathbb{F}_2$-polynomial f of degree d satisfies the bound $W_{1,2}[f] = O(d^2)$. In particular, a bound of $n^{1/2-o(1)}$ would already imply a generator for polynomials of degree $d = \Omega(\log n)$, a major breakthrough in complexity theory. Theorem 4 shows that their conjecture is true for the special case of *read-once* polynomials. In fact, it shows that $W_{1,t}[f] = O(d^t)$ for every positive integer t. Previous bound for read-once polynomials gives $W_{1,t}[f] = O(\log^4 n)^t$ [CHRT18].

The coin problem. Let $X_{n,\varepsilon} = (X_1,\ldots,X_n)$ be the distribution over n bits, where the variables X_i are independent and each X_i equals 1 with probability $(1-\varepsilon)/2$ and 0 otherwise. The *ε-coin problem* asks whether a given function f can distinguish between the distributions $X_{n,\varepsilon}$ and $X_{n,0}$ with advantage $1/3$.

This central problem has wide range of applications in computational complexity and has been studied extensively for different restricted classes of tests, including bounded-depth circuits [Ajt83, Val84, ABO84, Ama09, Vio09, SV10, Aar10, Vio14, CGR14], space-bounded algorithms [BV10, Ste13, CGR14], bounded-depth circuits with parity gates [SV10, KS18, RS17, LSS+18], $\mathbb{F}_2$-polynomials [LSS+18, CHLT19] and product tests [LV18].

It is known that if a function f has bounded L_1 Fourier tail, then it implies a lower bound on the smallest ε^* of ε that f can solve the ε-coin problem.

Fact 9. *Let $f\colon \{0,1\}^n \to \mathbb{C}_{\le 1}$ be any function. If for every integer $d \in \{0,\dots,n\}$ we have $W_{1,d}[f] \le b^d$, then f solves the ε-coin problem with advantage at most $2b\varepsilon$.*

Proof. We may assume $b\varepsilon \le 1/2$, otherwise the result is trivial. Observe that we have $\mathbb{E}[\chi_S(X_{n,\varepsilon})] = \varepsilon^{|S|}$ for every subset $S \subseteq [n]$. Thus,

$$\left|\mathbb{E}[f(X_{n,\varepsilon})] - \mathbb{E}[f(X_{n,0})]\right| = \left|\sum_{S\neq\emptyset} \hat{f}_S\,\mathbb{E}[X_{n,\varepsilon}]\right|$$
$$\le \sum_{d=1}^{n}\sum_{|S|=d} |\hat{f}_S|\cdot\varepsilon^d = \sum_{d=1}^{n}(b\varepsilon)^d \le b\varepsilon\cdot\sum_{d=1}^{n} 2^{-(d-1)} \le 2b\varepsilon. \qquad \square$$

Lee and Viola [LV18] showed that product tests with range $[-1,1]$ can solve the ε-coin problem with $\varepsilon^* = \Theta(1/\sqrt{m\log k})$. Hence, Fact 9 implies that Theorem 4 recovers their lower bound. Moreover, their upper bound implies that the dependence on m and k in Theorem 4 is tight up to constant factors when d is constant. Claim 5 complements this by showing that the dependence on d in Theorem 4 is also tight for some choice of k.

The work [LV18] also shows that when the range of the functions f_i is $\mathbb{C}_{\le 1}$, the right answer for ε^* is $\Theta(1/\sqrt{mk})$. Therefore, one cannot obtain for a better tail bound than the trivial bound of $(\sqrt{mk})^d$ when the range is $\mathbb{C}_{\le 1}$.

1.1 Techniques

We now explain how to obtain Theorems 3 and 4 and our pseudorandom generators for product tests (Theorems 6 and 8).

1.1.1 Fourier spectrum of product tests

The high-level idea of proving Theorems 3 and 4 is inspired from [LV18]. For intuition, let us first assume that the functions f_i have outputs $\{0,1\}$ and are all equal to f_1 (but defined on disjoint inputs). It will also be useful to think of the number of functions k being much larger than input length m of each function. We first explain how to bound above $W_{1,1}[f]$. (Recall in Definition 2 we defined $W_{q,d}[f]$ of a function f to be $\sum_{|S|=d}|\hat{f}_S|^q$.)

Bounding $W_{1,1}[f]$. Since the functions f_i of a product test f are defined on disjoint inputs, each Fourier coefficient of f is a product of the coefficients of the f_i, and so each weight-1 coefficent of f is a product of $k-1$ weight-0 and 1 weight-1 coefficients of the f_i. From this, we can see that $W_{1,1}[f]$ is equal to

$$\binom{k}{1}\cdot W_{1,1}[f_1]\cdot W_{1,0}[f_1]^{k-1} = k\cdot W_{1,1}[f_1]\cdot\mathbb{E}[f_1]^{k-1}. \tag{1}$$

Because of the term $\mathbb{E}[f_1]^{k-1}$, to maximize $W_{1,1}[f]$ it is natural to consider taking f_1 to be a function with expectation $\mathbb{E}[f_1]$ as close to 1 as possible, i.e. the OR function. In such case, one would hope for a better bound on $W_{1,1}[f_1]$. Indeed, Chang's inequality [Cha02] (see also [IMR14] for a simple proof) says that for a $[0,1]$-valued function g with expectation $\alpha \leq 1/2$, we have

$$W_{2,1}[g] \leq 2\alpha^2 \ln(1/\alpha).$$

(The condition $\alpha \leq 1/2$ is without loss of generality as one can instead consider $1 - g$.) It follows by a simple application of the Cauchy–Schwarz inequality that $W_{1,1}[g] \leq O(\sqrt{n}) \cdot \alpha\sqrt{\ln(1/\alpha)}$ (see Fact 12 below for a proof). Moreover, when the functions f_i are Boolean, we have $2^{-m} \leq \mathbb{E}[f_i] \leq 1 - 2^{-m}$, and so $\sqrt{\ln(1/\alpha)} \leq \sqrt{m}$. Plugging these bounds into Equation (1), we obtain a bound of $O(m) \cdot k(1 - \mathbb{E}[f_1]) \mathbb{E}[f_1]^{k-1}$. So indeed $\mathbb{E}[f_1]$ should be roughly $1 - 1/k$ in order to maximize $W_{1,1}[f]$, giving an upper bound of $O(m)$. For the case where the f_i can be different, a simple convexity argument shows that $W_{1,1}[f]$ is maximized when the functions f_i have the same expectation.

Bounding $W_{1,d}[f]$ for $d > 1$. To extend this argument to $d > 1$, one has to generalize Chang's inequality to bound above $W_{2,d}[g]$ for $d > 1$. The case $d = 2$ was already proved by Talagrand [Tal96]. Following Talagrand's argument in [Tal96] and inspired by the work of Keller and Kindler [KK13], which proved a similar bound in terms of a different measure than $\mathbb{E}[g]$, we prove the following bound on $W_{2,d}[g]$ in terms of its expectation.

Lemma 10. *Let $g\colon \{0,1\}^n \to [0,1]$ be any function. For every positive integer d, we have*

$$W_{2,d}[g] \leq 4\,\mathbb{E}[g]^2 \left(2e \ln(e/\mathbb{E}[g]^{1/d})\right)^d.$$

We note that the exponent $1/d$ of $\mathbb{E}[g]$ either did not appear in previous upper bounds (mentioned without proof in [IMR14]), or only holds for restricted values of d [O'D14]. This exponent is not important for proving Theorem 3 , but will be crucial in the proof of Theorem 4, which we will explain later on.

For $d > 1$, the expression for $W_{1,d}[f]$ becomes much more complicated than $W_{1,1}[f]$, as it involves $W_{1,z}[f_1]$ for different values of $z \in [m]$. So one has to formulate the expression of $W_{1,d}[f]$ carefully. (See Lemma 13.) Once we have obtained the right expression for $W_{1,d}[f]$, the proof of Theorem 3 follows the outline above by replacing Chang's inequality with Lemma 10. One can then handle functions f_i with outputs $\{-1,1\}$ by considering the translation $f_i \mapsto (1 - f_i)/2$, which only changes each $W_{1,d}[f_i]$ (for $d > 0$) by a factor of 2. We remark that Theorem 3 is sufficient for constructing the generator in Theorem 6.

Handling $[-1,1]$-valued f_i. Extending this argument to proving Theorem 4 poses several challenges. Following the outline above, after plugging in Lemma 10, we would like to show that $\mathbb{E}[f_1]$ should be roughly $1 - 1/k$ to maximize $W_{1,d}[f]$. However, it is no longer clear why this is the case even assuming the maximum is attained by functions f_i with the same expectation, as we now do not have the bound $\sqrt{\ln(1/\alpha)} \leq \sqrt{m}$, and so it cannot be used to simplify the expression of $W_{1,d}[f]$ as before. In fact, the above assumption is simply false if we plug in the upper bound in Lemma 10 with the exponent $1/d$ omitted to the $W_{1,z_i}[f_i]$.

Using Lemma 10 and the symmetry of the expression for $W_{1,d}[f]$, we reduce the problem of bounding above $W_{1,d}[f]$ with different f_i to bounding the same quantity but with the additional assumption that the f_i have the same expectation $\mathbb{E}[f_1]$. This uses Schur-convexity (see Section 2 for its definition). Then by another convexity argument we show that the maximum is attained when $\mathbb{E}[f_1]$ is roughly equal to $1 - d/k$. Both of these arguments critically rely on the aforementioned exponent of $1/d$ in Lemma 10.

1.1.2 Pseudorandom generators

We now discuss how to use Theorems 3 and 4 to construct our pseudorandom generators for product tests. Our construction follows the Ajtai–Wigderson framework [AW89] that was recently revived and refined by Gopalan, Meka, Reingold, Trevisan and Vadhan [GMR+12].

The high-level idea of this framework involves two steps. For the first step, we show that *derandomized bounded independence plus noise* fools f. More precisely, we will show that if we start with a small-bias or almost-bounded independent distribution D ("bounded independence"), and select roughly half of D's positions T pseudorandomly and set them to uniform U ("plus noise"), then this distribution, denoted by $D + T \wedge U$, fools product tests.

Forbes and Kelley [FK18] recently improved the analysis in [HLV18] and implicitly showed that δ-almost d-wise independent plus noise fools product tests, where $d = O(m + \log(k/\varepsilon))$ and $\delta = n^{-\Omega(d)}$. Using Theorem 4, we improved the dependence on δ to $(m \ln k)^{-\Omega(d)}$ and obtain the following theorem.

