THEORY OF COMPUTING, Volume 16 (7), 2020, pp. 1–50
*www.theoryofcomputing.org*

# More on Bounded Independence Plus Noise: Pseudorandom Generators for Read-Once Polynomials

Chin Ho Lee[*] Emanuele Viola[†]

*Received March 30, 2019; Revised December 26, 2019; Published October 9, 2020*

**Abstract.** We construct pseudorandom generators with improved seed length for several classes of tests. First, we consider the class of read-once polynomials over GF(2) in $m$ variables. For error $\varepsilon$ we obtain seed length $\tilde{O}(\log(m/\varepsilon)\log(1/\varepsilon))$. This is optimal up to a factor of $\log(1/\varepsilon)\cdot\operatorname{poly}\log\log(m/\varepsilon)$. The previous best seed length was polylogarithmic in $m$ and $1/\varepsilon$.

Second, we consider product tests $f\colon \{0,1\}^m \to \mathbb{C}_{\le 1}$. These tests are the product of $k$ functions $f_i\colon \{0,1\}^\ell \to \mathbb{C}_{\le 1}$, where the inputs of the $f_i$ are disjoint subsets of the $m$ variables and $\mathbb{C}_{\le 1}$ is the complex unit disk. Here we obtain seed length $\ell\cdot\operatorname{polylog}(m/\varepsilon)$. This implies better generators for other classes of tests. If moreover the $f_i$ have output range $\{-1,0,1\}$ then we obtain seed length $\tilde{O}((\log(k/\varepsilon)+\ell)(\log(1/\varepsilon)+\log\log m))$. This is again optimal up to a factor of $\log(1/\varepsilon)\cdot\operatorname{polylog}(\ell,\log k,\log m,\log(1/\varepsilon))$, while the previous best seed length was $\ge\sqrt{k}$.

A main component of our proofs is showing that these classes of tests are fooled by almost $d$-wise independent distributions perturbed with noise.

**ACM Classification:** G.3

**AMS Classification:** 68Q87, 68W20

**Key words and phrases:** pseudorandom generators, bounded independence plus noise, branching programs, read-once polynomials

[*] Supported by NSF CCF award 1813930, the Croucher Foundation and the Simons Collaboration on Algorithms and Geometry. Most of this work was carried out while the author was at Northeastern University.

[†] Supported by NSF CCF award 1813930.

DOI: 10.4086/toc.2020.v016a007

# 1 Introduction

A pseudorandom generator (PRG) is a function that stretches a few input bits to a longer output so that on a uniformly distributed input, the output distribution of the generator fools, i. e., looks random to, some tests.

**Definition 1.1.** A distribution $D$ over a finite set $\Omega$ *fools* a family $\mathcal{F}$ of functions $f\colon \Omega \to \mathbb{C}$ with error $\varepsilon$, if

$$\left| \mathop{\mathbb{E}}_{x\sim D}[f(x)] - \mathop{\mathbb{E}}_{u\sim U}[f(u)] \right| \le \varepsilon$$

for every $f \in \mathcal{F}$, where $U$ is the uniform distribution over $\Omega$.

We often use the letter $U$ to denote a random variable, uniformly distributed over the shared domain $\Omega$ of a family $\mathcal{F}$ of functions.

**Definition 1.2.** A *pseudorandom generator (PRG)* for a family $\mathcal{F}$ of functions with domain $\Omega$ is a function $G\colon \{0,1\}^s \to \Omega$ such that $G(U)$ fools $\mathcal{F}$, where $U$ is a random variable distributed uniformly over $\{0,1\}^s$.

The parameter $s$ is called the *seed length,* and sometimes we call the family $\mathcal{F}$ a *class of tests.* The construction of unconditional pseudorandom generators that fool restricted classes of tests is a fundamental research direction, as the generators constructed often have disparate applications in theoretical computer science. In this article we obtain new generators for several classes of tests. We start with the simplest one.

## 1.1 History and results

### 1.1.1 Fooling read-once polynomials

Pseudorandom generators for *polynomials* have been studied since at least the 1993 work by Luby, Veličković, and Wigderson [27], who gave a generator for GF(2) polynomials consisting of $s$ monomials with error $\varepsilon$ and seed length $2^{O(\sqrt{\log(s/\varepsilon)})}$. See [38] for an alternative proof. Servedio and Tan [34] recently improved the seed length to $2^{O(\sqrt{\log s})} \cdot \log(1/\varepsilon)$, and any significant improvement on the seed length would require a breakthrough on circuit lower bounds. For low-degree polynomials, better generators are known [6, 25, 39]. In this work we consider *read-once* polynomials on $m$ variables, which are a sum of monomials on disjoint sets of variables. For this class, a generator with seed length polylogarithmic in $m$ and $1/\varepsilon$ is given in [14] and it applies more generally to read-once $\mathrm{ACC}^0$. We obtain a seed length which is optimal up to a factor of $\log(1/\varepsilon) \cdot \operatorname{poly}\log\log(m/\varepsilon)$. In particular, when $\varepsilon$ is not too small, our generator has seed length optimal up to $\operatorname{poly}\log\log m$. Throughout this paper the notation $\tilde{O}(t)$ stands for $t \cdot \operatorname{polylog}(t)$.

**Theorem 1.3.** *There exists an explicit generator $G\colon \{0,1\}^s \to \{0,1\}^m$ that fools any read-once GF(2) polynomial with error $\varepsilon$ and seed length $s = \tilde{O}(\log(m/\varepsilon)\log(1/\varepsilon))$.*

A specific motivation for studying read-once polynomials comes from derandomizing space-bounded algorithms, a major line of research in pseudorandomness whose leading goal is proving RL = L. It is known that $\mathrm{RL} \subseteq \mathrm{DSPACE}((\log n)^{3/2})$ [33]. However, despite a lot of effort, in terms of PRGs for general space-bounded algorithms, there has been no improvement over the seed length $\geq \log^2 m$ since the classic 1992 paper by Nisan [31]. In fact, no improvement is known even under the restriction that the algorithm uses constant space. Read-once polynomials can be implemented by constant-space algorithms, and were specifically pointed out by several authors as a bottleneck for progress on space-bounded algorithms, see for example this survey talk by Trevisan [36]. Thus our work can be seen as progress towards derandomizing small-space algorithms. We note that the concurrent work of Chattopadhyay, Hatami, Reingold and Tal [9] gives a generator for space-bounded algorithms which implies a generator for polynomials with seed length $\tilde{O}(\log^3 m \log^2(m/\varepsilon))$.

Theorem 1.3 also holds for polynomials modulo $M$ for any fixed $M$; in fact we obtain it as an easy corollary of a more general generator (Theorem 1.7).

### 1.1.2 Fooling products

We consider tests on $m$ bits that can be written as the product of $k$ bounded functions on disjoint inputs of $\ell$ bits each. Such tests generalize the well-studied *combinatorial rectangles* [1, 31, 32, 22, 12, 3, 26, 40, 16, 19] as well as other classes of tests, see [15]. They were introduced in [15] by Gopalan, Kane, and Meka who call them *Fourier shapes*. However, in their definition the partition of the $m$-bit input into the $k$ blocks of $\ell$-bit inputs to the functions is fixed and known to the generator. Following a recent push for breaking the mold of "fixed-order" tests, we consider such tests under arbitrary order. We call them *product tests* and define them formally next.

**Definition 1.4** (Product tests)**.** A function $f\colon \{0,1\}^m \to \mathbb{C}_{\leq 1}$ is an $m$-bit *product test* with $k$ functions of input length $\ell$ if there exist $k$ disjoint subsets $I_1, I_2, \ldots, I_k \subseteq \{1,2,\ldots,m\}$ of size $\leq \ell$ each such that $f(x) = \prod_{i \leq k} f_i(x_{I_i})$ for some functions $f_i$ with range in $\mathbb{C}_{\leq 1}$. Here $\mathbb{C}_{\leq 1}$ is the complex unit disk $\{z \in \mathbb{C} : |z| \leq 1\}$, and $x_{I_i}$ is the string of $|I_i|$ bits of $x$ indexed by $I_i$.

Handling arbitrary order is significantly more challenging, because the classical space-bounded generators [31, 22] only work in fixed order [37, 5]. Our previous work with Haramaty [20] gave the first generators for this class, but its dependence on $k$ is poor: the seed length is always $\geq \sqrt{k}$. In this paper we improve the dependence on $k$ exponentially, though the results in [20] remain unsurpassed when $k$ is very small, e. g., $k = O(1)$. We actually obtain two incomparable generators.

**Theorem 1.5.** *There exists an explicit generator $G\colon \{0,1\}^s \to \{0,1\}^m$ that fools any $m$-bit product test with $k$ functions of input length $\ell$ with error $\varepsilon$ and seed length $s = \tilde{O}\big((\ell + \log k)(\log k)\log(1/\varepsilon) + \log\log m\big)$.*

Using the probabilistic method, one can show that there exist (non-explicit) generators with seed length $s = O(\ell + \log k + \log(1/\varepsilon) + \log\log m)$.

By the reductions in [15], we also obtain generators that fool variants of product tests where the outputs of the $f_i$ are not simply multiplied but combined in other ways. These variants include generalized halfspaces [18] and combinatorial shapes [17, 10], extended to arbitrary order. For those we obtain seed length $\tilde{O}(\ell + \log k)^2 \log(1/\varepsilon)\log k$, whereas the previous best was $\geq \ell\sqrt{k}$ [20]. As this application amounts to plugging the above theorem into previous reductions, we don't discuss it further in this paper and instead refer the reader to Section 6 in [20].

We then give another generator which has a seed length that is optimal up to a factor $\log(1/\varepsilon)\cdot\operatorname{polylog}(\ell,\log k,\log m,\log(1/\varepsilon))$, just like Theorem 1.3. However, for this we need each function $f_i$ in the definition of product tests to have expectation at most $1-\Omega(2^{-\ell})$. This condition is satisfied by Boolean and most natural functions. For simplicity one can think of the functions $f_i$ having outputs $\{-1,1\}$.

**Definition 1.6** (Nice product tests)**.** A product test as in Definition 1.4 is *nice* if each function $f_i$ has expectation at most $1-2^{-\ell}/4$.

Here, the constant 4 is chosen for ease of presentation and can be replaced by an arbitrary constant.

**Theorem 1.7.** *There exists an explicit generator $G\colon \{0,1\}^s \to \{0,1\}^m$ that fools any nice $m$-bit product test with $k$ functions of input length $\ell$ with error $\varepsilon$ and seed length $\tilde{O}\big((\log(k/\varepsilon)+\ell)(\log(1/\varepsilon)+\log\log m)\big)$.*

This is the result from which the generator for polynomials in Theorem 1.3 follows easily.

### 1.1.3 Bounded independence plus noise

The framework in which we develop these generators was first laid out by Ajtai and Wigderson in their pioneering paper [2] where they constructed generators for $\mathrm{AC}^0$ with polynomial seed length. The framework seems to have been forgotten for a while, possibly due to the spectacular successes by Nisan who gave better and arguably simpler generators [30, 31]. The interest in the Ajtai–Wigderson generators has recently been revived in a series of papers starting with the influential paper [16] by Gopalan, Meka, Reingold, Trevisan, and Vadhan who use it to obtain a generator for read-once CNF on $m$ bits with error $\varepsilon$ and seed length $\tilde{O}(\log(m/\varepsilon))$. This significantly improves the previously available seed length of $O(\log m)\log(1/\varepsilon)$ [8, 11] when $\varepsilon$ is small.

The Ajtai–Wigderson framework goes by showing that the test is fooled by a distribution with bounded independence [29] ($d$-wise independence for some small value of $d$), *if we perturb it with noise*. (Previous papers use the equivalent language of *restrictions*; we instead follow [20].) Then the high-level idea is to recurse on the function restricted to the positions perturbed by the noise. This has to be coupled with a separate, sometimes technical argument showing that each recursive step simplifies the test. We shall explain this in Section 1.2. Thus our goal is to understand if bounded independence plus noise fools product tests.

We say that $X$ is an *$n$-bit random variable* if it is a random variable that takes its values in $\{0,1\}^n$. We say that $U$ is a *uniformly distributed $n$-bit random variable* if $U$ is uniformly distributed over $\{0,1\}^n$.

We perform bitwise operations of $m$-bit random variables. Specifically, for $m$-bit random variables $X=(X_1,\dots,X_m)$ and $Y=(Y_1,\dots,Y_m)$, we write $X+Y=(X_1+Y_1,\dots,X_m+Y_m)$ (coordinatewise XOR) and $X\wedge Y=(X_1Y_1,\dots,X_mY_m)$.

Let $D$ and $T$ be $m$-bit random variables and $U$ a uniformly distributed $m$-bit variable. We write the perturbed version of $D=(D_1,\dots,D_m)$ as $D+(T\wedge U)$. So if $T=(T_1,\dots,T_m)$ and $T_i=1$ then we replace $D_i$ by a uniform random bit, and all these bits are independent. We refer to $T\wedge U$ as the *noise vector*. For the application it is important that $T$ be selected pseudorandomly, though the result is interesting even if $T$ is uniform in $\{0,1\}^m$. We now state the result from [20] after defining bounded near-independence.

**Definition 1.8** (Total variation distance of $m$-bit random variables)**.** Let $X = (X_1,\dots,X_m)$ and $Y = (Y_1,\dots,Y_m)$ be $m$-bit random variables. The total variation distance of $X$ and $Y$ is defined as

$$\delta_{\mathrm{TV}}(X,Y) = \max_{S\subseteq\{0,1\}^m} |\Pr(X \in S) - \Pr(Y \in S)|.$$

**Definition 1.9** ($(\delta,d)$-closeness)**.** The $m$-bit random variable $X = (X_1,\dots,X_m)$ is $(\delta,d)$-*close* to the $m$-bit random variable $Y = (Y_1,\dots,Y_m)$ if for every $i_1,\dots,i_d \in \{1,2,\dots,m\}$ the the $d$-bit random variables $(X_{i_1},\dots,X_{i_d})$ and $(Y_{i_1},\dots,Y_{i_d})$ have total variation distance at most $\delta$.

Note that the variables $X_i$ are $d$-wise independent exactly if $(X_1,\dots,X_m)$ is $(0,d)$-close to the uniform distribution.

**Theorem 1.10** ([20])**.** *Let $f\colon \{0,1\}^m \to \mathbb{C}_{\le 1}$ be an $m$-bit product test with $k$ functions of input length $\ell$. Let $D,T$ and $U$ be three independent $m$-bit random variables, where $D$ and $T$ are $(0,d\ell)$-close to uniform, and $U$ is uniform. Then*

$$\left|\mathbb{E}[f(D+T\wedge U)] - \mathbb{E}[f(U)]\right| \le k2^{-\Omega(d^2\ell/k)}.$$

Note that the dependence on $k$, the number of functions, is poor: when $k = \Omega(d^2\ell)$, the error bound does not give anything non-trivial. One of our main technical contributions is obtaining exponentially better dependence on $k$ using different techniques from [20]. Our theorem gives a non-trivial error bound even when $d = O(1)$ and $k$ is exponential in $\ell$.

**Theorem 1.11.** *Let $f\colon \{0,1\}^m \to \mathbb{C}_{\le 1}$ be an $m$-bit product test with $k$ functions of input length $\ell$. Let $D,T$ and $U$ be three independent $m$-bit random variables, where $D$ and $T$ are $(\delta,d\ell)$-close to uniform, and $U$ is uniform. Then*

$$\left|\mathop{\mathbb{E}}_{D,T,U}[f(D+T\wedge U)] - \mathop{\mathbb{E}}_{U}[f(U)]\right| \le 2^{-\Omega(d)} + B\delta,$$

*for the following choices of $B$:*

*i.* $B = (k2^\ell)^{O(d)}$*;*

*ii. if $f$ is nice, then* $B = (d2^\ell)^{O(d)}$.

Setting $\delta = 0$, Theorem 1.11 has a better bound than Theorem 1.10 when $k = \Omega(d\ell)$. An interesting feature of Theorem 1.11 is that for nice products the parameter $\delta$ can be independent of $k$. We complement this feature with a negative result showing that for general products a dependence on $k$ is necessary. Thus, the distinction between products and nice products is not an artifact of our proof but is inherent.

**Claim 1.12.** *For every sufficiently large $k$, there exists a random variable $D$ over $\{0,1\}^k$ that is $(k^{-\Omega(1)},k^{\Omega(1)})$-close to uniform, and a $k$-bit product test $f\colon \{0,1\}^k \to \mathbb{C}_{\le 1}$ with $k$ functions of input length $1$ such that*

$$\left|\mathbb{E}[f(D+T\wedge U)] - \mathbb{E}[f(U)]\right| \ge 1/100,$$

*where $D,T$ and $U$ are independent, and $D$ and $T$ are uniform over $\{0,1\}^k$.*

This claim also shows that for $\ell = 1$ and $\varepsilon = \Omega(1)$ one needs $\delta \leq k^{-\Omega(1)}$, and even for random variables which are $(\delta, k^{\Omega(1)})$-close to uniform, instead of just $(\delta, O(1))$-close.

For the class of combinatorial rectangles, which corresponds to product tests with each $f_i$ outputting values in $\{0,1\}$, the classic result [12] (extended in [8], for an exposition see Lecture 1 in [41]) shows that $d\ell$-wise independence alone fools rectangles with error $2^{-\Omega(d)}$ and this error bound is tight. So Theorem 1.11 does not give better bounds for rectangles, even with the presence of noise. We develop additional machinery and obtain an improvement on Theorem 1.11. While the improvement is modest, the machinery we develop may be useful for further improvements. Since this improvement is not used in our construction of PRGs, we only state and prove it for exact bounded independence. For technical reasons we restrict the range of the $f_i$.

**Theorem 1.13.** *Let $f$ be an $m$-bit product test with $k$ functions of input length $\ell$. Suppose the range of each function $f_i$ of $f$ is the set $\{0,1\}$, or the set of all $M$-th roots of unity for some fixed $M$. Let $D, T$ and $U$ be three independent $m$-bit random variables, where $D$ and $T$ are $d\ell$-wise independent, and $U$ is uniform. Then*

$$\left|\mathbb{E}[f(D + T \wedge U)] - \mathbb{E}[f(U)]\right| \leq \ell^{-\Omega(d)}.$$

Finally, it is natural to ask if similar techniques fool non-read-once polynomials. In this regard, we are able to show that small-bias distributions [29] plus noise fool $\mathbb{F}_2$-polynomials of degree 2.

**Claim 1.14.** *Let $p\colon \{0,1\}^m \to \{0,1\}$ be any $\mathbb{F}_2$-polynomial of degree 2. Let $D, T = (T_1, \ldots, T_m)$ and $U$ be three independent $m$-bit random variables, where $D$ is $\delta$-biased, $T_1, \ldots, T_m$ are independent and $\mathbb{E}[T_i] = 2/3$ for every $i$, and $U$ is uniform. Then*

$$\left|\mathbb{E}[p(D + T \wedge U)] - \mathbb{E}[p(U)]\right| \leq \delta.$$

### 1.1.4 Subsequent developments

This paper is one of a series of recent articles that construct new pseudorandom generators. One celebrated goal is to construct better generators for read-once bounded-width branching programs. As previously mentioned, the special case of read-once polynomials was an obstacle to achieving this goal that was noted by several researchers including Trevisan [36] and Vadhan (personal communication). This paper removes that obstacle. Building on this paper, subsequent work made progress on width-3 branching programs [28]. As of February 2019, the main generators in this paper are subsumed by subsequent work [28, 23], but in the case of products of complex-valued functions, the seed length in this paper remains unsurpassed, and we believe the techniques here may find further applications.

## 1.2 Techniques

We first explain how to prove that bounded independence plus noise fools product tests (Theorem 1.11). After that, we will explain the additional ideas that go into constructing our PRGs.

Following the literature [16, 17, 19], at a high level we do a case analysis based on the *total variance* of the product test $f$ we want to fool. This is defined as the sum of the variances $\mathrm{Var}[f_i]$ of the functions $f_i$ in the definition of product test. The variance of a function $g$ is $\mathbb{E}[|g(x)|^2] - |\mathbb{E}[g(x)]|^2$ where $x$ is uniform.

### 1.2.1 Low total variance

Our starting point is an important inequality by Gopalan, Kane and Meka [15] (cf. [16, 19]), who showed that bounded independence alone without noise fools low total-variance product tests. However, their result is only proved for *exact* bounded independence, namely, every $d$ bits are exactly uniform, whereas it is critical for our seed lengths to handle bounded *near-independence*, i. e., every $d$ bits are close to uniform.

One of the technical contributions of this paper is extending the inequality in [15] to work for bounded near-independence. The proof of the inequality in [15] is somewhat technical, and our extension introduces several complications. For example, the expectations of the $f_i$ under a bounded near-independent distribution and the uniform distribution are not guaranteed to be equal, and this requires additional arguments. However our proof follows the argument in [15], which we also present in a slightly different way that is possibly of interest to some readers. Finally we mention that Claim 1.12 shows that our error term is close to tight in certain regimes, cf. Section 7.

### 1.2.2 High total variance

Here we take a different approach from the ones in the literature: The papers [14, 15] essentially reduce the case of high total variance to the case of low total variance. However their techniques either blow up the seed length polynomially [14] or rely on space-bounded generators that only work in fixed order [15].

We instead observe that bounded independence plus noise fools even high total-variance product tests. We now give some details of our approach. A standard fact is that the expectation of a product test $f$ is bounded above by

$$\prod_i |\mathbb{E}[f_i]| \leq \prod_i (1-\mathrm{Var}[f_i])^{1/2} \leq \mathrm{e}^{-\sum_i \mathrm{Var}[f_i]/2}.$$

So if the total variance $\sum_i \mathrm{Var}[f_i]$ is large then the expectation of the product test under the uniform distribution is small. Thus, it suffices to show that the expectation is also small under bounded independence plus noise. To show this, we argue that typically, the total variance remains high even considering the $f_i$ as functions of the noise only. Specifically, we first show that on average over uniformly distributed $x$ and $t$, the variance of the functions $f_i'(y) := f_i(x+t \wedge y)$ is about as large as that of the $f_i$. This uses Fourier analysis. Then we use concentration inequalities for bounded near-independent random variables to derandomize this fact: we show that it also holds for typical $x$ and $t$ sampled from $D$ and $T$.

This suffices to prove Theorem 1.11.i. Proving Theorem 1.11.ii requires additional ideas.

We first note that the case of high total variance actually does not appear in the read-once CNF generator in [16]. This is because one can always *truncate* the CNF to have at most $2^w \log(1/\varepsilon)$ number of clauses of width $w$, which suffices to determine the expected value of the CNF up to an additive error of $\varepsilon$, and such a CNF has low total variance (for this one argues that noise helps reduce the variance a little.) To handle an arbitrary read-once CNF, [16] partition the clauses according to their width, and handle each partition separately.

However, one cannot truncate polynomials without noise. To see why, consider, for a simple example, the linear polynomial $x_1 + x_2 + \ldots + x_m$ (corresponding to a product test that computes the parity function). Here no strict subset of the monomials determines the expectation of the polynomial. Indeed, one can construct distributions which look random to $m-1$ monomials, but not to $m$.

### 1.2.3 Truncation using noise

Although we cannot truncate polynomials without noise, we show that something almost as good can still be done, and this idea is critical to obtaining our seed lengths. We show that the statistical closeness parameter in $D$ and $T$ can be selected *as if the polynomial was truncated*: it is independent from the number $k$ of functions. This is reflected in Theorem 1.11.ii, where $\delta$ is independent from $k$. The proof goes by showing that if the number $k$ of functions is much larger than $2^{3\ell}$ then noise alone will be enough to fool the test, regardless of anything else. This proof critically uses noise: without noise a dependence on $k$ is necessary, as shown in the parity example in our discussion. Also, for the proof to work the functions must have expectation at most $1-\Omega(2^{-\ell})$. As mentioned earlier, we further prove that this last requirement is necessary (Claim 1.12): we construct functions whose expectation is about $1-1/k$ but their product is not fooled by almost bounded independence plus noise, if the statistical closeness parameter is larger than $1/k^c$ for a suitable constant $c$.

### 1.2.4 Additional ideas for improved bound

To obtain the improved error bound in Theorem 1.13, we show that whenever the total variance of a product test lies below $dn^{0.1}$, we can use noise to bring it down below $d\ell^{-0.1}$. This produces a gap of $[d\ell^{-0.1}, d\ell^{0.1}]$ between cases of the high and low total variance, which gives the better bound using the previous arguments. Reducing the total variance requires a few additional ideas. First, we use Theorem 1.10 to handle the functions $f_i$ in the product test which have high variances. Then we use the hypercontractivity theorem to reduce the variances of the rest of the $f_i$ individually. [16] also uses noise to reduce variance, but the functions $f_i$ in [16] are just AND and so it does not need hypercontractivity. To combine both ideas, we prove a new "XOR Lemma" for bounded independence, a variant of an XOR lemma for small-bias, which was proved in [16].

### 1.2.5 Constructing our PRGs

We now explain how to use Theorem 1.11 to construct our PRGs. The high-level idea of our PRG construction is to apply Theorem 1.11 recursively following the Ajtai–Wigderson framework: Given $D+T\wedge U$, where $T=(T_1,\dots,T_m)$, we can think of each $T_i$ as an indicator random variable that selects the $i$-th position with probability $1/2$. For intuition, it would be helpful to assume each position is selected independently.

We will focus on how to construct a PRG using a seed of length $\tilde{O}(\log m)$ for read-once polynomials with constant error, as this simplifies the parameters and captures all the ideas. Without loss of generality, we can assume the degree of a polynomial to be $\ell=O(\log m)$, because the contribution of higher-degree terms can be shown to be negligible under a small-bias distribution. (See the proof of Theorem 1.3.)

Let $p\colon \{0,1\}^m \to \{0,1\}$ be a degree-$\ell$ read-once polynomial with $k$ monomials. It would be convenient to think of $p$ outputting values $\{-1,1\}$. Further, we can write $p$ as a product $\prod_{i=1}^k p_i$, where each $p_i$ is a monomial on at most $\ell$ bits (with outputs in $\{-1,1\}$.)

Now suppose we only assign the values in $D$ to the positions not chosen by $T$, that is, setting the input bits $x_i=D_i$ for $i\notin T$. This induces another polynomial $p_{D,T}$ defined on the positions in $T$. Clearly, $p_{D,T}$ also has degree at most $\ell$, and so we can reapply Theorem 1.11 to $p_{D,T}$.

Repeating the above argument $t$ times induces a polynomial defined on the positions $T_t := \bigwedge_{i=1}^{t} T_i$. One can think of $T_t = (T_{t1} \dots, T_{tm})$ as a single random variable, where each $T_i$ selects the $i$position with probability $2^{-t}$. Viewing $T_t$ this way, it is easy to see that we can terminate the recursion after $t := O(\log m)$ steps, as the set of positions selected by $T_t$ should become empty with high probability.

By standard constructions [29], it takes $s := \tilde{O}(\ell)$ bits to sample $D$ and $T$ in Theorem 1.11.ii each time. Therefore, we get a PRG of seed length $t \cdot s = \tilde{O}(\ell \log m)$.