Theorem 11. *Let $f\colon \{0,1\}^n \to [-1,1]$ be a product test with k functions of input length m. Let d be a positive integer. Let D and T be two independent δ-almost d-wise independent distributions over $\{0,1\}^n$, and U be the uniform distribution over $\{0,1\}^n$. Then*

$$\left|\mathbb{E}[f(D + T \wedge U)] - \mathbb{E}[f(U)]\right| \le k \cdot \left(\sqrt{\delta} \cdot (170 \cdot \sqrt{m \ln(ek)})^d + 2^{-(d-m)/2}\right),$$

where "+" and "∧" are bit-wise XOR and AND respectively.

The second step of the Ajtai–Wigderson framework builds a pseudorandom generator by applying the first step (Theorem 11) recursively. Let $f\colon \{0,1\}^n \to \{0,1\}$ be a product test with k functions of input length m. As product tests are closed under restrictions (and shifts), after applying Theorem 11 to f and fixing D and T in the theorem, the function $f_{D,T}\colon \{0,1\}^T \to \{0,1\}$ defined by $f_{D,T}(y) := f(D + T \wedge y)$ is also a product test. Thus one can apply Theorem 11 to $f_{D,T}$ again and repeat the argument recursively. We will use different progress measures to bound above the number of recursion steps in our constructions. We first describe the recursion in Theorem 8 as it is simpler.

Fooling $[-1,1]$-valued product tests. Here our progress measure is the maximum input length m of the functions f_i. We show that after $O(\log(k/\varepsilon))$ steps of the recursion, the functions f_i of the restricted product test have their input length halved with high probability. Therefore, repeating above for $O(\log m)$ steps, the product test is restricted to a constant function. This simple recursion gives our second PRG (Theorem 8).

Fooling Boolean-valued product tests. Our construction of the first generator (Theorem 6) is more complicated and uses two progress measures. The first one is again the maximum input length m of the functions f_i, and the second is the number k of the functions f_i. We reduce the number of recursion steps from $O(\log(k/\varepsilon))\log m$ to $O(\log m)$. This requires a more delicate construction and analysis that are similar to the recent work of Meka, Reingold and Tal [MRT18], which constructed a pseudorandom generator against XOR of disjoint constant-width read-once branching programs. There are two main ideas in their construction. First, they ensure $k \leq 16^m$ in each step of the recursion, by constructing another PRG to fool the test f for the case $k \geq 16^m$. We will also use this PRG in our construction. Next, throughout the recursion they allow one "bad" function f_i of the product test f to have a longer input length than m, but not longer than $O(\log(n/\varepsilon))$. Using these two ideas, they show that whenever $m \geq \log\log n$ during the recursion, then after $O(1)$ steps of the recursion all but the "bad" f_i have their input length restricted by a half, while the "bad" f_i always has length $O(\log(n/\varepsilon))$. This allows us to repeat $O(\log m)$ steps until we are left with a product test of $k' \leq \text{polylog}(n)$ functions, where all but one of the f_i have input length at most $m' = O(\log\log n)$.

Now we switch our progress measure to the number of functions. This part is different from [MRT18], in which their construction relies on the fact that the f_i are computable by read-once branching programs. Here because our functions f_i are arbitrary, by grouping c functions as one, we can instead think of the parameters k' and m' in the product test as $k'' = k'/c$ and $m'' = cm'$, respectively. Choosing c to be $O(\log n/\log\log n)$, we have $m'' = O(\log n)$ and so we can repeat the previous argument again. Because each time k' is reduced by a factor of c, after repeating this for $O(1)$ steps, we are left with a product test defined on $O(\log n)$ bits, which can be fooled using a small-bias distribution. This gives our first generator (Theorem 6).

Organization In Section 2 we prove Theorems 3 and 4. In Section 3 we construct our pseudorandom generators for product tests, proving Theorems 6 and 8. In Section 4 we prove Lemma 10, which is used in the proof of Theorem 4.

2 Fourier spectrum of product tests

In this section we prove Theorems 3 and 4. We first restate the theorems.

Theorem 3. *Let $f\colon \{0,1\}^n \to [-1,1]$ be a product test of k functions $f_1,\dots,f_k$ with input length m. Suppose there is a constant $c > 0$ such that $|\mathbb{E}[f_i]| \leq 1 - 2^{-cm}$ for every f_i. For every positive integer d, we have*

$$W_{1,d}[f] \leq \left(72(\sqrt{c}\cdot m)\right)^d.$$

Theorem 4. *Let $f\colon \{0,1\}^n \to [-1,1]$ be a product test of k functions $f_1,\dots,f_k$ with input length m. Let d be a positive integer. We have*

$$W_{1,d}[f] \leq \left(85\sqrt{m\ln(4ek)}\right)^d.$$

Both theorems rely on the following lemma which gives an upper bound on $W_{2,d}[g]$ in terms of the expectation of a $[0,1]$-valued function g. The case $d=1$ is known as Chang's inequality [Cha02]. (See also [IMR14] for a simple proof.) This was then generalized by Talagrand to $d=2$ [Tal96]. Using a similar argument to [Tal96], we extend this to $d>2$.

Lemma 10. *Let $g\colon \{0,1\}^n \to [0,1]$ be any function. For every positive integer d, we have*

$$W_{2,d}[g] \le 4\,\mathbb{E}[g]^2\big(2e\ln(e/\mathbb{E}[g]^{1/d})\big)^d.$$

We defer its proof to Section 4. We remark that a similar upper bound was proved by Keller and Kindler [KK13]. However, the upper bound in [KK13] was proved in terms of $\sum_{i=1}^n I_i[g]^2$, where $I_i[g]$ is the influence of the ith coordinate on g, instead of $\mathbb{E}[g]$. A similar upper bound in terms of $\mathbb{E}[g]$ can be found in [O'D14] under the extra condition $d \le 2\ln(1/\mathbb{E}[g])$.

We will also use the following well-known fact that bounds above $W_{1,d}[f]$ in terms of $W_{2,d}[f]$.

Fact 12. *Let $f\colon \{0,1\}^n \to \mathbb{R}$ be any function. We have $W_{1,d}[f] \le n^{d/2}\sqrt{W_{2,d}[f]}$.*

Proof. By the Cauchy–Schwarz inequality,

$$W_{1,d}[f] = \sum_{|S|=d} |\hat{f}_S| \le \sqrt{\binom{n}{d}\sum_{|S|=d}\hat{f}_S^2} \le n^{d/2}\sqrt{W_{2,d}[f]}. \qquad \square$$

Lemma 13. *Let $f\colon \{0,1\}^n \to [-1,1]$ be a product test of k functions $f_1,\dots,f_k$ with input length m, and $\alpha_i := (1-\mathbb{E}[f_i])/2$ for every $i \in [k]$. Let d be a positive integer. We have*

$$W_{1,d}[f] \le \big(\sqrt{32e^3m}\big)^d g(\alpha_1,\dots,\alpha_k),$$

where the function $g\colon (0,1]^k \to \mathbb{R}$ is defined by

$$g(\alpha_1,\dots,\alpha_k) := e^{-2\sum_{i=1}^k \alpha_i}\sum_{\ell=1}^{d}\sum_{\substack{S\subseteq[k]\\|S|=\ell}}\sum_{\substack{z\in[m]^S\\ \sum_i z_i=d}}\prod_{i\in S}\Big(\alpha_i\big(\ln\big(e/\alpha_i^{1/z_i}\big)\big)^{z_i/2}\Big).$$

Proof. For notational simplicity, we will use $W_d[f]$ to denote $W_{1,d}[f]$. Write $f=\prod_{i=1}^k f_i$. Without loss of generality we will assume each function f_i is non-constant. Since f_i and $-f_i$ have the same weight $W_d[f_i]$, we will further assume $\mathbb{E}[f_i]\in[0,1)$. Note that for a subset $S = S_1\times\cdots\times S_k \subseteq (\{0,1\}^m)^k$, we have $\hat{f}_S = \prod_{i=1}^k \hat{f}_{i S_i}$. So,

$$W_d[f] = \sum_{\substack{z\in\{0,\dots,m\}^k\\ \sum_i z_i=d}}\prod_{i=1}^k W_{z_i}[f_i] = \sum_{\ell=1}^{d}\sum_{\substack{S\subseteq[k]\\|S|=\ell}}\sum_{\substack{z\in[m]^S\\ \sum_i z_i=d}}\Big(\prod_{i\in S}W_{z_i}[f_i]\cdot\prod_{i\notin S}W_0[f_i]\Big).$$

Since $x = 1-(1-x) \le e^{-(1-x)}$ for every $x\in\mathbb{R}$, for every subset $S\subseteq[k]$ of size at most d, we have

$$\prod_{i\notin S}W_{z_i}[f_i] \le e^{-\sum_{i\notin S}(1-W_{z_i}[f_i])} \le e^{-\sum_{i\notin S}(1-W_{z_i}[f_i])}\cdot e^{\sum_{i\in S}W_{z_i}[f_i]} \le e^d\cdot e^{-\sum_{i=1}^k(1-W_{z_i}[f_i])}.$$

Hence,

$$W_d[f] = \sum_{\ell=1}^{d} \sum_{\substack{S\subseteq[k]\\|S|=\ell}} \sum_{\substack{z\in[m]^S\\ \sum_i z_i=d}} \Big(\prod_{i\in S} W_{z_i}[f_i]\cdot\prod_{i\notin S} W_0[f_i]\Big)$$
$$\le e^d\cdot e^{-\sum_{i=1}^k(1-W_0[f_i])}\sum_{\ell=1}^{d} \sum_{\substack{S\subseteq[k]\\|S|=\ell}} \sum_{\substack{z\in[m]^S\\ \sum_i z_i=d}} \prod_{i\in S} W_{z_i}[f_i]. \qquad (2)$$

Define $f'_i := (1-f_i)/2 \in [0,1]$. Let $\alpha_i := \mathbb{E}[f'_i] = (1-\mathbb{E}[f_i])/2 \in (0,1/2]$. Applying Lemma 10 and Fact 12 to the functions f'_i, we have for every subset $S\subseteq[k]$ of size at most d,

$$\sum_{\substack{z\in[m]^S\\ \sum_i z_i=d}} \prod_{i\in S} W_{z_i}[f'_i] \le \sum_{\substack{z\in[m]^S\\ \sum_i z_i=d}} \prod_{i\in S}\Big(2m^{z_i/2}\alpha_i\big(2e\ln(e/\alpha_i^{1/z_i})\big)^{z_i/2}\Big)$$
$$\le (\sqrt{8em})^d \sum_{\substack{z\in[m]^S\\ \sum_i z_i=d}} \prod_{i\in S}\Big(\alpha_i\big(\ln(e/\alpha_i^{1/z_i})\big)^{z_i/2}\Big).$$

Note that for every integer $d\ge 1$, we have $W_d[f_i] = 2W_d[f'_i]$. Plugging the bound above into Equation (2), we have

$$W_d[f] \le (2e)^d\cdot e^{-2\sum_{i=1}^k \alpha_i}\sum_{\ell=1}^{d} \sum_{\substack{S\subseteq[k]\\|S|=\ell}} \sum_{\substack{z\in[m]^S\\ \sum_i z_i=d}} \prod_{i\in S} W_{z_i}[f'_i] \le \big(\sqrt{32e^3m}\big)^d g(\alpha_1,\dots,\alpha_k),$$

where the function $g\colon (0,1]^k\to\mathbb{R}$ is defined by

$$g(\alpha_1,\dots,\alpha_k) := e^{-2\sum_{i=1}^k\alpha_i}\sum_{\ell=1}^{d} \sum_{\substack{S\subseteq[k]\\|S|=\ell}} \sum_{\substack{z\in[m]^S\\ \sum_i z_i=d}} \prod_{i\in S}\Big(\alpha_i\big(\ln(e/\alpha_i^{1/z_i})\big)^{z_i/2}\Big). \qquad \square$$

We now prove Theorems 3 and 4. For every $(\alpha_1,\dots,\alpha_k)\in(0,1]^k$, let $\alpha := \sum_{i=1}^k \alpha_i/k \in (0,1]$. We note that the upper bound in Theorem 3 is sufficient to prove Theorem 6.

Proof of Theorem 3. We will bound above $g(\alpha_1,\dots,\alpha_k)$ in Lemma 13. Recall that $\alpha_i = (1-\mathbb{E}[f_i])/2$. Since $|\mathbb{E}[f_i]|\le 1-2^{-cm}$, we have $\alpha_i \ge 2^{-(cm+1)}$, and so $\ln(1/\alpha_i)\le cm+1$. For every subset $S\subseteq[k]$, the set $\{z\in[m]^S : \sum_i z_i = d\}$ has size at most $\binom{d-1}{|S|-1}\le 2^d$. Hence,

$$\sum_{\substack{z\in[m]^S\\ \sum_i z_i=d}} \prod_{i\in S}\big(\ln(1/\alpha_i)\big)^{z_i/2} \le 2^d(cm+1)^{d/2}.$$

By Maclaurin's inequality (cf. [Ste04, Chapter 12]), we have

$$\sum_{\substack{S\subseteq[k]\\|S|=\ell}} \prod_{i\in S}\alpha_i \le (e/\ell)^\ell\Big(\sum_{i=1}^k \alpha_i\Big)^\ell = (e/\ell)^\ell(k\alpha)^\ell.$$

Because the function $x \mapsto e^{-2x}x^\ell$ is maximized when $x = \ell/2$, it follows that

$$\sum_{\ell=1}^{d} e^{-2k\alpha} \sum_{\substack{S\subseteq[k]\\|S|=\ell}} \prod_{i\in S}\alpha_i \le \sum_{\ell=1}^{d} e^{-2k\alpha}(e/\ell)^\ell(k\alpha)^\ell \le \sum_{\ell=1}^{d} e^{-\ell}(e/\ell)^\ell(\ell/2)^\ell = \sum_{\ell=1}^{d} 2^{-\ell} \le 1.$$

Therefore,

$$\begin{aligned} g(\alpha_1,\dots,\alpha_k) &= e^{-2\sum_{i=1}^k \alpha_i} \sum_{\ell=1}^{d} \sum_{\substack{S\subseteq[k]\\|S|=\ell}} \sum_{\substack{z\in[m]^S\\ \sum_i z_i = d}} \prod_{i\in S} \Big(\alpha_i \big(\ln(1/\alpha_i^{1/z_i})\big)^{z_i/2}\Big) \\ &\le 2^d(cm+1)^{d/2} \sum_{\ell=1}^{d} e^{-2k\alpha} \sum_{\substack{S\subseteq[k]\\|S|=\ell}} \prod_{i\in S}\alpha_i \\ &\le 2^d(cm+1)^{d/2}. \end{aligned}$$

Plugging this bound into Lemma 13, we have

$$W_{1,d}[f] \le \big(\sqrt{32e^3m}\big)^d \cdot \big(\sqrt{4(cm+1)}\big)^d \le \big(72(\sqrt{c}\cdot m)\big)^d. \qquad \square$$

We now prove Theorem 4. Recall that we let $\alpha := \sum_{i=1}^k \alpha_i/k \in (0,1]$ for every $(\alpha_1,\dots,\alpha_k) \in (0,1]^k$. We will show that the maximum of the function g defined in Lemma 13 is attained at the diagonal $(\alpha,\dots,\alpha)$. We state the claim now and defer the proof to the next section.

Claim 14. *Let g be the function defined in Lemma 13. For every $(\alpha_1,\dots,\alpha_k) \in (0,1]^k$, we have $g(\alpha_1,\dots,\alpha_k) \le g(\alpha,\dots,\alpha)$.*

Proof of Theorem 4. We first apply Claim 14 and obtain

$$g(\alpha_1,\dots,\alpha_k) \le g(\alpha,\dots,\alpha) = e^{-2k\alpha} \sum_{\ell=1}^{d} \sum_{\substack{S\subseteq[k]\\|S|=\ell}} \alpha^\ell \sum_{\substack{z\in[m]^S\\ \sum_i z_i = d}} \prod_{i\in S} \big(\ln\big(e/\alpha^{1/z_i}\big)\big)^{z_i/2}.$$

We next give an upper bound on $g(\alpha,\dots,\alpha)$ that has no dependence on the numbers z_i. By the weighted AM-GM inequality, for every subset $S \subseteq [k]$ of size ℓ and numbers z_i such that $\sum_{i\in S} z_i = d$,

$$\begin{aligned} \prod_{i\in S} \big(\ln\big(e/\alpha^{1/z_i}\big)\big)^{z_i/2} &\le \Big(\sum_{i\in S} \frac{z_i \ln\big(e/\alpha^{1/z_i}\big)}{d}\Big)^{d/2} \\ &= \Big(\frac{1}{d}\sum_{i\in S} z_i\Big(1 + \frac{1}{z_i}\ln(1/\alpha)\Big)\Big)^{d/2} \\ &= \Big(1 + \frac{\ell}{d}\ln(1/\alpha)\Big)^{d/2} \\ &= \big(\ln\big(e/\alpha^{\ell/d}\big)\big)^{d/2}. \end{aligned}$$

For every subset $S \subseteq [k]$, the set $\{z \in [m]^S : \sum_i z_i = d\}$ has size at most $\binom{d-1}{|S|-1} \leq 2^d$. Thus,

$$\begin{aligned}
g(\alpha, \dots, \alpha) &\leq e^{-2k\alpha} \sum_{\ell=1}^{d} \sum_{\substack{S \subseteq [k] \\ |S| = \ell}} \alpha^\ell \sum_{\substack{z \in [m]^S \\ \sum_i z_i = d}} \big(\ln\big(e/\alpha^{\ell/d}\big)\big)^{d/2} \\
&\leq 2^d \sum_{\ell=1}^{d} e^{-2k\alpha} \sum_{\substack{S \subseteq [k] \\ |S| = \ell}} \alpha^\ell \big(\ln\big(e/\alpha^{\ell/d}\big)\big)^{d/2} \\
&\leq 2^d \sum_{\ell=1}^{d} e^{-2k\alpha} \Big(\frac{ek\alpha}{\ell}\Big)^\ell \big(\ln\big(e/\alpha^{\ell/d}\big)\big)^{d/2}. \qquad (3)
\end{aligned}$$

For every $\ell \in [k]$, define $g_\ell \colon (0, 1] \to \mathbb{R}$ to be

$$g_\ell(x) := e^{-2kx} \Big(\frac{ekx}{\ell}\Big)^\ell \big(\ln\big(e/x^{\ell/d}\big)\big)^{d/2}.$$

We now bound above the maximum of g_ℓ over $x \in (0, 1]$. One can verify easily that the derivative of g is

$$g'_\ell(x) = \frac{g_\ell(x)}{2x \ln\big(e/x^{\ell/d}\big)} \big(\ln(1/x^{2\ell/d})(\ell - 2kx) + (\ell - 4kx)\big).$$

Observe that when $x \leq \ell/4k$, then $g'_\ell(x) \geq \frac{g_\ell(x)}{4x \ln(e/x^{\ell/d})}\big(\ell \ln(1/x^{2\ell/d})\big) \geq 0$. Likewise, when $x \geq \ell/2k$, then $g'_\ell(x) \leq \frac{g_\ell(x)}{2x \ln(e/x^{\ell/d})}(-\ell) \leq 0$. Also, we have $g_\ell(0) = 0$. Hence, $g_\ell(x) \leq g_\ell(\beta_\ell \ell/4k)$ for some $\beta_\ell \in [1, 2]$, which is at most

$$e^{-\ell/2} \cdot (e/2)^\ell \cdot \Big(\ln\big(e(4k/\ell)^{\ell/d}\big)\Big)^{d/2}.$$

(In the case when $\ell/4k \geq 1$, we have $g_\ell(x) \leq g_\ell(1) \leq e^{-2k}(ek/\ell)^\ell$.) Therefore, plugging this back into Equation (3),

$$\begin{aligned}
g(\alpha, \dots, \alpha) \leq 2^d \sum_{\ell=1}^{d} g_\ell(\alpha) \leq 2^d \sum_{\ell=1}^{d} g_\ell(\beta_\ell \ell/4k) &\leq 2^d \sum_{\ell=1}^{d} e^{-\ell/2} \cdot (e/2)^\ell \cdot \Big(\ln\big(e(4k/\ell)^{\ell/d}\big)\Big)^{d/2} \\
&\leq 2^d \big(e \ln(4ek)\big)^{d/2} \sum_{\ell=1}^{d} 2^{-\ell} \\
&\leq \big(\sqrt{4e \ln(4ek)}\big)^d.
\end{aligned}$$

Putting this back into the bound in Lemma 13, we conclude that

$$W_{1,d}[f] \leq \big(84\sqrt{m \ln(4ek)}\big)^d,$$

proving the theorem. □

2.1 Schur-concavity of g

We prove Claim 14 in this section. First recall that the function $g\colon (0,1]^k \to \mathbb{R}$ is defined as

$$g(\alpha_1,\ldots,\alpha_k) := \sum_{\ell=1}^{d} \sum_{\substack{S\subseteq[k]\\|S|=\ell}} \sum_{\substack{z\in[m]^S\\ \sum_i z_i=d}} \prod_{i\in S} \phi_{z_i}(\alpha_i),$$

where for every positive integer z, the function $\phi_z\colon (0,1] \to \mathbb{R}$ is defined by

$$\phi_z(x) = x\ln(e/x^{1/z})^{z/2}.$$

The proof of Claim 14 follows from showing that g is *Schur-concave*. Before defining it, we first recall the concept of majorization. Let $x,y \in \mathbb{R}^k$ be two vectors. We say that y *majorizes* x, denoted by $x \prec y$, if for every $j \in [k]$ we have

$$\sum_{i=1}^{j} x_{(i)} \le \sum_{i=1}^{j} y_{(i)},$$

and $\sum_{i=1}^{k}(x_i - y_i) = 0$, where $x_{(i)}$ and $y_{(i)}$ are the ith largest coordinates in x and y respectively.