To obtain a better seed length, we will instead apply Theorem 1.11 in *stages*. Our goal in each stage is to reduce the degree of the polynomial by half. In other words, we want the restricted polynomial defined on the positions in $\bigwedge_{i=1}^{t} T_i$ to have degree $\ell/2$. It is not difficult to see that in order to reduce the degree of the $m$ monomials of $p$ to $\ell/2$ with high probability, it suffices to apply our above argument recursively for $t := O(\log m)/\ell$ times. So in each stage, we use a seed of length

$$t \cdot s = \tilde{O}(\ell) \cdot \left(\frac{\log m}{\ell}\right) = \tilde{O}(\log m).$$

After repeating the same argument for $O(\log \ell) = O(\log\log m)$ stages, with high probability the restricted polynomial would have degree 0 and we are done. Therefore, the total seed length of our PRG is $\tilde{O}(\log m)$.

Here we remark that it is crucial in our argument that $D$ and $T$ are bounded near-independent, as opposed to being small-biased. Otherwise, we cannot have seed length $s = \tilde{O}(\ell)$ when $\ell = o(\log m)$. For example, when $\ell = O(1)$, with small-bias we would need $s = O(\log m)$ bits, whereas we just use $O(\log\log m)$ bits.

Motivated by the analysis in [20], Forbes and Kelley [13] show that $2t$-wise independence plus noise fools width-$w$ ROBPs on $m$ bits with error $2^{-t/2}mw$. Their work implicitly shows that $t$-wise independence plus noise fools product tests with $k$ functions of input length $\ell$ with error $k2^{-\Omega(t)+\ell-1}$, improving [20]. However, their result is incomparable to Theorems 1.11 and 1.13, as there is no dependence on $k$ in our error bounds for exact bounded independence, i. e., when $D$ is $(0, d\ell)$-close to uniform. By combining their result with the observation that noise alone fools nice product tests when $k$, the number of functions, is much larger than $2^{3\ell}$, we show that the dependence on $k$ in their error bound can be removed for nice product tests.

**Theorem 1.15.** *Let $f\colon \{0,1\}^m \to \mathbb{C}_{\le 1}$ be a nice product test with $k$ functions of input length $\ell$. Let $D, T$ and $U$ be three $m$-bit random variables, where $D$ and $T$ are $t$-wise independent and $U$ is uniform. Then*

$$\left| \mathop{\mathbb{E}}_{D,T,U}[f(D + T \wedge U)] - \mathop{\mathbb{E}}_{U}[f(U)] \right| \le 2^{8\ell - \Omega(t)}.$$

We note that for product tests this error bound is optimal up to the constant in the exponent, because the same distribution fools parities with error $2^{-(t+1)}$. On the other hand, [7, Theorem 8] shows that for ROBPs the dependence on $m$ in the error is inherent.

**Organization.** We prove bounded independence plus noise fools product (Theorem 1.11) in Section 2, except the proof of the case of low total variance, which we defer to Section 4. Then we give constructions of our PRGs in Section 3. In Section 5, we show how to obtain the modest improvement of Theorem 1.11 and the optimal error bound for nice product tests (Theorem 1.15) using [13]. After that, we prove our result on fooling degree-2 polynomials in Section 6. Finally, we prove Claim 1.12 in Section 7.

| | Conditions | Uses | Follows from | Error |
|---|---|---|---|---|
| (1) | $\sum_{i\le k}\mathrm{Var}[f_i]\le\alpha d$ | $D$ | Lemma 2.1 | $2^{-\Omega(d)}+(k2^{\ell})^{O(d)}\delta$ |
| (2) | $\sum_{i\le k}\mathrm{Var}[f_i]\ge\alpha d$ | $D+T\wedge U$ | Derandomized Claim 2.2 | $2^{-\Omega(d)}+k^{O(d)}\delta$ |
| (3) | $k\ge 2^{3\ell+1}d$, nice products | $T\wedge U$ | Claim 2.4 | $2^{-\Omega(d\ell)}+2^{O(d\ell)}\delta$ |

Table 1: Error bounds for fooling a product tests of $k$ functions of input length $\ell$ under different conditions. Here $D,T$ and $U$ are independent, where $D$ and $T$ are $(\delta,d\ell)$-close to uniform, $U$ is uniform, and $\alpha$ is a small constant.

## 2 Bounded independence plus noise fools products

In this section we prove Theorem 1.11. As we mentioned in the introduction, the proof consists of 3 parts: (1) Low total variance, (2) high total variance, and (3) truncation using noise for nice products. We summarize the conditions and the error bounds we obtain for these cases in Table 1. Let us now quickly explain how to put them together to prove Theorem 1.11. Clearly, combining (1) and (2) immediately gives us a bound of $2^{-\Omega(d)}+(k2^{\ell})^{O(d)}$ for product tests, proving Theorem 1.11.i. For nice product tests, we can apply (3) if $k\ge 2^{3\ell+1}d$, otherwise we can plug in $k\le 2^{3\ell+1}d$ in the previous bound, proving Theorem 1.11.ii.

We now discuss each of the 3 cases in order. Since the proof of the case of low total variance is quite involved, we only state the lemma in this section and defer its proof to Section 4.

**Lemma 2.1.** *Let $X_1,X_2,\dots,X_k$ be $k$ independent random variables over $\mathbb{C}_{\le 1}$ such that for every $i\in\{1,\dots,k\}$ and $z_i\in\mathrm{Supp}(X_i)$ we have $\Pr[X_i=z_i]\ge 2^{-\ell}$. Let $Y_1,Y_2,\dots,Y_k$ be $k$ random variables over $\mathbb{C}_{\le 1}$ that are $(\varepsilon,16d)$-close to $X_1,\dots,X_k$. For every $\sigma\ge\sqrt{\sum_{i=1}^k\mathrm{Var}[X_i]}$,*

$$\left|\mathbb{E}\left[\prod_{i=1}^k Y_i\right]-\mathbb{E}\left[\prod_{i=1}^k X_i\right]\right|\le 2^{O(d)}\left(\frac{\sigma^2}{d}\right)^{d/2}+(k2^{\ell})^{O(d)}\varepsilon.$$

We now prove a claim that handles the case of high total variance. This claim shows that for uniformly distributed $x$ and $t$, the variance of the function $g(y):=f(x+t\wedge y)$ is close to the variance of $f$ in expectation. Its proof follows from a simple calculation in Fourier analysis. Later, we will derandomize this claim in the proof of Theorem 1.11.

**Claim 2.2.** *Let $T=(T_1,\dots,T_\ell)$ be a $\ell$-bit random variable where the $T_j$'s are independent and $\mathbb{E}[T_j]=\eta$ for each $j$. Let $f\colon\{0,1\}^{\ell}\to\mathbb{C}$ be any function. Then*

$$\mathop{\mathbb{E}}_{U,T}\left[\mathop{\mathrm{Var}}_{U'}[f(U+T\wedge U')]\right]\ge\eta\,\mathrm{Var}[f],$$

*where $T,U$ and $U'$ are independent, and $U$ and $U'$ are uniform.*

*Proof of Claim 2.2.* By the definition of variance and linearity of expectation, we have

$$\begin{aligned}\mathop{\mathbb{E}}_{U,T}\Big[\mathop{\mathrm{Var}}_{U'}[f(U+T\wedge U')]\Big] &= \mathop{\mathbb{E}}_{U,T}\Big[\mathop{\mathbb{E}}_{U'}\big[|f(U+T\wedge U')|^2\big]-\Big|\mathop{\mathbb{E}}_{U'}[f(U+T\wedge U')]\Big|^2\Big]\\ &= \mathop{\mathbb{E}}_{U,T}\Big[\mathop{\mathbb{E}}_{U'}\big[\big|f(U+T\wedge U')\big|^2\big]\Big]-\mathop{\mathbb{E}}_{U,T}\Big[\Big|\mathop{\mathbb{E}}_{U'}[f(U+T\wedge U')]\Big|^2\Big].\end{aligned}$$

The first term is equal to

$$\mathop{\mathbb{E}}_{U}[|f(U)|^2]=\sum_{\alpha,\alpha'}\hat{f}_\alpha\overline{\hat{f}_{\alpha'}}\mathop{\mathbb{E}}_{U}[\chi_{\alpha-\alpha'}(U)]=\sum_\alpha|\hat{f}_\alpha|^2.$$

The second term is equal to

$$\begin{aligned}&\mathop{\mathbb{E}}_{U,T}\Big[\mathop{\mathbb{E}}_{U'}\Big[\sum_\alpha\hat{f}_\alpha\chi_\alpha(U+T\wedge U')\Big]\overline{\mathop{\mathbb{E}}_{U''}\Big[\sum_{\alpha'}\hat{f}_{\alpha'}\chi_{\alpha'}(U+T\wedge U'')\Big]}\Big]\\ &=\mathop{\mathbb{E}}_{U,T}\Big[\sum_{\alpha,\alpha'}\hat{f}_\alpha\overline{\hat{f}_{\alpha'}}\mathop{\mathbb{E}}_{U'}[\chi_\alpha(U+T\wedge U')]\mathop{\mathbb{E}}_{U''}[\chi_{\alpha'}(U+T\wedge U'')]\Big]\\ &=\sum_{\alpha,\alpha'}\hat{f}_\alpha\overline{\hat{f}_{\alpha'}}\mathop{\mathbb{E}}_{U}[\chi_{\alpha+\alpha'}(U)]\mathop{\mathbb{E}}_{T}\Big[\mathop{\mathbb{E}}_{U'}[\chi_\alpha(T\wedge U')]\mathop{\mathbb{E}}_{U''}[\chi_{\alpha'}(T\wedge U'')]\Big]\\ &=\sum_\alpha|\hat{f}_\alpha|^2\mathop{\mathbb{E}}_{T,U',U''}[\chi_\alpha(T\wedge(U'+U''))]\\ &=\sum_\alpha|\hat{f}_\alpha|^2(1-\eta)^{|\alpha|}.\end{aligned}$$

Therefore,

$$\mathop{\mathbb{E}}_{U,T}\Big[\mathop{\mathrm{Var}}_{U'}[f(U+T\wedge U')]\Big]=\sum_\alpha|\hat{f}_\alpha|^2\Big(1-(1-\eta)^{|\alpha|}\Big)\geq\eta\sum_{\alpha\neq0}|\hat{f}_\alpha|^2=\eta\,\mathrm{Var}[f],$$

where the inequality is because $1-(1-\eta)^{|\alpha|}\geq 1-(1-\eta)\geq\eta$ for any $\alpha\neq 0$. □

With Lemma 2.1 and Claim 2.2, we now prove Theorem 1.11.

*Proof of Theorem 1.11.i.* Let $\sigma$ be exactly $(\sum_{i\leq k}\mathrm{Var}[f_i])^{1/2}$. We will consider two cases: $\sigma^2\leq\alpha d$ and $\sigma^2>\alpha d$, where $\alpha>0$ is a sufficiently small constant.

If $\sigma^2\leq\alpha d$, we use Lemma 2.1. Specifically, since $\Pr[f_i(U)=z_i]\geq 2^{-\ell}$ for every $i$ and $z_i\in\mathrm{Supp}(f_i)$, it follows from Lemma 2.1 that

$$\Big|\mathbb{E}\Big[\prod_{i=1}^k f_i(D)\Big]-\mathbb{E}\Big[\prod_{i=1}^k f_i(U)\Big]\Big|\leq 2^{-\Omega(d)}+(k2^\ell)^{O(d)}\delta,$$

and the desired bound holds for every fixing of $T$ and $U$.

If $\sigma^2\geq\alpha d$, then the expectation of $f$ under the uniform distribution is small. More precisely, we have

$$\Big|\prod_{i\leq k}\mathop{\mathbb{E}}_{U}[f_i(U)]\Big|\leq\prod_{i\leq k}(1-\mathrm{Var}[f_i])^{1/2}\leq \mathrm{e}^{-\frac{1}{2}\sigma^2}\leq 2^{-\Omega(d)}. \tag{2.1}$$

Thus, it suffices to show that its expectation under $D+T\wedge U$ is at most $2^{-\Omega(d)}+(k2^{\ell})^{O(d)}\delta$. We now use Claim 2.2 to show that

$$\Big|\mathop{\mathbb{E}}_{D,T,U}\Big[\prod_{i=1}^{k} f_i(D+T\wedge U)\Big]\Big| \le 2^{-\Omega(d)}+k^{O(d)}\delta.$$

For each $t,x\in\{0,1\}^m$, and each $i\in\{1,2,\ldots,k\}$, let $\sigma^2_{t,x,i}$ denote $\mathrm{Var}_{U'}[f_i(x+t\wedge U')]$. We claim that $\sum_{i\le k}\sigma^2_{t,x,i}$ is large for most $x$ and $t$ sampled from $D$ and $T$ respectively. From Claim 2.2 we know that this quantity is large in expectation for uniform $x$ and $t$. By a tail bound for bounded near-independent random variables, we show that the same is true for most $x\in D$ and $t\in T$. By a similar calculation to (2.1) we show that for these $x$ and $t$ we have that $|\mathbb{E}[f(x+t\wedge U)]|$ is small.

To proceed, let $T'$ be a random variable distributed uniformly over $\{0,1\}^m$. Applying Claim 2.2 with $\eta=1/2$, we have $\mathbb{E}_{T',U}[\sigma^2_{T',U,i}]\ge \mathrm{Var}[f_i]/2$. So by linearity of expectation,

$$\mathop{\mathbb{E}}_{T',U}\Big[\sum_{i\le k}\sigma^2_{T',U,i}\Big]\ge \sigma^2/2\ge \alpha d/2.$$

Since $T$ and $D$ are both $(\delta,d\ell)$-close to uniform, the random variables $\sigma^2_{T,D,1},\ldots,\sigma^2_{T,D,k}$ are $(2\delta,d)$-close to $\sigma^2_{T',U,1},\ldots,\sigma^2_{T',U,k}$. We now use the following tail bound on the $\sigma^2_{T,D,i}$. Its proof can be found in Section 8.