A function $f\colon D \to \mathbb{R}$ where $D \subseteq \mathbb{R}^k$ is *Schur-concave* if whenever $x \prec y$ we have $f(x) \ge f(y)$. We will show that g is Schur-concave using the Schur–Ostrowski criterion.

Theorem 15 (Schur–Ostrowski criterion (Theorem 12.25 in [PPT92]))**.** *Let $f\colon D \to \mathbb{R}$ be a function where $D \subseteq \mathbb{R}^k$ is permutation-invariant, and assume that the first partial derivatives of f exist in D. Then f is Schur-concave in D if and only if*

$$(x_j - x_i)\left(\frac{\partial f}{\partial x_i} - \frac{\partial f}{\partial x_j}\right) \ge 0$$

for every $x \in D$, and every $1 \le i \ne j \le k$.

Claim 14 then follows from the observation that $(\sum_i x_i/k,\ldots,\sum_i x_i/k) \prec x$ for every $x \in [0,1]^k$.

Claim 16. *For every $x \in (0,1]$ we have*

1. *$\phi_z(x) \ge 0$;*
2. *$\phi_z'(x) = \frac{1}{2}\ln\left(\frac{e}{x^{2/z}}\right)\ln\left(\frac{e}{x^{1/z}}\right)^{z/2-1} > 0$, and*
3. *$\phi_z''(x) = -\frac{1}{2xz}\ln\left(\frac{e}{x^{1/z}}\right)^{z/2-2}\left(2\ln\left(\frac{e}{x^{1/z}}\right) + (\frac{z}{2}-1)\ln\left(\frac{e}{x^{2/z}}\right)\right) \le 0$.*

Proof. The derivatives of ϕ_z and the non-negativity of ϕ_z and ϕ_z' can be verified easily. It is also clear that ϕ_z'' is non-positive when $z \ge 2$. Thus it remains to verify $\phi_1''(x) \le 0$ for every x. We have

$$\phi_1''(x) = -\frac{1}{2x}\ln\left(\frac{e}{x}\right)^{-3/2}\left(2\ln\left(\frac{e}{x}\right) - \frac{1}{2}\ln\left(\frac{e}{x^2}\right)\right).$$

It follows from $\frac{1}{2}\ln(e/x^2) \le \ln(e^2/x^2) = 2\ln(e/x)$ that $\phi_1''(x) \le 0$. □

Lemma 17. *g is Schur-concave.*

Proof. Fix $1 \le u \ne v \le k$ and write $g = g_1 + g_2$, where

$$g_1(\alpha_1, \ldots, \alpha_k) := \sum_{\ell=1}^{d} \sum_{\substack{S \subseteq [k], |S|=\ell \\ (S \ni u \wedge S \not\ni v) \vee (S \not\ni u \wedge S \ni v)}} \sum_{\substack{z \in [m]^S \\ \sum_i z_i = d}} \prod_{i \in S} \phi_{z_i}(\alpha_i)$$

and

$$g_2(\alpha_1, \ldots, \alpha_k) := \sum_{\ell=1}^{d} \sum_{\substack{S \subseteq [k], |S|=\ell \\ (S \ni u \wedge S \ni v) \vee (S \not\ni u \wedge S \not\ni v)}} \sum_{\substack{z \in [m]^S \\ \sum_i z_i = d}} \prod_{i \in S} \phi_{z_i}(\alpha_i).$$

We will show that for every $\alpha \in (0,1]^k$, whenever $\alpha_v \le \alpha_u$ we have (1) $\left(\frac{\partial g_1}{\partial \alpha_u} - \frac{\partial g_1}{\partial \alpha_v}\right)(\alpha) \le 0$ and (2) $\left(\frac{\partial g_2}{\partial \alpha_u} - \frac{\partial g_2}{\partial \alpha_v}\right)(\alpha) \le 0$, from which the lemma follows from Theorem 15.

For g_1, since $\phi_z'' \le 0$ and $\alpha_v \le \alpha_u$, we have $\phi_{z_u}'(\alpha_v) \ge \phi_{z_u}'(\alpha_u)$. Moreover, as $\phi_z \ge 0$ and $\phi_z' > 0$, we have

$$\begin{aligned}
\frac{\partial g_1}{\partial \alpha_u}(\alpha) &\le \sum_{\ell=1}^{d} \sum_{\substack{S \subseteq [k], |S|=\ell \\ (S \ni u \wedge S \not\ni v)}} \sum_{\substack{z \in [m]^S \\ \sum_i z_i = d}} \prod_{\substack{i \in S \\ i \ne u}} \phi_{z_i}(\alpha_i) \cdot \phi_{z_u}'(\alpha_u) \cdot \frac{\phi_{z_u}'(\alpha_v)}{\phi_{z_u}'(\alpha_u)} \\
&= \sum_{\ell=1}^{d} \sum_{\substack{S \subseteq [k], |S|=\ell \\ (S \ni u \wedge S \not\ni v)}} \sum_{\substack{z \in [m]^S \\ \sum_i z_i = d}} \prod_{\substack{i \in S \\ i \ne u}} \phi_{z_i}(\alpha_i) \cdot \phi_{z_u}'(\alpha_v) \\
&= \sum_{\ell=1}^{d} \sum_{\substack{S \subseteq [k], |S|=\ell \\ (S \ni v \wedge S \not\ni u)}} \sum_{\substack{z \in [m]^S \\ \sum_i z_i = d}} \prod_{\substack{i \in S \\ i \ne v}} \phi_{z_i}(\alpha_i) \cdot \phi_{z_v}'(\alpha_v) = \frac{\partial g_1}{\partial \alpha_v}(\alpha),
\end{aligned}$$

where in the second equality we simply renamed z_u to z_v.

We now show that $\left(\frac{\partial g_2}{\partial \alpha_u} - \frac{\partial g_2}{\partial \alpha_v}\right)(\alpha) \le 0$ whenever $\alpha_v \le \alpha_u$. For all positive integers z and w, define $\psi_{z,w} \colon (0,1]^2 \to \mathbb{R}$ by

$$\psi_{z,w}(x,y) := \phi_z'(x)\phi_w(y) + \phi_w'(x)\phi_z(y) - \phi_z(x)\phi_w'(y) - \phi_w(x)\phi_z'(y).$$

Note that when $x = y$ we have $\psi_{z,w}(x,x) = 0$. Moreover, when $z = w$ we have $\psi_{z,z}(x,y) = 2(\phi_z'(x)\phi_z(y) - \phi_z(x)\phi_z'(y))$. For every $x, y \in (0,1]$, by Claim 16 we have

$$\frac{\partial}{\partial y}\psi_{z,w}(x,y) = \phi_z'(x)\phi_w'(y) + \phi_w'(x)\phi_z'(y) - \phi_z(x)\phi_w''(y) - \phi_w(x)\phi_z''(y) \ge 0.$$

Since $\psi_{z_u,z_v}(\alpha_u,\alpha_u) = 0$, we have $\psi_{z_u,z_v}(\alpha_u,\alpha_v) \le 0$ whenever $\alpha_v \le \alpha_u$, and so

$$\left(\frac{\partial g_2}{\partial \alpha_u} - \frac{\partial g_2}{\partial \alpha_v}\right)(\alpha) =$$

$$\sum_{\ell=2}^{d} \sum_{\substack{S \subseteq [k] \\ |S|=\ell \\ S \ni u \wedge S \ni v}} \Bigg(\sum_{\substack{z \in [m]^S \\ \sum_i z_i = d \\ z_u = z_v}} \prod_{\substack{i \in S \\ i \ne u \\ i \ne v}} \phi_{z_i}(\alpha_i) \cdot \psi_{z_u,z_v}(\alpha_u,\alpha_v)/2 + \sum_{\substack{z \in [m]^S \\ \sum_i z_i = d \\ z_u < z_v}} \prod_{\substack{i \in S \\ i \ne u \\ i \ne v}} \phi_{z_i}(\alpha_i) \cdot \psi_{z_u,z_v}(\alpha_u,\alpha_v) \Bigg) \le 0$$

because the values ϕ_{z_i} are non-negative. □

2.2 Lower bound

In this section we prove Claim 5. We first restate our claim.

Claim 5. *For all positive integers m and d, there exists a product test $f\colon \{0,1\}^{mk} \to \{0,1\}$ with $k = d \cdot 2^m$ functions of input length m such that*

$$W_{1,d}[f] \geq (m/e^{3/2})^d.$$

Proof. Let $k = d \cdot 2^m$ and $f_1, \ldots, f_k\colon \{0,1\}^{mk} \to \{0,1\}$ be the OR function on k disjoint sets of m bits. It is easy to verify that $\hat{f}_i(\emptyset) = 1 - 2^{-m}$ and $|\hat{f}_i(S)| = 2^{-m}$ for every $S \neq \emptyset$. Consider the product test $f := \prod_{i=1}^k f_i$. Using the fact that $1 - x \geq e^{-x(1+x)}$ for $x \in [0, 1/2]$, we have

$$(1 - 2^{-m})^k \geq e^{-2^m(1+2^{-m})k} \geq e^{-d(1+2^{-m})} \geq e^{-3d/2}.$$

Hence,

$$\begin{aligned}
W_{1,d}[f] &= \sum_{\substack{z \in \{0,\ldots,m\}^k \\ \sum_i z_i = d}} \prod_{i=1}^k W_{z_i}[f_i] \\
&\geq \sum_{|S|=d} \Bigl(\prod_{i \in S} W_{1,1}[f_i] \prod_{i \notin S} W_{1,0}[f_i] \Bigr) \\
&= \binom{k}{d} \cdot (m2^{-m})^d \cdot (1 - 2^{-m})^{k-d} \\
&\geq \left(\frac{d \cdot 2^m}{d} \right)^d \cdot (m2^{-m})^d \cdot e^{-3d/2} \\
&= (m/e^{3/2})^d.
\end{aligned}$$

□

3 Pseudorandom generators

In this section, we use Theorem 4 to construct two pseudorandom generators for product tests. The first one (Theorem 8) has seed length $\tilde{O}(m + \log(k/\varepsilon)) \log(k/\varepsilon)$. The second one (Theorem 6) has a seed length of $\tilde{O}(m + \log(n/\varepsilon))$ but only works for product tests with outputs $\{-1, 1\}$ and their variants (see Corollary 7). We note that Theorem 6 can also be obtained using Theorem 3 in place of Theorem 4.

Both constructions use the Ajtai–Wigderson framework [AW89, GMR+12], and follow from recursively applying the following theorem, which roughly says that $2^{-\tilde{\Omega}(m+\log(k/\varepsilon))}$-almost $O(m + \log(k/\varepsilon))$-wise independence plus constant fraction of noise fools product tests.

Theorem 11. *Let $f\colon \{0,1\}^n \to [-1,1]$ be a product test with k functions of input length m. Let d be a positive integer. Let D and T be two independent δ-almost d-wise independent distributions over $\{0,1\}^n$, and U be the uniform distribution over $\{0,1\}^n$. Then*

$$\bigl| \mathbb{E}[f(D + T \wedge U)] - \mathbb{E}[f(U)] \bigr| \leq k \cdot \bigl(\sqrt{\delta} \cdot (170 \cdot \sqrt{m \ln(ek)})^d + 2^{-(d-m)/2} \bigr),$$

where "+" and "$\wedge$" are bit-wise XOR and AND respectively.

Theorem 11 follows immediately by combining Theorem 4 and Lemma 18 below.

Lemma 18. *Let $f\colon \{0,1\}^n \to [-1,1]$ be a product test with k functions of input length m. Let d be a positive integer. Let D, T, U be a δ-almost $(d+m)$-wise independent, a γ-almost $(d+m)$-wise independent, and the uniform distributions over $\{0,1\}^n$, respectively. Then*

$$\bigl|\mathbb{E}[f(D + T \wedge U)] - \mathbb{E}[f(U)]\bigr| \le k \cdot \bigl(\sqrt{\delta} \cdot W_{1,\le d+m}[f] + 2^{-d/2} + \sqrt{\gamma}\bigr),$$

where "+" and "$\wedge$" are bit-wise XOR and AND respectively.

Proof. We slightly modify the decomposition in [FK18, Proposition 6.1] as follows. Let f be a product test and write $f = \prod_{i=1}^{k} f_i$. As the distribution $D + T \wedge U$ is symmetric, we can assume the function f_i is defined on the ith m bits. For every $i \in \{1,\ldots,k\}$, let $f^{\le i} = \prod_{j \le i} f_j$ and $f^{>i} = \prod_{j>i} f_j$. We decompose f into

$$f = \hat{f}_\emptyset + L + \sum_{i=1}^{k} H_i f^{>i}, \tag{4}$$

where

$$L := \sum_{\substack{\alpha \in \{0,1\}^{mk} \\ 0<|\alpha|<d}} \hat{f}_\alpha \chi_\alpha$$

and

$$H_i := \sum_{\substack{\alpha=(\alpha_1,\ldots,\alpha_i)\in\{0,1\}^{mi}: \\ \text{the } d\text{th } 1 \text{ in } \alpha \text{ appears in } \alpha_i}} \hat{f}^{\le i}_\alpha \chi_\alpha.$$

We now show that the expressions on both sides of Equation (4) are identical. Clearly, every Fourier coefficient on the right hand side is a coefficient of f. To see that every coefficient of f appears on the right hand side exactly once, let $\alpha = (\alpha_1, \ldots, \alpha_k) \in \{0,1\}^{mk}$ and $\hat{f}_\alpha = \prod_{i=1}^{k} \hat{f}_i(\alpha_i)$ be a coefficient of f. If $|\alpha| < d$, then $\hat{f}_\alpha$ appears in $\hat{f}_\emptyset$ or L. Otherwise, $|\alpha| \ge d$. Then the dth 1 in α must appear in one of $\alpha_1, \ldots, \alpha_k$. Say it appears in α_i. Then we claim that α appears in $H_i f^{>i}$. This is because the coefficient indexed by $(\alpha_1, \ldots, \alpha_i)$ appears in H_i, and the coefficient indexed by $(\alpha_{i+1}, \ldots, \alpha_k)$ appears in $f^{>i}$. Note that all the coefficients in each function H_i have weights between d and $d+m$, and because our distributions D and T are both almost $(d+m)$-wise independent, we get an error of $2^{-d} + \gamma$ in Lemma 7.1 in [FK18]. The rest of the analysis follows from [FK18] or [HLV18]. □

3.1 Generator for product tests

We now prove Theorem 8.

Theorem 8. *There exists an explicit generator $G\colon \{0,1\}^\ell \to \{0,1\}^n$ that fools any product test with k functions of input length m with error ε and seed length $O(m + \log(k/\varepsilon))\log(k/\varepsilon)(\log m + \log\log n) = \tilde{O}(m + \log(k/\varepsilon))\log(k/\varepsilon)$.*

The high-level idea is very simple. Let f be a product test. For every choice of D and T in Theorem 11, the function $f'\colon \{0,1\}^T \to [-1,1]$ defined by $f'(y) := f(D + T \wedge y)$ is also a product test. So we can apply Theorem 11 again and recurse. In Lemma 19 below we show that if we repeat this argument for $t = O(\log(k/\varepsilon))$ times with t independent copies of D and T, then for every fixing of $D_1, \ldots, D_t$ and with high probability over the choice of $T_1, \ldots, T_t$, the restricted product test defined on $\{0,1\}^{\bigwedge_{i=1}^t T_i}$ is a product test with k functions of input length $m/2$. Now we simply repeat above for $O(\log m) = \tilde{O}(1)$ steps so that f becomes a constant function and we are done.

Lemma 19. *If there is an explicit generator $G'\colon \{0,1\}^{\ell'} \to \{0,1\}^n$ that fools product tests with k functions of input length $m/2$ with error ε' and seed length ℓ', then there is an explicit generator $G\colon \{0,1\}^{\ell} \to \{0,1\}^n$ that fools product tests with k functions of input length m with error $\varepsilon' + \varepsilon$ and seed length*

$$\ell' + O(\log(k/\varepsilon))\big((m + \log(k/\varepsilon))(\log m + \log\log(k/\varepsilon)) + \log\log n\big) = \ell' + \tilde{O}(m + \log(k/\varepsilon))\log(k/\varepsilon).$$

Proof. Let C be a sufficiently large constant. Let $d = C(m + \log(k/\varepsilon))$, $\delta = d^{-2d}$, and $t = C\log(k/\varepsilon)$. Let $D_1, \ldots, D_t, T_1, \ldots, T_t$ be $2t$ independent δ-almost d-wise independent distributions over $\{0,1\}^n$. Define $D^{(1)} := D_1$ and $D^{(i+1)} := D_{i+1} + T_i \wedge D^{(i)}$.

Let $D := D^{(t)}$, $T := \bigwedge_{i=1}^t T_i$. For a subset $S \subseteq [n]$, define the function $\mathrm{PAD}_S(x)\colon \{0,1\}^{|S|} \to \{0,1\}^n$ to output n bits of which the positions in S are the first $|S|$ bits of $x0^{|S|}$ and the rest are 0. Our generator G outputs

$$D + T \wedge \mathrm{PAD}_T(G').$$

We first look at the seed length of G. By [NN93, Lemma 4.2], sampling the distributions D_i and T_i takes a seed of length

$$\begin{aligned} s &:= t \cdot O(d \log d + \log\log n) \\ &= t \cdot O\big((m + \log(k/\varepsilon))(\log m + \log\log(k/\varepsilon)) + \log\log n\big) \\ &= t \cdot \tilde{O}\big(m + \log(k/\varepsilon)\big). \end{aligned}$$

Hence the total seed length of G is $\ell' + s = \ell' + \tilde{O}(m + \log(k/\varepsilon))\log(k/\varepsilon)$.

We now look at the error of G. By our choice of δ and applying Theorem 11 recursively for t times, we have

$$\begin{aligned} \big|\mathbb{E}[f(D + T \wedge U)] - \mathbb{E}[f(U)]\big| &\le t \cdot k \cdot \Big(\sqrt{\delta} \cdot \big(170 \cdot \sqrt{m \ln(ek)}\big)^d + 2^{-(d-m)/2}\Big) \\ &\le t \cdot k \cdot \Big(\Big(\frac{170\sqrt{m \ln(ek)}}{d}\Big)^d + 2^{-\Omega(d)}\Big) \\ &\le t \cdot 2^{-\Omega(d)} \le \varepsilon/2. \end{aligned}$$

Next, we show that for every fixing of D and most choices of T, the function $f_{D,T}(y) := f(D + T \wedge y)$ is a product test with k functions of input length $m/2$, which can be fooled by G'.

Because the variables T_i are independent and each of them is δ-almost d-wise independent, for every subset $I \subseteq [n]$ of size at most $m \le d$, we have

$$\Pr\big[|T \cap I| \ge m/2\big] \le \binom{|I|}{m/2}(2^{-m/2}+\delta)^t \le 2^m \cdot 2^{-\Omega(mt)} \le \varepsilon/2k.$$

It follows by a union bound over the k subsets $I_1, \dots, I_k$ that for every fixing of D, with probability at least $1-\varepsilon/2$ over the choice of T, the function $f_{D,T}$ is a product test with k functions of input length $m/2$, which can be fooled by G' with error ε'. Hence G fools f with error $\varepsilon' + \varepsilon$. □

Proof of Theorem 8. We apply Lemma 19 recursively for $r := O(\log m) = \tilde{O}(1)$ times. Note that a product test of input length 0 is a constant function, which can always be fooled with zero error. Hence we have a generator that fools product tests with k functions of input length m, with error $r \cdot \varepsilon$ and seed length

$$r \cdot O\big((m+\log(k/\varepsilon))(\log m+\log\log(k/\varepsilon))+\log\log n\big)\log(k/\varepsilon) = \tilde{O}(\log(k/\varepsilon)+m)\log(k/\varepsilon). \quad \square$$

Replacing ε with ε/r proves the theorem.

3.2 Almost-optimal generator for XOR of Boolean functions

In this section, we construct our generator for product tests with outputs $\{-1,1\}$, which correspond to the XOR of Boolean functions f_i defined on disjoint inputs. Throughout this section we will call these tests $\{-1,1\}$-*products*. We first restate our theorem.