**Lemma 2.3.** *Let $X_1,X_2,\ldots,X_k\in[0,1]$ be independent random variables. Let $d$ be an even positive integer. Let $Y_1,Y_2,\ldots,Y_k\in[0,1]$ be random variables that are $(\varepsilon,d)$-close to $X_1,\ldots,X_k$. Let $Y=\sum_{i\le k}Y_i$ and $\mu=\mathbb{E}[\sum_i X_i]$. Then,*

$$\Pr[|Y-\mu|\ge\delta\mu]\le 4\cdot 2^d\left(\frac{\sqrt{\mu d}+d}{\delta\mu}\right)^d+4\left(\frac{2k}{\delta\mu}\right)^d\varepsilon.$$

*In particular, if $\mu\ge 25d$, we have* $\Pr[|Y-\mu|\ge\mu/2]\le 2^{-\Omega(d)}+k^d\varepsilon$.

Let $\mu$ be $\mathbb{E}_{T',U}[\sum_{i\le k}\sigma^2_{T',U,i}]\ge\alpha d/2$. By Lemma 2.3,

$$\Pr_{T',U}\Big[\sum_{i\le k}\sigma^2_{T',U,i}\le\mu/2\Big]\le 2^{-\Omega(d)}+k^{O(d)}\delta. \tag{2.2}$$

Hence, except with probability $2^{-\Omega(d)}+k^{O(d)}\delta$ over $t\in T$ and $x\in D$, we have

$$\sum_{i\le k}\sigma^2_{t,x,i}=\sum_{i\le k}\mathop{\mathrm{Var}}_{U'}[f_i(x+t\wedge U')]\ge\alpha d/4.$$

For every such $t$ and $x$, we have

$$\begin{aligned}\Big|\prod_{i\le k}\mathop{\mathbb{E}}_{U}[f_i(x+t\wedge U)]\Big| &\le \prod_{i\le k}\Big|\mathop{\mathbb{E}}_{U}[f_i(x+t\wedge U)]\Big|\\ &\le\prod_{i\le k}(1-\sigma^2_{t,x,i})^{1/2}\\ &\le \mathrm{e}^{-\frac12\sum_{i\le k}\sigma^2_{t,x,i}}\le 2^{-\Omega(d)}.\end{aligned} \tag{2.3}$$

In addition, we always have $|f| \le 1$. Hence, summing the right hand side of (2.2) and (2.3), we have

$$\left|\mathop{\mathbb{E}}_{D,T,U}\Big[\prod_{i\le k} f_i(D+T\wedge U)\Big]\right| \le \mathop{\mathbb{E}}_{D,T}\Big[\Big|\prod_{i\le k}\mathop{\mathbb{E}}_{U}[f_i(D+T\wedge U)]\Big|\Big] \le 2^{-\Omega(d)} + k^{O(d)}\delta. \qquad \square$$

To prove Theorem 1.11.ii, we use the following additional observation that noise alone fools nice products when $k$ is suitably larger than $2^{2\ell}$. The high-level idea is that in such a case there will be at least $k2^{-\ell} \ge 2^{\ell}$ functions $f_i$ whose inputs are completely set to uniform by the noise. Since the expectation of each $f_i$ is bounded by $1-2^{-\ell}/4$, the expectation of their product becomes small when $k$ is suitably larger than $2^{2\ell}$. On the other hand, $\mathbb{E}[f(U)]$ can only get smaller under the uniform distribution, and so the expectations under uniform and noise are both small.

**Claim 2.4** (Noise fools nice products with large $k$)**.** *Let $f\colon \{0,1\}^m \to \mathbb{C}_{\le 1}$ be a* nice *$m$-bit product test with $k \ge 2^{3\ell+1}d$ functions of input length $\ell$. Let $T,U$ be two independent $m$-bit random variables where $T$ is $(\delta, d\ell)$-close to uniform, and $U$ is uniform.*

*Then*

$$\left|\mathop{\mathbb{E}}_{T,U}[f(T\wedge U)] - \mathbb{E}[f(U)]\right| \le 2^{-\Omega(d\ell)} + 2^{O(d\ell)}\delta.$$

*Proof.* We will bound above both expectations in absolute value. Let $k' := 2^{3\ell+1}d \le k$. Write $f = \prod_{i=1}^{k} f_i$, where $f_i\colon \{0,1\}^{I_i} \to \mathbb{C}_{\le 1}$. Since $f$ is nice, we have $|\mathbb{E}[f_i(U)]| \le 1-2^{-\ell}/4$ for every $i \in \{1,\dots,k\}$. Under the uniform distribution, we have

$$\big|\mathbb{E}[f(U)]\big| = \prod_{i=1}^{k}\big|\mathbb{E}[f_i(U)]\big| \le (1-2^{-\ell}/4)^k \le \mathrm{e}^{-\Omega(k2^{-\ell})} \le 2^{-\Omega(d\ell)}. \tag{2.4}$$

It suffices to show that the expectation under $T\wedge U$ is at most $2^{-\Omega(d\ell)} + 2^{O(d\ell)}\delta$. Note that

$$\big|\mathbb{E}[f(T\wedge U)]\big| \le \mathop{\mathbb{E}}_{T}\Big[\prod_{i=1}^{k}\big|\mathop{\mathbb{E}}_{U}[f_i(T\wedge U)]\big|\Big] \le \mathop{\mathbb{E}}_{T}\Big[\prod_{i=1}^{k'}\big|\mathop{\mathbb{E}}_{U}[f_i(T\wedge U)]\big|\Big].$$

We now show that the right hand side is at most $2^{-\Omega(d\ell)} + 2^{O(d\ell)}\delta$. We first show that the expected number of $f_i$ whose inputs are all selected by $T$ when $T$ is uniform is large, and then apply Lemma 2.3 to show that it holds for most $t \in T$. Let $T'$ be a random variable distributed uniformly over $\{0,1\}^m$. Then

$$\mathbb{E}\Big[\sum_{i=1}^{k'} \mathbb{1}(T'_{I_i} = 1^{|I_i|})\Big] = \sum_{i=1}^{k'} \Pr[T'_{I_i} = 1^{|I_i|}] \ge k'2^{-\ell} = 2^{2\ell+1}d.$$

Since $T$ is $(\delta, d\ell)$-close to uniform, the $T_{I_i}$ are $(\delta, d)$-close to uniform. By Lemma 2.3,

$$\Pr_{T}\Big[\sum_{i=1}^{k'} \mathbb{1}(T_{I_i} = 1^{|I_i|}) \le 2^{2\ell}d\Big] \le 2^{-\Omega(d\ell)} + 2^{O(d\ell)}\delta. \tag{2.5}$$

Note that if $T_{I_i} = 1^{|I_i|}$, then $|\mathbb{E}_U[f_i(T \wedge U)]| = |\mathbb{E}[f_i]| \le 1 - 2^{-\ell}/4$. Thus, conditioned on $\sum_{i=1}^{k'} \mathbb{1}(T_{I_i} = 1^{|I_i|}) \ge 2^{2\ell} d$, we have

$$\prod_{i=1}^{k'} \big|\mathbb{E}[f_i(T \wedge U)]\big| \le (1 - 2^{-\ell}/4)^{2^{2\ell} d} \le 2^{-\Omega(d\ell)}. \tag{2.6}$$

Since we always have $|f| \le 1$, the error bound follows from summing the right hand side of (2.4), (2.5) and (2.6). □

Theorem 1.11.ii now follows easily from Claim 2.4 and Theorem 1.11.i.

*Proof of Theorem 1.11.ii.* Since $f$ is nice, $|\mathbb{E}[f_i]| \le 1 - 2^{-\ell}/4$. If $k \ge 2^{3\ell+1} d$, then the theorem follows from Claim 2.4. Otherwise, $k \le 2^{3\ell+1} d$ and the theorem follows from Theorem 1.11.i. □

## 3 Pseudorandom generators

In this section we construct our generators. As explained in the introduction, all constructions follow from applying the Theorem 1.11 recursively. We obtain our generator for arbitrary product tests (Theorem 1.5) by applying Theorem 1.11 for $O(\log \ell k)$ times recursively. Our progress measure for the recursion is the number of bits the restricted product test is defined on. We show that after $O(\log \ell k)$ steps of the recursion we are left with a product test that is defined on $m' := O(\ell \log(1/\varepsilon))$ bits, which can be fooled by a distribution that is $(\varepsilon, m')$-close to uniform. As a first read, we suggest the readers to refer to the $\tilde{O}$ notations in the statements and proofs, i. e., ignore polylogarithmic factors in $\ell$, $\log k$, $\log(1/\varepsilon)$ and $\log m$, and think of $k$ as $m$ and $\varepsilon$ as some small constant.

*Proof of Theorem 1.5.* Let $C$ be a sufficiently large constant. Let $t = C \log(\ell k)$, $d = C \log(t/\varepsilon)$ and $\delta = (k 2^{\ell})^{-Cd}$. Let $D_1, \ldots, D_t, T_1, \ldots, T_t$ and $G'$ be $2t+1$ independent $m$-bit random variables that are $(\delta, d\ell)$-close to uniform. Define $D^{(1)} := D_1$ and $D^{(i+1)} := D_{i+1} + T_i \wedge D^{(i)}$. Let $D := D^{(t)}$, $T := \bigwedge_{i=1}^{t} T_i$. Our generator $G$ outputs

$$D + T \wedge G'.$$

We first look at the seed length of $G$. By [29, Lemma 4.2], sampling $G'$ and each of the random variables $D_i$ and $T_i$ takes a seed of length

$$\begin{aligned} &O\big(d\ell + \log(1/\delta) + \log\log m\big) \\ &= O\big((\ell + \log k)\log(t/\varepsilon) + \log\log m\big) \\ &= \tilde{O}(\ell + \log k)\log(1/\varepsilon) + O(\log\log m). \end{aligned}$$

Hence the total seed length of $G$ is

$$(2t+1) \cdot \big(\tilde{O}(\ell + \log k)\log(1/\varepsilon) + O(\log\log m)\big) = \tilde{O}\big((\ell + \log k)(\log k)\log(1/\varepsilon) + \log\log m\big).$$

We now look at the error of $G$. By our choice of $\delta$ and applying Theorem 1.11.i recursively for $t$ times, we have

$$\left|\mathbb{E}[f(D+T\wedge U)]-\mathbb{E}[f(U)]\right| \le t\cdot\big(2^{-\Omega(d)}+(k2^{\ell})^{O(d)}\delta\big) \le \varepsilon/2.$$

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 defined on $d\ell$ bits, which can be fooled by $G'$.

Let $I=\bigcup_{i=1}^{k} I_i$. Note that $|I|\le \ell k$. Because the variables $T_i$ are independent and each of them is $(\delta, d\ell)$-close to uniform, we have

$$\Pr\big[|I\cap T|\ge d\ell\big] \le \binom{|I|}{d\ell}(2^{-d\ell}+\delta)^t \le 2^{d\ell\log(\ell k)}\cdot 2^{-\Omega(Cd\ell\log(\ell k))} \le \varepsilon/4.$$

It follows that for every fixing of $D$, with probability at least $1-\varepsilon/4$ over the choice of $T$, the function $f_{D,T}$ is a product test defined on at most $d\ell$ bits, which can be fooled by $G'$ with error $\varepsilon/4$. Hence $G$ fools $f$ with error $\varepsilon$. □

Our generator for nice product tests (Theorem 1.7) uses the maximum input length of the functions $f_i$ as the progress measure. We will use the following lemma, which captures the trade-off between the number of recursive steps and the simplification on a product test measured in terms of the maximum input length of the $f_i$.

**Lemma 3.1.** *Given an explicit generator $G'\colon \{0,1\}^{s'}\to\{0,1\}^m$ that fools nice $m$-bit product tests with $k$ functions of input length $r$ with error $\varepsilon'$ and seed length $s'$, one can construct another explicit generator $G\colon \{0,1\}^{s}\to\{0,1\}^m$ that fools nice $m$-bit product tests with $k$ functions of input length $\ell$ with error $\varepsilon'+t\varepsilon$, where*

$$t = O\left(\frac{\log(k/\varepsilon)}{r+1}+\log\left(\frac{\ell}{r+1}\right)\right),$$

*and seed length*

$$s = s'+t\cdot O((\ell+\log\log(1/\varepsilon))\log(1/\varepsilon)+\log\log m) = s'+t\cdot\tilde{O}\big((\ell\log(1/\varepsilon))+\log\log m\big).$$

We defer its proof to the end. Theorem 1.7 requires applying the lemma in stages, where in each stage we apply the lemma with a different value of $\ell$. XORing its output with a random variable sampled according to a small-bias distribution gives our generator for polynomials (Theorem 1.3).

We will apply Lemma 3.1 in $O(\log\ell)$ stages. In each stage our goal is to halve the input length of the product test.

*Proof of Theorem 1.7.* Let $f$ be a nice $m$-bit product test with $k$ functions of input length $\ell$. Note that by applying Lemma 3.1 with $r=\ell/2$ and error $\varepsilon/(t\log\ell)$, where $t=O(\log(k/\varepsilon)/\ell+1)$, we can halve its input length by incurring an error of $\varepsilon/O(\log\ell)$ and using a seed of length

$$\begin{aligned}&t\cdot O\big((\ell+\log\log((t\log\ell)/\varepsilon))\log((t\log\ell)/\varepsilon)+\log\log m\big)\\ &=\tilde{O}\big((\log(k/\varepsilon)+\ell)(\log(1/\varepsilon)+\log\log m)\big).\end{aligned}$$

Now we repeat the argument for $s=O(\log\ell)$ steps until the input length is zero, which is a constant function and can be fooled with zero error. So we have a generator that fools nice $m$-bit product tests with $k$ functions of input length $\ell$, with error $\varepsilon$ and seed length $s\cdot\tilde{O}\big((\log(k/\varepsilon)+\ell)(\log(1/\varepsilon)+\log\log m)\big) = \tilde{O}\big((\log(k/\varepsilon)+\ell)(\log(1/\varepsilon)+\log\log m)\big)$. □

Theorem 1.3 follows from XORing the output of the above generator with a small-bias distribution.