Theorem 6. *There exists an explicit generator $G\colon \{0,1\}^\ell \to \{0,1\}^n$ that fools the XOR of any k Boolean functions on disjoint inputs of length $\le m$ with error ε and seed length $O(m+\log(n/\varepsilon))(\log m+\log\log(n/\varepsilon))^2 = \tilde{O}(m+\log(n/\varepsilon))$.*

Theorem 6 relies on applying the following lemma recursively in different ways. From now on, we will relax our tests to allow one of the k functions to have input length greater than m, but bounded by $O(m+\log(n/\varepsilon))$.

Lemma 20. *There exists a constant C such that the following holds. Let m and s be two integers such that $m \ge C\log\log(n/\varepsilon)$ and $s = 5(m+\log(n/\varepsilon))$. If there is an explicit generator $G'\colon \{0,1\}^{\ell'} \to \{0,1\}^n$ that fools $\{-1,1\}$-products with $k' \le 16^{m+1}$ functions, $k'-1$ of which have input lengths $\le m/2$ and one has length $\le s$, with error ε' and seed length ℓ', then there is an explicit generator $G\colon \{0,1\}^\ell \to \{0,1\}^n$ that fools $\{-1,1\}$-products with $k \le 16^{2m+1}$ functions, $k-1$ of which have input lengths $\le m$ and one has length $\le s$, with error $\varepsilon'+\varepsilon$ and seed length $\ell = \ell' + O(m+\log(n/\varepsilon))(\log m+\log\log(n/\varepsilon)) = \ell' + \tilde{O}(m+\log(n/\varepsilon))$.*

The proof of Lemma 20 closely follows a construction by Meka, Reingold and Tal [MRT18]. First of all, we will use the following generator in [MRT18]. It fools any $\{-1,1\}$-products when the number of functions k is significantly greater than the input length m of the functions f_i.

Lemma 21 (Lemma 6.2 in [MRT18])**.** *There exists a constant C such that the following holds. Let n, k, m, s be integers such that $C \log\log(n/\varepsilon) \leq m \leq \log n$ and $16^m \leq k \leq 2 \cdot 16^{2m}$. There exists an explicit pseudorandom generator $G_{\oplus\text{Many}} \colon \{0,1\}^\ell \to \{0,1\}^n$ that fools $\{-1,1\}$-products with k non-constant functions, $k-1$ of which have input lengths $\leq m$ and one has length $\leq s$, with error ε and seed length $O(s + \log(n/\varepsilon))$.*

Here is the high-level idea of proving Lemma 20. We consider two cases depending on whether k is large with respect to m. If $k \geq 16^m$, then by Lemma 21, the generator $G_{\oplus\text{Many}}$ fools f. Otherwise, we show that for every fixing of D and most choices of T, the restriction of f under (D,T) is a $\{-1,1\}$-product with k functions, $k-1$ of which have input length $\leq m/2$ and one has length $\leq s$. More specifically, we will show that for most choices of T, the following would happen: for the function with input length $\leq s$, at most $s/2$ of its inputs remain in T; for the rest of the functions with input length $\leq m$, after being restricted by (D,T), at most $\lceil s/2m \rceil$ of them have input length $> m/2$, and so they are defined on a total of $s/2$ positions in T. Now we can think of these "bad" functions as one function with input length $\leq s$, and the rest of the at most k "good" functions have input length $m/2$. So we can apply the generator G' in our assumption.

Proof of Lemma 20. Let C be the constant in Lemma 21 and C' be a sufficiently large constant.

Let $d = C's$ and $\delta = d^{-2d}$. Let $D_1, \ldots, D_{50}, T_1, \ldots, T_{50}$ be 100 independent δ-almost d-wise independent distributions over $\{0,1\}^n$. Define $D^{(1)} := D_1$ and $D^{(i+1)} := D_{i+1} + T_i \wedge D^{(i)}$.

Let $D := D^{(50)}$, $T := \bigwedge_{i=1}^{50} T_i$ and $G_{\oplus\text{Many}}$ be the generator in Lemma 21 with respect to the values of n, k, m, s given in this lemma. For a subset $S \subseteq [n]$, define the function $\text{PAD}_S(x) \colon \{0,1\}^{|S|} \to \{0,1\}^n$ to output n bits of which the positions in S are the first $|S|$ bits of $x0^{|S|}$ and the rest are 0. Our generator G outputs

$$(D + T \wedge \text{PAD}_T(G')) + G_{\oplus\text{Many}}.$$

We first look at the seed length of G. By Lemma 21, $G_{\oplus\text{Many}}$ uses a seed of length $O(s + \log(n/\varepsilon)) = O(m + \log(n/\varepsilon))$. By [NN93, Lemma 4.2], sampling the distributions D_i and T_i takes a seed of length

$$O(s \log s) = O\big(m + \log(n/\varepsilon)\big)(\log m + \log\log(n/\varepsilon)) = \tilde{O}(m + \log(n/\varepsilon)).$$

Hence the total seed length of G is $\ell' + O(m + \log(n/\varepsilon))(\log m + \log\log(n/\varepsilon)) = \ell' + \tilde{O}(m + \log(n/\varepsilon))$.

We now show that G fools f. Write $f = \prod_{i=1}^k f_i$, where $f_i \colon \{0,1\}^{I_i} \to \{-1,1\}$. Without loss of generality we can assume each function f_i is non-constant. We consider two cases.

k is large: If $k \geq 16^m$, then for every fixing of D, T and G', the function $f'(y) := f(D + T \wedge \text{PAD}_T(G') + y)$ is also a $\{-1,1\}$-product with the same parameters as f. Note that we always have $k \leq n$ and so $m \leq \log n$. Hence it follows from Lemma 21 that the generator $G_{\oplus\text{Many}}$ fools f' with error ε. Averaging over D, T and G' shows that G fools f with error ε.

k **is small:** Now suppose $k \le 16^m$. For every fixing of $G_{\oplus\text{Many}}$, consider $f'(y) := f(y + G_{\oplus\text{Many}})$. Again, f' is a $\{-1,1\}$-product with the same parameters as f. In particular, it is a $\{-1,1\}$-product with k functions with input length s. So, by our choice of δ and applying Theorem 11 recursively for 50 times, we have

$$\begin{aligned}\left|\mathbb{E}[f'(D + T \wedge U)] - \mathbb{E}[f'(U)]\right| &\le 50 \cdot k \cdot \left(\sqrt{\delta} \cdot \left(170 \cdot \sqrt{s \ln(ek)}\right)^d + 2^{-(d-s)/2}\right) \\ &\le 50 \cdot 2^s \cdot \left((170s/d)^d + 2^{-\Omega(s)}\right) \\ &\le 2^{-\Omega(s)} \le \varepsilon/2.\end{aligned}$$

Next, we show that for every fixing of D and most choices of T, the function $f'_{D,T}(y) := f'(D + T \wedge y)$ is a $\{-1,1\}$-product with k functions, $k-1$ of which have input lengths $\le m/2$ and one has length $\le s$, which can be fooled by G'.

Because the variables T_i are independent and each of them is δ-almost d-wise independent, for every subset $I \subseteq [n]$ of size at most d, we have

$$\Pr[T \cap I = I] = \prod_{i=1}^{50} \Pr[T_i \cap I = I] \le (2^{-|I|} + \delta)^{50} \le (3/4)^{-50|I|}.$$

Without loss of generality, we assume $I_1, \ldots, I_{k-1}$ are the subsets of size at most m and I_k is the subset of size at most s. We now look at which subsets $T \cap I_i$ have length at most $m/2$ and which subsets do not. For the latter, we collect the indices in these subsets.

Let $G := \{i \in [k-1] : |T \cap I_i| \le m/2\}$, $B := \{i \in [k-1] : |T \cap I_i| > m/2\}$ and $BV := \{j \in [n] : j \in \bigcup_{i \in B}(T \cap I_i)\}$. We claim that with probability $1 - \varepsilon/2$ over the choice of T, we have $|BV| \le s$. Note that the indices in BV either come from I_k, or I_i for $i \in [k-1]$. For the first case, the probability that at least $s/2$ of the indices in I_k appear in BV is at most

$$\binom{|I_k|}{s/2}(3/4)^{-25s} \le 2^s \cdot (3/4)^{-25s} \le \varepsilon/4.$$

For the second case, note that if at least $s/2$ of the variables in $\bigcup_{i \in [k-1]} I_i$ appear in BV, then they must appear in at least $\lceil s/2m \rceil$ of the subsets $T \cap I_1, \ldots, T \cap I_{k-1}$. The probability of the former is at most the probability of the latter, which is at most

$$\binom{k-1}{\lceil s/2m \rceil}\binom{m \cdot \lceil s/2m \rceil}{s/2}(3/4)^{-25s} \le 16^{m \cdot (s/2m+1)} \cdot 2^{m \cdot (s/2m+1)} \cdot (3/4)^{-25s} \le \varepsilon/4,$$

because $k \le 16^m$ and $m \le s$. Hence with probability $1 - \varepsilon/2$ over the choice of T, the function $f'_{D,T}$ is a product $g \cdot h$, where g is a product of $|G| \le k-1$ functions of input length $m/2$, and h is a product of $|B|+1$ functions defined on a total of $|BV| \le s$ bits. Recall that $k \le 16^m$, so by our assumption G' fools $f'_{D,T}$ with error ε'. Therefore G fools f with error $\varepsilon + \varepsilon'$. □

We obtain Theorem 6 by applying Lemma 20 repeatedly in different ways.

Proof of Theorem 6. Given a $\{-1,1\}$-product $f \colon \{0,1\}^n \to \{-1,1\}$ with k functions of input length m, we will apply Lemma 20 in stages. In each stage, we start with a $\{-1,1\}$-product

f with k_1 functions, $k_1 - 1$ of which have input lengths $\leq m_1 = \max\{m, 2\log(n/\varepsilon)\}$ and one has length $\leq s := 5(m + \log(n/\varepsilon))$. Note that $k_1 \leq 16^{2m_1+1}$. Let C be the constant in Lemma 20. We apply Lemma 20 for $t = O(\log m_1)$ times until f is restricted to a $\{-1,1\}$-product f' with k_2 functions, $k_2 - 1$ of which have input lengths $\leq m_2$ and one has length $\leq s$, where $m_2 = C \log\log(n/\varepsilon)$, $k_2 \leq 16^{2m_2+1} \leq (\log(n/\varepsilon))^r$, and $r := 8C + 4$ is a constant. This uses a seed of length

$$\begin{aligned} t \cdot O(m + \log(n/\varepsilon))(\log m + \log\log(n/\varepsilon)) &\leq O(m + \log(n/\varepsilon))(\log m + \log\log(n/\varepsilon))^2 \\ &= \tilde{O}(m + \log(n/\varepsilon)). \end{aligned}$$

At the end of each stage, we repeat the above argument by grouping every $\lceil \log(n/\varepsilon)/m_2 \rceil$ functions of f' that have input lengths $\leq m_2$ as one function of input length $\leq 2\log(n/\varepsilon)$, so we can think of f' as a $\{-1,1\}$-product with $k_3 := k_2/\lceil m_2/(\log n) \rceil \leq (\log(n/\varepsilon))^{r-1} \log\log n$ functions, $k_3 - 1$ of which have input lengths $\leq \log(n/\varepsilon)$ and one has length $\leq s$.