*Proof of Theorem 1.3.* Let $c$ be a sufficiently large constant. Let $D$ and $G$ be two independent $m$-bit random variables, where $D$ is sampled from a $(\varepsilon/m)^c$-biased distribution [29], and $G$ is sampled from the output distribution of the generator in Theorem 1.7 that fools $m$-bit product tests with $m$ functions and input length $c\log(m/\varepsilon)$ with error $\varepsilon/2$. The generator outputs $D+G$. By [29] and Theorem 1.7, it takes a seed of length

$$O(\log(m/\varepsilon)) + \tilde{O}\big(\log(m/\varepsilon) + c\log(m/\varepsilon)\big)\log(1/\varepsilon) = \tilde{O}(\log(m/\varepsilon))\log(1/\varepsilon).$$

Let $p\colon \{0,1\}^m \to \{-1,1\}$ be any read-once GF(2) polynomial. Consider the polynomial $p'$ obtained from $p$ by removing all the monomials with degree greater than $c\log(m/\varepsilon)$ in $p$. We claim that the expectation of $p$ and $p'$ under $D+G$ differs by at most $\varepsilon$. Note that for any random variable $X$ that is sampled according any $(\varepsilon/m)^c$-biased distribution $X$, the probability that any $c\log(m/\varepsilon)$ bits of $X$ are 1 is at most $\varepsilon/4m$, and so by a union bound we have $\Pr[p(X) \neq p'(X)] \leq \varepsilon/4$. In particular, this holds for $D+G$ and $U$. It follows that

$$\big|\mathbb{E}[p(D+G)] - \mathbb{E}[p(U)]\big| \leq \big|\mathbb{E}[p'(D+G)] - \mathbb{E}[p'(U)]\big| + \varepsilon/2 \leq \varepsilon,$$

where the last inequality holds for any fixed $D$ because of Theorem 1.7. □

We now prove Lemma 3.1. First we make an observation that will be used in the proof to reduce the input length of the product test.

**Claim 3.2.** *Let $T^{(1)},\dots,T^{(t)}$ be $t$ independent and identical random variables over $\{0,1\}^\ell$ that are $\delta$-close to uniform. Then* $\Pr[\mathrm{wt}(\bigwedge_{i=1}^t T^{(i)}) > r] \leq \binom{\ell}{r+1}(2^{-(r+1)}+\delta)^t$.

*Proof.* We have

$$\begin{aligned}
\Pr\Big[\mathrm{wt}\Big(\bigwedge_{i=1}^t T^{(i)}\Big) > r\Big] &\leq \sum_{S:|S|=r+1} \Pr\Big[\bigwedge_{i=1}^t \wedge_{j\in S}(T_j^{(i)} = 1)\Big] \\
&= \sum_{S:|S|=r+1} \prod_{i=1}^t \Pr\Big[\bigwedge_{j\in S}(T_j^{(i)} = 1)\Big] \\
&\leq \sum_{S:|S|=r+1} (2^{-(r+1)}+\delta)^t = \binom{\ell}{r+1}(2^{-(r+1)}+\delta)^t.
\end{aligned}$$

□

*Proof of Lemma 3.1.* Let $C$ be a sufficiently large constant. The generator $G$ will output $H^{(1)}$, where we define the distribution of the random variable $H^{(i)}$ recursively for

$$t = O\left(\frac{\log(k/\varepsilon)}{r+1} + \log\left(\frac{\ell}{r+1}\right)\right)$$

steps: At the $i$-th step, $H^{(i)}$ samples two independent $m$-bit random variables, $D^{(i)}, T^{(i)}$, that are $(\delta, C\ell\log(1/\varepsilon))$-close to uniform, where $\delta = 2^{-C(\ell+\log\log(1/\varepsilon))\log(1/\varepsilon)}$, and independent from $H^{(i+1)}$. Then output

$$H^{(i)} := D^{(i)} + T^{(i)} \wedge H^{(i+1)}.$$

We define $H^{(t+1)}$ to be $G'(U_{s'})$.

By [29, Lemma 4.2], sampling $D^{(i)}$ and $T^{(i)}$ takes a seed of length

$$\begin{aligned} u &:= O(\ell \log(1/\varepsilon) + \log(1/\delta) + \log\log m) \\ &= O((\ell + \log\log(1/\varepsilon)) \log(1/\varepsilon) + \log\log m) \\ &= \tilde{O}(\ell \log(1/\varepsilon)) + O(\log\log m). \end{aligned}$$

The total seed length of $G$ is therefore $s = s' + tu = s' + t \cdot \tilde{O}\big((\ell \log(1/\varepsilon)) + \log\log m\big)$.

We now analyze the error of $G$. For $i \in \{1, 2, \dots, t\}$, consider the variant $H_U^{(i)}$ of $H^{(1)}$, which is the same as $H^{(1)}$ but at the $i$-th step replace $H^{(i+1)}$ with $U_m$. Let $H_U^{(0)} = U_m$.

For every $i \in \{1, \dots, t\}$, for every fixed $D^{(1)}, \dots, D^{(i-1)}$ and $T^{(1)}, \dots, T^{(i-1)}$, the function $f$ restricted to $\bigwedge_{j<i} T^{(j)}$ remains a product test with $k$ functions of input length $\ell$, and remains nice if $f$ is nice. Call the restricted function $g$. Then, by Theorem 1.11, we have

$$\big|\mathbb{E}[f(H_U^{(i-1)})] - \mathbb{E}[f(H_U^{(i)})]\big| = \big|\mathbb{E}[g(U)] - \mathbb{E}[g(D^{(i)} + T^{(i)} \wedge U_m)]\big| \le \varepsilon.$$

Hence, summing over $i$ we have

$$\big|\mathbb{E}[f(U_m)] - \mathbb{E}[f(H_U^{(t)}]\big| \le \sum_{i=1}^{t} \big|\mathbb{E}[f(H_U^{(i-1)})] - \mathbb{E}[f(H_U^{(i)})]\big| \le t\varepsilon.$$

We now prove that $|\mathbb{E}[f(H_U^{(t)})] - \mathbb{E}[f(H^{(1)})]| \le \varepsilon' + 2\varepsilon$. We will show that except with probability $\varepsilon$, the function $f$ restricted to $\bigwedge_{j \le t} T^{(j)}$ is an $r$-bit product test and so we can fool the restricted function using $G'$ given by our assumption.

Write $f = \prod_{i \le k} f_i$, where each $f_i$ is defined on $\{0,1\}^{I_i}$ with $|I_i| \le \ell$. We claim that

$$\Pr\Big[\operatorname{wt}\Big(\bigwedge_{i=1}^{t} T_{I_j}^{(i)}\Big) > r \text{ for some } j \in \{1, \dots, k\}\Big] \le \varepsilon.$$

It suffices to analyze $\Pr[\operatorname{wt}(\bigwedge_{i=1}^{t} T_{I_j}^{(i)}) > r]$ for each $j$ and take a union bound over $j \le k$.

Since $|I_j| \le \ell$, $T_{I_j}^{(i)}$ is $2^{-C\ell}$-close to uniform, by Claim 3.2 and a union bound over $j \le k$, the probability that some $f_i$ has input length $> r$ is at most

$$k \binom{\ell}{r+1} \big(2^{-(r+1)} + 2^{-C\ell}\big)^t \le k \cdot \left(\frac{\ell \cdot \mathrm{e}}{r+1}\right)^{r+1} \big(2^{-r}\big)^{\Omega\left(\frac{\log(k/\varepsilon)}{r+1} + \log(\frac{\ell}{r+1})\right)} \le \varepsilon.$$

Hence, for every $D^{(1)}, \dots, D^{(t)}$, with probability $1 - \varepsilon$ over the choice of $T^{(1)}, \dots, T^{(t)}$, the function $f$ restricted to $\bigwedge_{i=1}^{t} T^{(i)}$ becomes a product with $k$ functions of input length $r$, and remains nice if $f$ is nice. Conditioned on this, we have by the definition of $G'$ that $|\mathbb{E}[f(H_U^{(t)})] - \mathbb{E}[f(H^{(1)})]| \le \varepsilon'$. Otherwise, as $|f|$ is bounded by 1, the absolute difference is always at most 2. Hence, $|\mathbb{E}[f(H_U^{(t)})] - \mathbb{E}[f(H^{(1)})]| \le \varepsilon' + 2\varepsilon$, and so the total error is at most $\varepsilon' + (t+2)\varepsilon$. □

## 4 On almost $k$-wise independent variables with small total variance

In this section we prove Lemma 2.1. We first restate the lemma.

**Lemma 4.1.** *Let $X_1, X_2, \dots, X_k$ be $k$ independent random variables over $\mathbb{C}_{\le 1}$ such that for every $i \in \{1, \dots, k\}$ and $z_i \in \operatorname{Supp}(X_i)$ we have $\Pr[X_i = z_i] \ge 2^{-\ell}$. Let $Y_1, Y_2, \dots, Y_k$ be $k$ random variables over $\mathbb{C}_{\le 1}$ that are $(\varepsilon, 16d)$-close to $X_1, \dots, X_k$. For every $\sigma \ge \sqrt{\sum_{i=1}^k \operatorname{Var}[X_i]}$,*

$$\left| \mathbb{E}\Big[\prod_{i=1}^k Y_i\Big] - \mathbb{E}\Big[\prod_{i=1}^k X_i\Big] \right| \le 2^{O(d)} \left(\frac{\sigma^2}{d}\right)^{d/2} + (k2^{\ell})^{O(d)} \varepsilon.$$

Our proof follows closely to the one in [15], which proves the lemma for $\varepsilon = 0$, that is, when the $Y_i$'s are $d$-wise independent. We first give an overview of their proof.

For independent random variables $Z_1, \dots, Z_k$, we will use $\sigma(Z)$ to denote $(\sum_{i=1}^k \operatorname{Var}[Z_i])^{1/2}$.

As a first step, let us assume each $\mathbb{E}[X_i]$ is nonzero and normalize the variables $X_i$ by writing

$$\prod_i X_i = \prod_i (\mathbb{E}[X_i] + (X_i - \mathbb{E}[X_i])) = \prod_i \mathbb{E}[X_i] \left(1 + \frac{X_i - \mathbb{E}[X_i]}{\mathbb{E}[X_i]}\right).$$

Let $Z_i$ denote $(X_i - \mathbb{E}[X_i])/\mathbb{E}[X_i]$. If $|Z_i|$ is small, then intuitively a low-order Taylor's expansion of $\prod_i (1 + Z_i)$ should approximate the original function well. To write down its Taylor's expansion, a convenient way is to rewrite $\prod_i (1 + Z_i)$ as $\mathrm{e}^{\sum_i \log(1+Z_i)}$. It suffices to bound above its error term in expectation. This is equivalent to bounding the $d$-th moment of $\sum_i \log(1 + Z_i)$. A standard calculation gives a bound in terms of the norm and variance of the functions $\log(1 + Z_i)$. Since $|Z_i|$ is small, $\log(1 + Z_i)$ behaves similarly as $Z_i$. So we can relate the error term in terms of $|Z_i|$ and $\sum_i \operatorname{Var}[Z_i] = \sigma(Z)^2$. In particular if $|Z_i| \le B$ for all $i$ then we would get an error bound of the form $2^{O(d)}(\sqrt{\sigma(Z)^2/d} + B)^{O(d)}$. For now let's think of $\mathbb{E}[X_i]$ being bounded away from 0 so that $\operatorname{Var}[Z_i] = \Theta(\operatorname{Var}[X_i])$.

Now we handle the case where $|Z_i|$ is large. Note that this implies either (1) $|X_i - \mathbb{E}[X_i]|$ is large, or (2) $\mathbb{E}[X_i]$ is small. We will handle the two conditions separately by a reduction to the case where the $|Z_i|$'s are small.

The recurring idea throughout is that we can always tolerate a few bad variables that violate the conditions, provided with high probability there can be at most $O(d)$ of them. This is because by affording an extra $O(d)$ amount of independence in the beginning, we can condition on the values of these variables and work with the remaining ones.

As a simple illustration of this idea, throughout the proof we can assume each $\operatorname{Var}[X_i]$ is bounded by $\sigma(X)^2/d$, as there can be at most $d$ bad variables $X_i$ that violate this inequality, and so we can start with $2d$-wise independence, then conditioned on the values of the bad variables $X_i$, each of the rest of the $X_i$ would satisfy the bound.

We first assume the $|\mathbb{E}[X_i]|$'s are large and handle (1), we will round the $X_i$ to $\mathbb{E}[X_i]$ whenever $|X_i - \mathbb{E}[X_i]| \ge B$. Note that by Chebyshev's inequality an $X_i$ gets rounded with probability $\operatorname{Var}[X_i]/B^2$. It follows that the probability that there are more than $d$ such $X_i$'s is at most $(\sigma(X)/Bd)^d$. This suggests taking $B$ to be $(\sigma(X)/d)^\alpha$ for some constant $\alpha \in (0,1)$ to balance the error terms.