Repeating above for $r + 1 = O(1)$ stages, we are left with a $\{-1,1\}$-product of two functions, one has input length $\leq C \log\log(n/\varepsilon)$, and one has length $\leq s$, which can then be fooled by a $2^{-\Omega(s)}$-biased distribution that can be sampled using $O(m+\log(n/\varepsilon))$ bits [NN93]. So the total seed length is $O(m + \log(n/\varepsilon))(\log m + \log\log(n/\varepsilon))^2 = \tilde{O}(m + \log(n/\varepsilon))$, and the error is $(r + 1) \cdot t \cdot \varepsilon$. Replacing ε with $\varepsilon/(r + 1)t$ proves the theorem. □

4 Level-k inequalities

In this section, we prove Lemma 10 that gives an upper bound on the dth level Fourier weight of a $[0,1]$-valued function in L_2-norm. We first restate the lemma.

Lemma 10. *Let $g\colon \{0,1\}^n \to [0,1]$ be any function. For every positive integer d, we have*

$$W_{2,d}[g] \leq 4\,\mathbb{E}[g]^2 \big(2e \ln(e/\,\mathbb{E}[g]^{1/d})\big)^d.$$

Our proof closely follows the argument in [Tal96].

Claim 22. *Let $f\colon \{0,1\}^n \to \mathbb{R}$ have Fourier degree at most d and $\|f\|_2 = 1$. Let $g\colon \{0,1\}^n \to [0,1]$ be any function. If $t_0 \geq 2e^{d/2}$, then*

$$\mathbb{E}\big[g(x)|f(x)|\big] \leq \mathbb{E}[g]t_0 + 2et_0^{1-2/d} e^{-\frac{d}{2e}t_0^{2/d}}.$$

To prove this claim, we will use the following concentration inequality for functions with Fourier degree k from [DFKO07].

Theorem 23 (Lemma 2.2 in [DFKO07])**.** *Let $f\colon \{0,1\}^n \to \mathbb{R}$ have Fourier degree at most d and assume that $\|f\|_2 := \sum_S \hat{f}_S^2 = 1$. Then for any $t \geq (2e)^{d/2}$,*

$$\Pr\big[|f| \geq t\big] \leq e^{-\frac{d}{2e}t^{2/d}}.$$

We also need to bound above the integral of $e^{-\frac{d}{2e}t^{2/d}}$.

Claim 24. *Let d be any positive integer. If $t_0 \geq (2e)^{d/2}$, then we have*

$$\int_{t_0}^{\infty} e^{-\frac{d}{2e}t^{2/d}}dt \leq 2et_0^{1-2/d}e^{-\frac{d}{2e}t_0^{2/d}}.$$

Proof. First we apply the following change of variable to the integral. We set $s = \frac{d}{2e}t^{2/d}$ and obtain

$$\int_{t_0}^{\infty} e^{-\frac{d}{2e}t^{2/d}}dt = e\Big(\frac{2e}{d}\Big)^{d/2-1}\int_{s_0}^{\infty} s^{d/2-1}e^{-s}ds,$$

where $s_0 = \frac{d}{2e}t_0^{2/d}$. Define

$$\Gamma_{s_0}(d) = \int_{s_0}^{\infty} s^{d-1}e^{-s}ds.$$

(Note that when $s_0 = 0$ then $\Gamma_0(d)$ is the Gamma function.) Using integration by parts, we have

$$\Gamma_{s_0}(d) = s_0^{d-1}e^{-s_0} + (d-1)\Gamma_{s_0}(d-1). \tag{5}$$

Moreover, when $d \leq 1$, we have $\Gamma_{s_0}(d) \leq s_0^{d-1}\int_{s_0}^{\infty} e^{-s}ds = s_0^{d-1}e^{-s_0}$.

Note that if $t_0 \geq (2e)^{d/2}$, then $s_0 \geq d-2$. Hence, if we open the recursive definition of $\Gamma_{s_0}(d/2)$ in Equation (5), we have

$$\begin{aligned}\Gamma_{s_0}(d/2) &\leq e^{-s_0}\sum_{i=0}^{\lceil\frac{d}{2}\rceil-1} s_0^{d/2-1-i}\prod_{j=1}^{i}(d/2-j)\\ &\leq e^{-s_0}s_0^{d/2-1}\sum_{i=0}^{\lceil\frac{d}{2}\rceil-1}\Big(\frac{d/2-1}{s_0}\Big)^i\\ &\leq 2e^{-s_0}s_0^{d/2-1},\end{aligned}$$

because the summation is a geometric sum with ratio at most $1/2$. Substituting s_0 with t_0, we obtain

$$\begin{aligned}e\Big(\frac{2e}{d}\Big)^{d/2-1}\int_{s_0}^{\infty} s^{d/2-1}e^{-s}ds &\leq 2e\Big(\frac{2e}{d}\Big)^{d/2-1}e^{-s_0}s_0^{d/2-1}\\ &= 2et_0^{1-2/d}e^{-\frac{d}{2e}t_0^{2/d}}.\end{aligned}$$

□

Proof of Claim 22. We rewrite $|f(x)|$ as $\int_0^{|f(x)|}\mathbb{1}dt = \int_0^{\infty}\mathbb{1}(|f(x)| \geq t)dt$ and obtain

$$\begin{aligned}\mathop{\mathbb{E}}_{x\sim\{0,1\}^n}[g(x)|f(x)|] &= \mathop{\mathbb{E}}_{x\sim\{0,1\}^n}\Big[\int_0^{\infty} g(x)\mathbb{1}(|f(x)| \geq t)dt\Big]\\ &\leq \mathop{\mathbb{E}}_{x\sim\{0,1\}^n}\Big[\int_0^{\infty}\min\big\{g(x), \mathbb{1}(|f(x)| \geq t)\big\}dt\Big]\\ &= \int_0^{\infty}\min\Big\{\mathbb{E}[g], \Pr_x[|f(x)| \geq t]\Big\}dt\\ &\leq \int_0^{t_0}\mathbb{E}[g]dt + \int_{t_0}^{\infty}\Pr\big[|f(x)| \geq t\big]dt\\ &\leq \mathbb{E}[g]t_0 + \int_{t_0}^{\infty} e^{-\frac{d}{2e}t^{2/d}}dt.\end{aligned}$$

Since $t_0 \geq (2e)^{d/2}$, by Claim 24 this is at most $\mathbb{E}[g]t_0 + 2et_0^{1-2/d}e^{-\frac{d}{2e}t_0^{2/d}}$. □

Proof of Lemma 10. Define f to be $f(x) := \sum_{|S|=d} \hat{f}_S \chi_S(x)$, where $\hat{f}_S = \hat{g}_S \big(\sum_{|T|=d} \hat{g}_T^2\big)^{-1/2}$. Note that $\|f\|_2 = 1$, and we have

$$\mathbb{E}[g(x)f(x)] = \frac{\sum_S \hat{g}_S \, \mathbb{E}[g(x)\chi_S(x)]}{\big(\sum_{|T|=d} \hat{g}_T^2\big)^{1/2}} = \Big(\sum_{|S|=d} \hat{g}_S^2\Big)^{1/2}.$$

Let $t_0 = (2e \ln(e/\,\mathbb{E}[g]^{1/d}))^{d/2} \geq (2e)^{d/2}$. By Claim 22,

$$\Big(\sum_{|S|=d} \hat{g}_S^2\Big)^{1/2} = \mathbb{E}[g(x)f(x)] \leq \mathbb{E}[g(x)|f(x)|] \leq \mathbb{E}[g]t_0 + 2et_0^{1-2/d}e^{-\frac{d}{2e}t_0^{2/d}}.$$

By our choice of t_0, the second term is at most

$$2et_0^{1-2/d}e^{-\frac{d}{2e}t_0^{2/d}} \leq \Big(2e \ln\Big(\frac{e}{\mathbb{E}[g]^{1/d}}\Big)\Big)^{d/2} \frac{\mathbb{E}[g]}{e^d} \leq (2/e)^{d/2}\, \mathbb{E}[g] \ln\Big(\frac{e}{\mathbb{E}[g]^{1/d}}\Big)^{d/2},$$

which is no greater than the first term. So

$$\Big(\sum_{|S|=d} \hat{g}_S^2\Big)^{1/2} \leq 2\,\mathbb{E}[g]\big(2e \ln(e/\,\mathbb{E}[g]^{1/d})\big)^{d/2}.$$

and the lemma follows. □

Acknowledgement

I thank Salil Vadhan for asking about the Fourier spectrum of product tests. I also thank Andrej Bogdanov, Gil Cohen, Amnon Ta-Shma, Avishay Tal and Emanuele Viola for very helpful conversations. I am grateful to Emanuele Viola for his invaluable comments on the write-up.