It remains to handle condition (2), for $Z_i$ to be bounded by $B = (\sigma(X)^2/d)^{\Omega(1)}$, as explained above it suffices to show that all but $O(d)$ of the $X_i$'s satisfy $|\mathbb{E}[X_i]| \ge (\sigma(X)^2/d)^{O(1)}$. If $|\mathbb{E}[X_i]| \ge (\sigma(X)^2/d)^{O(1)}$

for $\Omega(d)$ of the $X_i$'s, then by a similar argument as above one can show that with high probability at least half of them is bounded by $(\sigma(X)^2/d)^{\Omega(1)}$. Hence, $\mathbb{E}[\prod_i X_i]$ is at most $(\sigma(X)^2/d)^{\Omega(d)}$ when the $X_i$'s are $d$-wise independent. This finishes the proof.

Note that in the case of $\varepsilon > 0$, each $X_i$ is only $\varepsilon$-close to the corresponding $Y_i$ and they are not exactly identical. As a result, throughout the proof we will often have to introduce hybrid terms to move from functions of $X_i$ to functions of $Y_i$, and vice versa, and we will show that each of these steps introduces an error of at most $k^{O(d)}\varepsilon$.

Also, there is some loss in $\varepsilon$ whenever we condition on the values of any subset of the $Y_i$'s, see Claim 4.9 for a formal claim. This introduces the extra condition that each $X_i$ must put a certain mass on each outcome.

## 4.1 Preliminaries

In this section, we prove several claims that will be used in the proof of Lemma 2.1.

**Lemma 4.2.** *For any $z \in \mathbb{C}$ with $|z| \le 1/2$, $|\log(1+z)| \le 2|z|$, where we take the principle branch of the logarithm (phase between $(-\pi,\pi)$).*

*Proof.* From the Taylor series expansion of the complex-valued log function we have

$$|\log(1+z)| = \left|\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} z^n\right| \le \sum_{n=1}^{\infty} |z|^n \le |z| \sum_{n=0}^{\infty} (1/2)^n = 2|z|. \qquad \square$$

**Lemma 4.3.** *Let $Z \in \mathbb{C}$ be a random variable with $|Z| \le 1/2$, $\mathbb{E}[Z] = 0$ and $W = \log(1+Z)$. We have* $\operatorname{Var}[W] \le 4\operatorname{Var}[Z]$.

*Proof.* Using the definition of Variance, the assumption that $\mathbb{E}[Z] = 0$ and Lemma 4.2,

$$\begin{aligned}\operatorname{Var}[W] &= \mathbb{E}[|W|^2] - |\mathbb{E}[W]|^2 \\ &\le \mathbb{E}[|W|^2] \\ &\le 4\,\mathbb{E}[|Z|^2] \\ &= 4\operatorname{Var}[Z]. \qquad \square\end{aligned}$$

**Lemma 4.4** (Taylor's approximation)**.** *For $w \in \mathbb{C}$ and $d > 0$,*

$$\left|e^w - \sum_{j=0}^{d-1} \frac{w^j}{j!}\right| \le O(1)\frac{|w|^d}{d!} \cdot \max\{1, e^{\Re(w)}\}.$$

*Proof.* We will prove the inequality $|e^w - 1| \le O(1) \cdot |w| \max\{1, e^{\Re(w)}\}$. The rest then follows from the observations that

$$e^w - \sum_{j=0}^{d} \frac{w^j}{j!} = \int_0^w \Big(e^z - \sum_{j=0}^{d-1} \frac{z^j}{j!}\Big) dz$$

and

$$\int_0^w \frac{|z|^{d-1}}{(d-1)!} \max\{1, \mathrm{e}^{\Re(z)}\} dz \le \frac{|w|^d}{d!} \cdot \max\{1, \mathrm{e}^{\Re(w)}\}.$$

We now prove the inequality at the beginning of the proof. Note that $|\mathrm{e}^w| = \mathrm{e}^{\Re(w)}$. Suppose $|w| > 1$. If $\Re(w) > 0$, then we have

$$\frac{|\mathrm{e}^w - 1|}{|w| \max\{1, \mathrm{e}^{\Re(w)}\}} \le \frac{\mathrm{e}^{\Re(w)} + 1}{\mathrm{e}^{\Re(w)}} \le 2.$$

Otherwise, $\Re(w) \le 0$ and we have

$$\frac{|\mathrm{e}^w - 1|}{|w| \max\{1, \mathrm{e}^{\Re(w)}\}} \le \mathrm{e}^{\Re(w)} + 1 \le 2.$$

Now suppose $|w| \le 1$. Then $|\mathrm{e}^w - 1| = \left| w \int_0^1 \mathrm{e}^{tw} dt \right| \le |w| \cdot e$.

□

**Lemma 4.5.** *For any random variable $W \in \mathbb{C}$, $|\mathrm{e}^{\mathbb{E}[W]}| \le \mathbb{E}[|\mathrm{e}^W|]$.*

*Proof.* By Jensen's inequality, we have

$$|\mathrm{e}^{\mathbb{E}[W]}| = |\mathrm{e}^{\mathbb{E}[\Re(W)]}| \le |\mathbb{E}[\mathrm{e}^{\Re(W)}]| = \mathbb{E}[|\mathrm{e}^W|].$$ □

**Claim 4.6.** $|\mathrm{e}^{z_1} - \mathrm{e}^{z_2}| \le |\mathrm{e}^{z_2}| \cdot O(|z_1 - z_2|)$ *whenever* $|z_1 - z_2| \le 1$.

*Proof.* By Lemma 4.4 with $d = 1$,

$$|\mathrm{e}^{z_1 - z_2} - 1| \le O(1) \cdot |z_1 - z_2| \cdot \max\{1, \mathrm{e}^{\Re(z_1 - z_2)}\} = O(|z_1 - z_2|),$$

because $\Re(z_1 - z_2) \le |z_1 - z_2| \le 1$. Therefore,

$$\begin{aligned} |\mathrm{e}^{z_1} - \mathrm{e}^{z_2}| &= |\mathrm{e}^{z_2}(\mathrm{e}^{z_1 - z_2} - 1)| \\ &= |\mathrm{e}^{z_2}||\mathrm{e}^{z_1 - z_2} - 1| \\ &\le |\mathrm{e}^{z_2}| \cdot O(|z_1 - z_2|). \end{aligned}$$ □

**Claim 4.7.** *Let $X, Y \in \Omega$ be two discrete random variables such that $\mathrm{sd}(X,Y) \le \varepsilon$. Let $f \colon \Omega \to \mathbb{C}$ be any function. We have $|\mathbb{E}[f(X)] - \mathbb{E}[f(Y)]| \le 2\max_z |f(z)| \cdot \mathrm{sd}(X,Y)$.*

*Proof.* Let $p$ and $q$ be the probability mass function of $X$ and $Y$. Using the fact that $\mathrm{sd}(X,Y) = \frac{1}{2}\sum_z |p(z) - q(z)|$, we have

$$\begin{aligned} \Big|\mathbb{E}[f(X)] - \mathbb{E}[f(Y)]\Big| &= \Big|\sum_z p(z)f(z) - \sum_z q(z)f(z)\Big| \\ &\le \sum_z |f(z)||p(z) - q(z)| \\ &\le \max_z |f(z)| \cdot \sum_z |p(z) - q(z)| \\ &= 2\max_z |f(z)| \cdot \mathrm{sd}(X,Y). \end{aligned}$$ □

**Claim 4.8** (Maclaurin's inequality (cf. [35])). *Let $z_1,\dots,z_k$ be $k$ non-negative real numbers. For any $i \in \{0,\dots,k\}$, we have*

$$S_i(z_1,\dots,z_k) := \sum_{S:|S|=i} \prod_{j\in S} z_j \le (\mathrm{e}/i)^i \Big(\sum_{j=1}^{k} z_j\Big)^i.$$

## 4.2 Proof of Lemma 2.1

We now prove Lemma 2.1. Recall that for independent random variables $Z_1,\dots,Z_k$, we use $\sigma(Z)$ to denote $(\sum_{i=1}^{k} \mathrm{Var}[Z_i])^{1/2}$. We will also denote $\sigma^2/d$ by $v$ for notational simplicity. As hinted in the overview above, throughout the proof we will assume $\mathrm{Var}[X_i] \le v = \sigma^2/d$ for every $i \in \{1,\dots,k\}$. This assumption will be used in the proof of Lemma 4.13 to give a uniform bound on how close the rounded $X_i$'s and $X_i$'s are in expectation. We will show how this assumption can be removed right after the proof of Lemma 4.20.

### 4.2.1 Assuming the variances are not too small

We first prove a claim showing that the $Y_i$ remain close to the $X_i$ even if we condition on the values of a few of the $Y_i$'s. This claim will be used multiple times throughout the proof. Note that this claim is immediate for exact independence ($\varepsilon = 0$) but less for almost independence. We shall use the assumption that each $X_i$ takes any value with probability at least $2^{-\ell}$.

**Claim 4.9.** *Let $X_1, X_2,\dots,X_k$ be $k$ independent random variables over $\mathbb{C}_{\le 1}$. Let $Y_1, Y_2,\dots,Y_k$ be $k$ random variables over $\mathbb{C}_{\le 1}$ that are $(\varepsilon, d)$-close to $X_1, X_2,\dots,X_k$. Let $S \subseteq \{1,\dots,k\}$ be a subset of size $t$. Suppose for every $i \in S$ and $z_i \in \mathrm{Supp}(X_i)$, we have $\Pr[X_i = z_i] \ge 2^{-\ell}$. Then conditioned on any values of the $Y_i$ for $i \in S$, the $Y_i$ for $i \notin S$ are $(2^{t\ell}\varepsilon, d-t)$-close to the $X_i$ for $i \notin S$.*

*Proof.* For every subset $W \subseteq [k]$ denote by $Z_W = z_W$ the event $\bigwedge_{j\in W}(Z_j = z_j)$. Let $T \subseteq [k] - S$ be a subset of size at most $d-t$. We have for every choice of $z$,

$$\begin{aligned}
&\big|\Pr[Y_T = z_T \mid Y_S = z_S] - \Pr[X_T = z_T]\big| \\
&= \left|\frac{\Pr[Y_{S\cup T} = z_{S\cup T}]}{\Pr[Y_S = z_S]} - \frac{\Pr[X_{S\cup T} = z_{S\cup T}]}{\Pr[X_S = z_S]}\right| \\
&\le \left|\frac{1}{\Pr[Y_S = z_S]} - \frac{1}{\Pr[X_S = z_S]}\right| \Pr[Y_{S\cup T} = z_{S\cup T}] + \frac{|\Pr[Y_{S\cup T} = z_{S\cup T}] - \Pr[X_{S\cup T} = z_{S\cup T}]|}{\Pr[X_S = z_S]}.
\end{aligned}$$

Summing over all possible choices of $z_T$ we have

$$\begin{aligned}
&\sum_{z_T}\big|\Pr[Y_T = z_T \mid Y_S = z_S] - \Pr[X_T = z_T]\big| \\
&\le \left|\frac{1}{\Pr[Y_S = z_S]} - \frac{1}{\Pr[X_S = z_S]}\right| \Pr[Y_S = z_S] + \frac{\varepsilon}{\Pr[X_S = z_S]} \\
&= \frac{|\Pr[X_S = z_S] - \Pr[Y_S = z_S]|}{\Pr[X_S = z_S]} + \frac{\varepsilon}{\Pr[X_S = z_S]} \\
&\le \frac{2\varepsilon}{\Pr[X_S = z_S]},
\end{aligned}$$

where the inequalities follow because $|S \cup T| \leq d$ and the variables $Y_i$ are $(\varepsilon, d)$-close to the $X_i$. Since the $X_i$ are independent and $\Pr[X_i = z_i] \geq 2^{-\ell}$, we have $\Pr[X_S = z_S] \geq 2^{-t\ell}$. This completes the proof. □

#### 4.2.2 Assuming the variables are close to their expectations and the expectations are large

**Lemma 4.10.** *Let $X_1, X_2, \ldots, X_k$ be $k$ independent discrete random variables over $\mathbb{C}_{\leq 1}$. Let $Y_1, Y_2, \ldots, Y_k$ be $k$ discrete random variables over $\mathbb{C}_{\leq 1}$ that are $(\varepsilon, d)$-close to $X_1, \ldots, X_k$. Assume for each $X_i$ and $Y_i$, there exist $Z_i$ and $Z_i'$ such that*

$$X_i = \mathbb{E}[X_i](1 + Z_i) \quad \textit{and} \quad Y_i = \mathbb{E}[X_i](1 + Z_i'),$$

*where $|Z_i| \leq B \leq 1/2$ and $|Z_i'| \leq B \leq 1/2$. Then for every $\sigma_Z \geq \sigma(Z)$*

$$\left| \mathbb{E}\left[\prod_{i=1}^{k} X_i\right] - \mathbb{E}\left[\prod_{i=1}^{k} Y_i\right] \right| \leq 2^{O(d)} \left( \frac{\sigma_Z \sqrt{d} + Bd}{d} \right)^d + (kB)^{O(d)} \varepsilon.$$

**Remark 4.11.** Note that we define $Z_i'$ above in terms of $\mathbb{E}[X_i]$ but not $\mathbb{E}[Y_i]$. The random variables $Z_i$ are independent, but the variables $Z_i'$ may not be. Also, later we will take $B$ to be $v^{1/3}$.

*Proof.* Define $W_i, \hat{W}_i$ such that

$$W_i = \log(1 + Z_i) \quad \text{and} \quad \hat{W}_i = W_i - \mathbb{E}[W_i].$$

Also define $W_i', \hat{W}_i'$ such that

$$W_i' = \log(1 + Z_i') \quad \text{and} \quad \hat{W}_i' = W_i' - \mathbb{E}[W_i].$$

Let $\hat{W} = \sum_i \hat{W}_i$ and $\hat{W}' = \sum_i \hat{W}_i'$. Note that $X_i = \mathbb{E}[X_i]\mathrm{e}^{\hat{W}_i + \mathbb{E}[W_i]}$ and $Y_i = \mathbb{E}[X_i]\mathrm{e}^{\hat{W}_i' + \mathbb{E}[W_i]}$. We have

$$\prod_{i=1}^{k} X_i = \left(\prod_{i=1}^{k} \mathbb{E}[X_i]\mathrm{e}^{\mathbb{E}[W_i]}\right)\mathrm{e}^{\hat{W}} \quad \text{and} \quad \prod_{i=1}^{k} Y_i = \left(\prod_{i=1}^{k} \mathbb{E}[X_i]\mathrm{e}^{\mathbb{E}[W_i]}\right)\mathrm{e}^{\hat{W}'}.$$

Hence the difference is

$$\prod_{i=1}^{k} X_i - \prod_{i=1}^{k} Y_i = \left(\prod_{i=1}^{k} \mathbb{E}[X_i]\mathrm{e}^{\mathbb{E}[W_i]}\right)\left(\mathrm{e}^{\hat{W}} - \mathrm{e}^{\hat{W}'}\right).$$

We rewrite $\mathrm{e}^{\hat{W}} - \mathrm{e}^{\hat{W}'}$ as a sum of 3 terms:

$$\mathrm{e}^{\hat{W}} - \mathrm{e}^{\hat{W}'} = \left(\mathrm{e}^{\hat{W}} - \sum_{j=0}^{d-1} \hat{W}^j / j!\right) + \left(\sum_{j=0}^{d-1} (\hat{W}^j - \hat{W}'^j)/j!\right) + \left(\sum_{j=0}^{d-1} \hat{W}'^j / j! - \mathrm{e}^{\hat{W}'}\right).$$

It suffices to bound above the expectation of each term multiplied by $\gamma := \prod_{i=1}^{k} \mathbb{E}[X_i]\mathrm{e}^{\mathbb{E}[W_i]}$. We bound the first and last terms using Taylor's approximation (Lemma 4.4), and the second term using $(\varepsilon, d)$-closeness of the variables. Specifically, we will show the following:

$$\mathbb{E}\left[\left|\gamma\cdot\left(\mathrm{e}^{\hat{W}'}-\sum_{j=0}^{d-1}\hat{W}'^{j}/j!\right)\right|\right]\le 2^{O(d)}\left(\frac{\sigma_Z\sqrt{d}+Bd}{d}\right)^{d}+(kB)^{O(d)}\varepsilon \tag{4.1}$$

$$\mathbb{E}\left[\left|\gamma\cdot\left(\mathrm{e}^{\hat{W}}-\sum_{j=0}^{d-1}\hat{W}^{j}/j!\right)\right|\right]\le 2^{O(d)}\left(\frac{\sigma_Z\sqrt{d}+Bd}{d}\right)^{d} \tag{4.2}$$

$$\left|\gamma\cdot\mathbb{E}\Big[\sum_{j=0}^{d-1}(\hat{W}^{j}-\hat{W}'^{j})/j!\Big]\right|\le 2k^{d}\varepsilon. \tag{4.3}$$

For (4.1), by Lemma 4.4 we have

$$\left|\gamma\cdot\left(\mathrm{e}^{\hat{W}'}-\sum_{j=0}^{d-1}\hat{W}'^{j}/j!\right)\right|\le|\gamma|\cdot O(1)\frac{|\hat{W}'|^{d}}{d!}\cdot\max\{1,\mathrm{e}^{\Re(\hat{W}')}\}.$$

We now bound above $|\gamma\cdot\max\{1,\mathrm{e}^{\Re(\hat{W}')}\}|$ by 1. We have

$$\begin{aligned}|\gamma|&=\left|\prod_{i=1}^{k}\mathbb{E}[X_i]\mathrm{e}^{\mathbb{E}[W_i]}\right|\\&=\left|\prod_{i=1}^{k}\mathbb{E}[X_i]\right|\cdot\left|\mathrm{e}^{\mathbb{E}[\sum_i W_i]}\right|\\&\le\left|\prod_{i=1}^{k}\mathbb{E}[X_i]\right|\cdot\mathbb{E}\big[|\mathrm{e}^{\sum_i W_i}|\big]&&\text{(Jensen's inequality, see Lemma 4.5)}\\&=\mathbb{E}\Big[\Big|\prod_{i=1}^{k}\mathbb{E}[X_i]\cdot\mathrm{e}^{\sum_i W_i}\Big|\Big]\\&=\mathbb{E}\Big[\Big|\prod_{i=1}^{k}X_i\Big|\Big]\\&\le 1.\end{aligned}$$

Moreover,

$$|\gamma\cdot\mathrm{e}^{\Re(\hat{W}')}|=\left|\prod_{i=1}^{k}\mathbb{E}[X_i]\mathrm{e}^{\mathbb{E}[W_i]}\right|\cdot\mathrm{e}^{\Re(\hat{W}')}=\left|\prod_{i=1}^{k}\mathbb{E}[X_i]\mathrm{e}^{\mathbb{E}[W_i]}\mathrm{e}^{\hat{W}'}\right|=\left|\prod_{i=1}^{k}Y_i\right|\le 1.$$

Hence, it suffices to bound above $\mathbb{E}[|\hat{W}'|^{d}]$. Note that the $\hat{W}_i'$'s are $(\varepsilon,d)$-close to the $\hat{W}_i$'s. So we bound above $|\hat{W}_i|$ and $\mathrm{Var}[\hat{W}_i]$ and then apply the following lemma. The same lemma and its proof can be found in Section 8 as Lemma 8.1.)

**Lemma 4.12.** *Let $\hat{W}_1,\hat{W}_2,\ldots,\hat{W}_k\in\mathbb{C}$ be independent random variables with $\mathbb{E}[\hat{W}_i]=0$, $|\hat{W}_i|<B$. Let $d$ be an even positive integer. Let $\hat{W}_1',\hat{W}_2',\ldots,\hat{W}_k'\in\mathbb{C}$ be random variables that are $(\varepsilon,d)$-close to $\hat{W}_1,\ldots,\hat{W}_k$. Then,*

$$\mathbb{E}\left[\Big|\sum_{i=1}^{k}\hat{W}_i'\Big|^{d}\right]\le 4\cdot 2^{d}\Big(\Big(\sum_i\mathrm{Var}[\hat{W}_i]\cdot d\Big)^{1/2}+dB\Big)^{d}+4(2kB)^{d}\varepsilon.$$

First, since $|Z_i| \le B$, we have $|W_i| \le 2B$ because of Lemma 4.2, and so $|\hat{W}_i| \le |W_i| + |\mathbb{E}[W_i]| \le 4B$. Next, we have $\mathrm{Var}[\hat{W}_i] \le 4\,\mathrm{Var}[Z_i]$ because of Lemma 4.3, and so $\sigma(\hat{W}) \le 2\sigma(Z)$. Therefore, applying Lemma 4.12 to $\mathbb{E}[|\hat{W}'|^d]$ and using the inequality $d! \ge (d/e)^d$, we have

$$\begin{aligned}\mathbb{E}\left[\left|\Big(\prod_{i=1}^{k}\mathbb{E}[X_i]\mathrm{e}^{\mathbb{E}[W_i]}\Big)\Big(\mathrm{e}^{\hat{W}'}-\sum_{j=0}^{d-1}\hat{W}'^{j}/j!\Big)\right|\right] &\le O(1)\frac{\mathbb{E}[|\hat{W}'|^d]}{d!}\\ &\le 2^{O(d)}\left(\frac{\sigma(\hat{W})\sqrt{d}+Bd}{d}\right)^{d}+(kB)^{O(d)}\varepsilon\\ &\le 2^{O(d)}\left(\frac{\sigma_Z\sqrt{d}+Bd}{d}\right)^{d}+(kB)^{O(d)}\varepsilon. \qquad (4.4)\end{aligned}$$

It follows by Inequality (4.4) by considering $\varepsilon = 0$ that

$$\mathbb{E}\left[\left|\Big(\prod_{i=1}^{k}\mathbb{E}[X_i]\mathrm{e}^{\mathbb{E}[W_i]}\Big)\Big(\mathrm{e}^{\hat{W}}-\sum_{j=0}^{d-1}\hat{W}^{j}/j!\Big)\right|\right] \le 2^{O(d)}\left(\frac{\sigma_Z\sqrt{d}+Bd}{d}\right)^{d}.$$

Finally we prove Inequality (4.3). By linearity of expectation,

$$\mathbb{E}\Big[\sum_{j=0}^{d-1}(\hat{W}^j-\hat{W}'^{j})/j!\Big] = \sum_{j=0}^{d-1}\big(\mathbb{E}_X[\hat{W}^j]-\mathbb{E}_Y[\hat{W}'^{j}]\big)/j!\,.$$

Note that $\hat{W}^j = (\sum_i \hat{W}_i)^j$ can be written as a sum of $k^j$ terms where each term is a product of at most $j \le d$ different $W_i$'s. Moreover, we have $|W_i| \le 2B \le 1$ for each $i$ because of Lemma 4.2. So by Claim 4.7 we have $|\mathbb{E}[\hat{W}^j] - \mathbb{E}[\hat{W}'^{j}]| \le 2k^j\varepsilon$. Hence,

$$\begin{aligned}\left|\mathbb{E}\Big[\sum_{j=0}^{d-1}(\hat{W}^j-\hat{W}'^{j})/j!\Big]\right| &\le \sum_{j=0}^{d-1}\big|\mathbb{E}[\hat{W}^j]-\mathbb{E}[\hat{W}'^{j}]\big|\\ &\le 2\sum_{j=0}^{d-1}k^j\varepsilon\\ &\le 2k^d\varepsilon.\end{aligned}$$

Recall that $|\gamma| \le 1$, this concludes (4.3). □

### 4.2.3 Assuming large expectations and small variances

We now prove the main lemma assuming each random variable $X_i$ has expectation far from zero and small variance.

**Lemma 4.13.** *Let $\sigma$ be a real number. Let $X_1, X_2, \dots, X_k$ be $k$ independent random variables over $\mathbb{C}_{\le 1}$ such that for every $i \in \{1, \dots, k\}$ and $z_i \in \operatorname{Supp}(X_i)$ we have $\Pr[X_i = z_i] \ge 2^{-\ell}$. Let $Y_1, Y_2, \dots, Y_k$ be $k$ random variables over $\mathbb{C}_{\le 1}$ that are $(\varepsilon, 9d)$-close to $X_1, \dots, X_k$. Assume for each $i$ we have $|\mathbb{E}[X_i]| \ge v^{1/6}$ and $\operatorname{Var}[X_i] \le v$, where $v = \sigma^2/d$. We have*

$$\left| \mathbb{E}\left[\prod_{i=1}^{k} X_i\right] - \mathbb{E}\left[\prod_{i=1}^{k} Y_i\right] \right| \le 2^{O(d)} v^d + (k2^\ell)^{O(d)} \varepsilon.$$

*Proof.* We will assume $v$ is less than a sufficiently small constant and $\varepsilon \le (k2^\ell)^{-Cd}$ for a sufficiently large $C$; otherwise the right hand side of the inequality is greater than 2 and there is nothing to prove.

For each $i \in \{1, 2, \dots, k\}$, we define a function $\mathrm{rd}_i \colon \mathbb{C}_{\le 1} \to \mathbb{C}_{\le 1}$ that will be used to round the variables $X_i$ and $Y_i$. We define $\mathrm{rd}_i$ as

$$\mathrm{rd}_i(z) := \begin{cases} z & \text{if } |z - \mathbb{E}[X_i]| \le v^{1/3} \\ \mathbb{E}[X_i] & \text{otherwise.} \end{cases}$$

Let $\tilde{X}_i = \mathrm{rd}_i(X_i)$ and $\tilde{Y}_i = \mathrm{rd}_i(Y_i)$. We will write both $\prod_i X_i$ and $\prod_i Y_i$ as

$$\prod_{i=1}^{k} X_i = \prod_{i=1}^{k} (X_i - \tilde{X}_i + \tilde{X}_i) = \sum_{S \subseteq \{1,2,\dots,k\}} \prod_{i \in S} (X_i - \tilde{X}_i) \prod_{i \notin S} \tilde{X}_i,$$

and

$$\prod_{i=1}^{k} Y_i = \prod_{i=1}^{k} (Y_i - \tilde{Y}_i + \tilde{Y}_i) = \sum_{S \subseteq \{1,2,\dots,k\}} \prod_{i \in S} (Y_i - \tilde{Y}_i) \prod_{i \notin S} \tilde{Y}_i.$$

Let $m = 3d$. Define

$$P_m(z_1, z_2, \dots, z_k) = \sum_{|S| < m} \prod_{i \in S} (z_i - \mathrm{rd}_i(z_i)) \prod_{i \notin S} \mathrm{rd}_i(z_i).$$

We will prove two claims below. Claim 4.14 shows that $P_m$ is a good approximation of the product in expectation under both $Y_i$'s and $X_i$'s (by setting $\varepsilon$ to 0). Claim 4.15 shows that the expectations of $P_m$ under $X_i$'s and $Y_i$'s are close.

**Claim 4.14.** $\left| \mathbb{E}\left[ \prod_i Y_i - P_m(Y_1, \dots, Y_k) \right] \right| \le 2^{O(d)} v^d + k^{O(d)} \varepsilon.$

**Claim 4.15.** $|\mathbb{E}[P_m(X_1, \dots, X_k)] - \mathbb{E}[P_m(Y_1, \dots, Y_k)]| \le 2^{O(d)} v^d + (k2^\ell)^{O(d)} \varepsilon$ .

Combining these two claims proves Lemma 4.13. □

We now prove Claims 4.14 and 4.15. We will use the following inequalities repeatedly. Recall that $v = \sigma^2/d$.