References

[Aar10] Scott Aaronson. BQP and the polynomial hierarchy. In *42nd ACM Symp. on the Theory of Computing (STOC)*, pages 141–150. ACM, 2010. 1

[ABO84] Miklós Ajtai and Michael Ben-Or. A theorem on probabilistic constant depth computations. In *16th ACM Symp. on the Theory of Computing (STOC)*, pages 471–474, 1984. 1

[Ajt83] Miklós Ajtai. Σ_1^1-formulae on finite structures. *Annals of Pure and Applied Logic*, 24(1):1–48, 1983. 1

[AKS87] Miklós Ajtai, János Komlós, and Endre Szemerédi. Deterministic simulation in logspace. In *19th ACM Symp. on the Theory of Computing (STOC)*, pages 132–140, 1987. 1

[Ama09] Kazuyuki Amano. Bounds on the size of small depth circuits for approximating majority. In *36th Coll. on Automata, Languages and Programming (ICALP)*, pages 59–70. Springer, 2009. 1

[ASWZ96] Roy Armoni, Michael E. Saks, Avi Wigderson, and Shiyu Zhou. Discrepancy sets and pseudorandom generators for combinatorial rectangles. In *37th IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 412–421, 1996. 1

[AW89] Miklos Ajtai and Avi Wigderson. Deterministic simulation of probabilistic constant-depth circuits. *Advances in Computing Research - Randomness and Computation*, 5:199–223, 1989. 1, 1.1.2, 3

[BPW11] Andrej Bogdanov, Periklis A. Papakonstantinou, and Andrew Wan. Pseudorandomness for read-once formulas. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 240–246, 2011. 1

[BV10] Joshua Brody and Elad Verbin. The coin problem, and pseudorandomness for branching programs. In *51th IEEE Symp. on Foundations of Computer Science (FOCS)*, 2010. 1

[CGR14] Gil Cohen, Anat Ganor, and Ran Raz. Two sides of the coin problem. In *Workshop on Randomization and Computation (RANDOM)*, pages 618–629, 2014. 1

[Cha02] Mei-Chu Chang. A polynomial bound in Freiman's theorem. *Duke Math. J.*, 113(3):399–419, 2002. 1.1.1, 2

[CHHL18] Eshan Chattopadhyay, Pooya Hatami, Kaave Hosseini, and Shachar Lovett. Pseudorandom generators from polarizing random walks. *Electronic Colloquium on Computational Complexity*, Technical Report TR18-015, 2018. www.eccc.uni-trier.de/. 1

[CHLT19] Eshan Chattopadhyay, Pooya Hatami, Shachar Lovett, and Avishay Tal. Pseudorandom generators from the second fourier level and applications to ac0 with parity gates. In *ITCS'19—Proceedings of the 2019 ACM Conference on Innovations in Theoretical Computer Science*. 2019. 1, 1

[CHRT18] Eshan Chattopadhyay, Pooya Hatami, Omer Reingold, and Avishay Tal. Improved pseudorandomness for unordered branching programs through local monotonicity. In *ACM Symp. on the Theory of Computing (STOC)*, 2018. 1, 1, 1, 1

[CRS00] Suresh Chari, Pankaj Rohatgi, and Aravind Srinivasan. Improved algorithms via approximations of probability distributions. *J. Comput. System Sci.*, 61(1):81–107, 2000. 1

[CSV15] Sitan Chen, Thomas Steinke, and Salil P. Vadhan. Pseudorandomness for read-once, constant-depth circuits. *CoRR*, abs/1504.04675, 2015. 1

[DETT10] Anindya De, Omid Etesami, Luca Trevisan, and Madhur Tulsiani. Improved pseudorandom generators for depth 2 circuits. In *Workshop on Randomization and Computation (RANDOM)*, pages 504–517, 2010. 1

[DFKO07] Irit Dinur, Ehud Friedgut, Guy Kindler, and Ryan O'Donnell. On the Fourier tails of bounded functions over the discrete cube. *Israel J. Math.*, 160:389–412, 2007. 4, 23

[DHH18] Dean Doron, Pooya Hatami, and William Hoza. Near-optimal pseudorandom generators for constant-depth read-once formulas. 2018. ECCC TR18-183. 1

[EGL$^+$98] Guy Even, Oded Goldreich, Michael Luby, Noam Nisan, and Boban Velickovic. Efficient approximation of product distributions. *Random Struct. Algorithms*, 13(1):1–16, 1998. 1

[FK18] Michael A. Forbes and Zander Kelley. Pseudorandom generators for read-once branching programs, in any order. In *47th IEEE Symposium on Foundations of Computer Science (FOCS)*, 2018. 1, 1.1.2, 3, 3

[GKM15] Parikshit Gopalan, Daniel Kane, and Raghu Meka. Pseudorandomness via the discrete fourier transform. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 903–922, 2015. 1, 1, 1

[GMR$^+$12] Parikshit Gopalan, Raghu Meka, Omer Reingold, Luca Trevisan, and Salil Vadhan. Better pseudorandom generators from milder pseudorandom restrictions. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, 2012. 1, 1, 1.1.2, 3

[GSW16] Parikshit Gopalan, Rocco A. Servedio, and Avi Wigderson. Degree and sensitivity: tails of two distributions. In *31st Conference on Computational Complexity*, volume 50 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 13, 23. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2016. 1

[GY14] Parikshit Gopalan and Amir Yehudayoff. Inequalities and tail bounds for elementary symmetric polynomials. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:19, 2014. 1, 1

[HLV18] Elad Haramaty, Chin Ho Lee, and Emanuele Viola. Bounded independence plus noise fools products. *SIAM J. Comput.*, 47(2):493–523, 2018. 1, 1, 1.1.2, 3

[HT18] Pooya Hatami and Avishay Tal. Pseudorandom generators for low-sensitivity functions. In *9th Innovations in Theoretical Computer Science*, volume 94 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 29, 13. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2018. 1

[IMR14] Russell Impagliazzo, Cristopher Moore, and Alexander Russell. An entropic proof of Chang's inequality. *SIAM J. Discrete Math.*, 28(1):173–176, 2014. 1.1.1, 1.1.1, 2

[IMZ12] Russell Impagliazzo, Raghu Meka, and David Zuckerman. Pseudorandomness from shrinkage. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 111–119, 2012. 1

[INW94] Russell Impagliazzo, Noam Nisan, and Avi Wigderson. Pseudorandomness for network algorithms. In *26th ACM Symp. on the Theory of Computing (STOC)*, pages 356–364, 1994. 1

[KK13] Nathan Keller and Guy Kindler. Quantitative relation between noise sensitivity and influences. *Combinatorica*, 33(1):45–71, 2013. 1.1.1, 2

[KS18] Swastik Kopparty and Srikanth Srinivasan. Certifying polynomials for $\mathrm{AC}^0[\oplus]$ circuits, with applications to lower bounds and circuit compression. *Theory Comput.*, 14:Article 12, 24, 2018. 1

[LSS$^+$18] Nutan Limaye, Karteek Sreenivasiah, Srikanth Srinivasan, Utkarsh Tripathi, and S Venkitesh. The coin problem in constant depth: Sample complexity and parity gates. In *Electronic Colloquium on Computational Complexity (ECCC)*, number TR18–157, 2018. 1

[Lu02] Chi-Jen Lu. Improved pseudorandom generators for combinatorial rectangles. *Combinatorica*, 22(3):417–433, 2002. 1

[LV17] Chin Ho Lee and Emanuele Viola. More on bounded independence plus noise: Pseudorandom generators for read-once polynomials. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 24, page 167, 2017. 1, 1, 1, 1

[LV18] Chin Ho Lee and Emanuele Viola. The coin problem for product tests. *ACM Trans. Comput. Theory*, 10(3):Art. 14, 10, 2018. 1, 1, 1.1.1

[Man95] Yishay Mansour. An $O(n^{\log\log n})$ learning algorithm for DNF under the uniform distribution. *J. Comput. System Sci.*, 50(3, part 3):543–550, 1995. Fifth Annual Workshop on Computational Learning Theory (COLT) (Pittsburgh, PA, 1992). 1

[MRT18] Raghu Meka, Omer Reingold, and Avishay Tal. Pseudorandom generators for width-3 branching programs. *arXiv preprint arXiv:1806.04256*, 2018. 1, 1, 1, 1.1.2, 3.2, 21

[Nis92] Noam Nisan. Pseudorandom generators for space-bounded computation. *Combinatorica*, 12(4):449–461, 1992. 1

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: efficient constructions and applications. *SIAM J. on Computing*, 22(4):838–856, 1993. 3.1, 3.2, 3.2

[NZ96] Noam Nisan and David Zuckerman. Randomness is linear in space. *J. of Computer and System Sciences*, 52(1):43–52, February 1996. 1

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions.* Cambridge University Press, 2014. 1.1.1, 2

[PPT92] Josip E. Pečarić, Frank Proschan, and Y. L. Tong. *Convex functions, partial orderings, and statistical applications*, volume 187 of *Mathematics in Science and Engineering*. Academic Press, Inc., Boston, MA, 1992. 15

[RS17] Benjamin Rossman and Srikanth Srinivasan. Separation of $\mathrm{AC}^0[\oplus]$ formulas and circuits. In *44th International Colloquium on Automata, Languages, and Programming*, volume 80 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 50, 13. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2017. 1

[RSV13] Omer Reingold, Thomas Steinke, and Salil P. Vadhan. Pseudorandomness for regular branching programs via Fourier analysis. In *Workshop on Randomization and Computation (RANDOM)*, pages 655–670, 2013. 1, 1, 1

[ST18] Rocco A. Servedio and Li-Yang Tan. Improved pseudorandom generators from pseudorandom multi-switching lemmas. *CoRR*, abs/1801.03590, 2018. 1

[Ste04] J. Michael Steele. *The Cauchy-Schwarz master class.* MAA Problem Books Series. Mathematical Association of America, Washington, DC; Cambridge University Press, Cambridge, 2004. 2

[Ste13] John P. Steinberger. The distinguishability of product distributions by read-once branching programs. In *IEEE Conf. on Computational Complexity (CCC)*, pages 248–254, 2013. 1

[SV10] Ronen Shaltiel and Emanuele Viola. Hardness amplification proofs require majority. *SIAM J. on Computing*, 39(7):3122–3154, 2010. 1

[SVW14] Thomas Steinke, Salil P. Vadhan, and Andrew Wan. Pseudorandomness and fourier growth bounds for width-3 branching programs. In *Workshop on Randomization and Computation (RANDOM)*, pages 885–899, 2014. 1, 1, 1

[Tal96] Michel Talagrand. How much are increasing sets positively correlated? *Combinatorica*, 16(2):243–258, 1996. 1.1.1, 2, 4

[Tal17] Avishay Tal. Tight bounds on the fourier spectrum of AC0. In *Conf. on Computational Complexity (CCC)*, pages 15:1–15:31, 2017. 1, 1

[TX13] Luca Trevisan and TongKe Xue. A derandomized switching lemma and an improved derandomization of ac0. In *Computational Complexity (CCC), 2013 IEEE Conference on*, pages 242–247. IEEE, 2013. 1

[Val84] Leslie G. Valiant. Short monotone formulae for the majority function. *J. Algorithms*, 5(3):363–366, 1984. 1

[Vio09] Emanuele Viola. On approximate majority and probabilistic time. *Computational Complexity*, 18(3):337–375, 2009. 1

[Vio14] Emanuele Viola. Randomness buys depth for approximate counting. *Computational Complexity*, 23(3):479–508, 2014. 1, 1

[Wat13] Thomas Watson. Pseudorandom generators for combinatorial checkerboards. *Computational Complexity*, 22(4):727–769, 2013. 1