**Claim 4.16.** $\Pr[\tilde{X}_i \ne X_i] \le \operatorname{Var}[X_i] v^{-2/3} \le v^{1/3}$. *In particular,* $\sum_i \Pr[\tilde{X}_i \ne X_i] \le (d\sigma)^{2/3}$.

*Proof.* The first inequality follows from Chebyshev's inequality and second follows from the assumption $\operatorname{Var}[X_i] \le v$. The last sentence is implied by the first inequality. □

*Proof of Claim 4.14.* Consider the product $\prod_{i\in S}(Y_i-\tilde{Y}_i)$. Let $N'$ be the number of $i\in\{0,1,2,\dots,k\}$ such that $\tilde{Y}_i\neq Y_i$. If $N'<m$ then any set $S$ of size at least $m$ must contain an $i$ such that $\tilde{Y}_i=Y_i$. In this case the product is 0 and thus

$$\prod_i Y_i - P_m(Y_1,\dots,Y_k) = \sum_{|S|\geq m}\prod_{i\in S}(Y_i-\tilde{Y}_i)\prod_{i\notin S}\tilde{Y}_i = 0.$$

So,

$$\begin{aligned}\left|\mathbb{E}\left[\prod_i Y_i - P_m(Y_1,\dots,Y_k)\right]\right| &= \left|\mathbb{E}\left[\mathbb{1}(N'\geq m)\cdot\Big(\prod_i Y_i - P_m(Y_1,\dots,Y_k)\Big)\right]\right|\\ &\leq \mathbb{E}\left[\mathbb{1}(N'\geq m)\cdot\Big(\Big|\prod_i Y_i\Big| + |P_m(Y_1,\dots,Y_k)|\Big)\right]\\ &= \mathbb{E}\left[\mathbb{1}(N'\geq m)\cdot\Big|\prod_i Y_i\Big|\right] + \mathbb{E}\left[\mathbb{1}(N'\geq m)\cdot|P_m(Y_1,\dots,Y_k)|\right].\end{aligned}$$

If $N'\geq m$ then there can be at most $\sum_{s=0}^{m-1}\binom{N'}{s}\leq\sum_{s=0}^{m-1}\binom{N'}{m}\binom{m}{s}\leq 2^m\binom{N'}{m}$ subsets in the sum in $P_m$ for which the product is nonzero, and each such product can have magnitude at most $2^m$ because $|S|<m$. Thus,

$$\begin{aligned}\mathbb{1}(N'\geq m)\cdot\Big|P_m(Y_1,\dots,Y_k)\Big| &\leq \mathbb{1}(N'\geq m)\cdot 2^m\sum_{s=0}^{m-1}\binom{N'}{s}\\ &\leq \mathbb{1}(N'\geq m)\cdot 2^m\cdot 2^m\binom{N'}{m}\\ &\leq 2^{2m}\binom{N'}{m}.\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{E}\left[\mathbb{1}(N'\geq m)\cdot\Big(\Big|\prod_i Y_i\Big| + |P_m(Y_1,\dots,Y_k)|\Big)\right] &\leq \mathbb{E}\left[\mathbb{1}(N'\geq m)\cdot\Big|\prod_i Y_i\Big|\right] + 2^{2m}\,\mathbb{E}\left[\binom{N'}{m}\right]\\ &\leq \mathbb{E}[\mathbb{1}(N'\geq m)] + 2^{2m}\,\mathbb{E}\left[\binom{N'}{m}\right].\end{aligned}$$

**Claim 4.17.** $\Pr[N'\geq m]\leq\mathbb{E}\left[\binom{N'}{m}\right]\leq v^d + k^{O(d)}\varepsilon.$

*Proof.* The first inequality is clear. To see the second one, note that

$$\begin{aligned}
\mathbb{E}\left[\binom{N'}{m}\right] &\le \sum_{|S|=m} \Pr\Big[\bigwedge_{i\in S} Y_i \neq \tilde{Y}_i\Big] \\
&\le \sum_{|S|=m}\Big(\prod_{i\in S}\Pr[X_i\neq \tilde{X}_i]+\varepsilon\Big) && \text{(each } Y_i \text{ is } \varepsilon\text{-close to } X_i) \\
&\le \sum_{|S|=m}\prod_{i\in S}\Pr[X_i\neq\tilde{X}_i]+k^m\varepsilon \\
&\le \left(\frac{\mathrm{e}\cdot\sum_{i=1}^k \Pr[X_i\neq\tilde{X}_i]}{m}\right)^m + k^m\varepsilon && \text{(Maclaurin's inequality)} \\
&\le \left(\frac{\mathrm{e}\cdot(d\sigma)^{2/3}}{3d}\right)^{3d} + k^m\varepsilon && \text{(Claim 4.16)} \\
&\le v^d + k^{O(d)}\varepsilon.
\end{aligned}$$

□

Therefore,

$$\begin{aligned}
\Big|\mathbb{E}\Big[\mathbb{1}(N'\ge m)\cdot\Big(\prod_i Y_i - P_m(Y_1,\dots,Y_k)\Big)\Big]\Big| &\le \mathbb{E}[\mathbb{1}(N'\ge m)] + 2^{2m}\,\mathbb{E}\left[\binom{N'}{m}\right] \\
&\le (1+2^{6d})(v^d+k^{O(d)}\varepsilon) && (m=3d) \\
&\le 2^{O(d)}v^d + k^{O(d)}\varepsilon,
\end{aligned}$$

proving the claim. □

We have just shown that $P_m$ approximates the product well under both $X_i$ and $Y_i$ in expectation. It remains to show that $P_m(Y_1,\dots,Y_k)$ is close to $P_m(X_1,\dots,X_k)$ in expectation.

*Proof of Claim 4.15.* The difference between $P_m(X_1,\dots,X_k)$ and $P_m(Y_1,\dots,Y_k)$ equals

$$P_m(X_1,\dots,X_k) - P_m(Y_1,\dots,Y_k) = \sum_{|S|<m}\Big(\prod_{i\in S}(X_i-\tilde{X}_i)\prod_{i\notin S}\tilde{X}_i - \prod_{i\in S}(Y_i-\tilde{Y}_i)\prod_{i\notin S}\tilde{Y}_i\Big).$$

We can rewrite the right hand side as

$$\sum_{|S|<m}\bigg(\Big(\prod_{i\in S}(X_i-\tilde{X}_i) - \prod_{i\in S}(Y_i-\tilde{Y}_i)\Big)\prod_{i\notin S}\tilde{X}_i + \prod_{i\in S}(Y_i-\tilde{Y}_i)\Big(\prod_{i\notin S}\tilde{X}_i - \prod_{i\notin S}\tilde{Y}_i\Big)\bigg).$$

It suffices to show that

$$\Big|\mathbb{E}\Big[\sum_{|S|<m}\Big(\prod_{i\in S}(X_i-\tilde{X}_i)-\prod_{i\in S}(Y_i-\tilde{Y}_i)\Big)\prod_{i\notin S}\tilde{X}_i\Big]\Big| \le k^{O(d)}\varepsilon \tag{4.5}$$

$$\Big|\mathbb{E}\Big[\sum_{|S|<m}\prod_{i\in S}(Y_i-\tilde{Y}_i)\Big(\prod_{i\notin S}\tilde{X}_i-\prod_{i\notin S}\tilde{Y}_i\Big)\Big]\Big| \le 2^{O(d)}v^d + (k2^\ell)^{O(d)}\varepsilon. \tag{4.6}$$

We first prove Inequality (4.5). Because the $X_i$'s are independent, the left hand side of the inequality equals

$$\begin{aligned}
&\Big|\sum_{|S|<m}\Big(\mathbb{E}\Big[\prod_{i\in S}(X_i-\tilde{X}_i)\Big]-\mathbb{E}\Big[\prod_{i\in S}(Y_i-\tilde{Y}_i)\Big]\Big)\mathbb{E}\Big[\prod_{i\notin S}\tilde{X}_i\Big]\Big|\\
&\le\sum_{s=1}^{m-1}\sum_{|S|=s}\Big|\mathbb{E}\Big[\prod_{i\in S}(X_i-\tilde{X}_i)\Big]-\mathbb{E}\Big[\prod_{i\in S}(Y_i-\tilde{Y}_i)\Big]\Big|\cdot\Big|\mathbb{E}\Big[\prod_{i\notin S}\tilde{X}_i\Big]\Big|\\
&\le\sum_{s=1}^{m-1}\sum_{|S|=s}\Big|\mathbb{E}\Big[\prod_{i\in S}(X_i-\tilde{X}_i)\Big]-\mathbb{E}\Big[\prod_{i\in S}(Y_i-\tilde{Y}_i)\Big]\Big|\\
&\le\sum_{s=1}^{m-1}\sum_{|S|=s}2\cdot 2^s\varepsilon\\
&\le\sum_{s=1}^{m-1}k^s\cdot 2\cdot 2^s\varepsilon\\
&\le 2(2k)^m\varepsilon\\
&=k^{O(d)}\varepsilon.
\end{aligned}$$

To see the third inequality, note that $|z-\mathrm{rd}_i(z)|\le 2$, and so $|\prod_{i\in S}(z_i-\mathrm{rd}_i(z_i))|\le 2^{|S|}$. So we can apply Claim 4.7 to bound above the absolute difference by $2\cdot 2^{|S|}\varepsilon$.

Now we prove Inequality (4.6). As the $Y_i$'s are $(\varepsilon,9d)$-close to $X_i$'s, for each $S$ with $|S|\le m=3d$, if we conditioned on the values of $\tilde{Y}_i$ for which $i\in S$, then by Claim 4.9 the remaining $\tilde{Y}_i$'s for which $i\notin S$ are still $(2^{O(m\cdot\ell)}\varepsilon,6d)$-close to their corresponding $\tilde{X}_i$'s. (Recall that we can assume $\varepsilon=(k2^\ell)^{-Cd}$ for a sufficiently large $C$.) We will apply Lemma 4.10 to these $X_i$ and $Y_i$.

Define $Z_i,Z_i'$ such that $\tilde{X}_i=\mathbb{E}[\tilde{X}_i](1+Z_i)$ and $\tilde{Y}_i=\mathbb{E}[\tilde{X}_i](1+Z_i')$. To apply Lemma 4.10, we need the following two claims to bound above $|Z_i|,|Z_i'|$ and $\sigma(Z)^2$. We defer their proofs to the end.

**Claim 4.18.** *Let $B=6v^{1/6}$. Then $|Z_i|\le B$ and $|Z_i'|\le B$.*

**Claim 4.19.** $\sigma(Z)^2\le 4\sigma(X)^2v^{-1/3}\le 4\sigma^2v^{-1/3}$.

Therefore, by Lemma 4.10 with $\varepsilon'=2^{O(m\cdot\ell)}\varepsilon$ and $B=6v^{1/6}\le 1/2$ (Recall that we can assume $v$ less than a sufficiently small constant),

$$\Big|\mathbb{E}\Big[\sum_{|S|<m}\prod_{i\in S}(Y_i-\tilde{Y}_i)\Big(\prod_{i\notin S}\tilde{X}_i-\prod_{i\notin S}\tilde{Y}_i\Big)\Big]\Big|\le\sum_{|S|<m}\mathbb{E}\Big[\Big|\prod_{i\in S}(Y_i-\tilde{Y}_i)\Big|\Big]\cdot M,$$

where

$$\begin{aligned}
M &\le 2^{O(d)} \left( \frac{\sigma v^{-1/6}\sqrt{d} + dB}{d} \right)^{6d} + (Bk)^{O(d)}\varepsilon' \\
&\le 2^{O(d)} (v^{1/3} + B)^{6d} + (Bk2^{\ell})^{O(d)}\varepsilon \\
&= 2^{O(d)} \left(v^{1/3} + v^{1/6}\right)^{6d} + (k2^{\ell})^{O(d)}\varepsilon \\
&= 2^{O(d)} v^{d} + (k2^{\ell})^{O(d)}\varepsilon .
\end{aligned}$$

We now bound above $\mathbb{E}[|\prod_{i\in S}(Y_i - \tilde{Y}_i)|]$. Note that $|\prod_{i\in S}(z_i - \mathrm{rd}_i(z_i))| \le 2^{|S|}$. Hence by Claim 4.7,

$$\mathbb{E}\Big[\Big|\prod_{i\in S}(Y_i - \tilde{Y}_i)\Big|\Big] \le \mathbb{E}\Big[\Big|\prod_{i\in S}(X_i - \tilde{X}_i)\Big|\Big] + 2\cdot 2^{|S|}\varepsilon .$$

Let $N$ be the number of $i \in \{0, 1, \dots, k\}$ such that $\tilde{X}_i \ne X_i$. Note that

$$\begin{aligned}
\sum_{|S|<m} \mathbb{E}\Big[\Big|\prod_{i\in S}(X_i - \tilde{X}_i)\Big|\Big] &\le \sum_{s=0}^{m-1} \left( 2^s\, \mathbb{E}\left[\binom{N}{s}\right]\right) \\
&\le 2^m\, \mathbb{E}[2^N] \\
&= 2^m \prod_{i=1}^{k}\Big(1 + \Pr[X_i \ne \tilde{X}_i]\Big) \\
&\le 2^m \mathrm{e}^{\sum_i \Pr[X_i \ne \tilde{X}_i']} && \text{(Claim 4.16)} \\
&\le 2^m \mathrm{e}^{(d\sigma)^{2/3}} \\
&\le 2^{O(d)},
\end{aligned}$$

where the last inequality is because $\sigma^2/d \le 1$ and so $\sigma^{2/3} \le d^{1/3}$. Therefore,

$$\sum_{|S|<m} \mathbb{E}\Big[\Big|\prod_{i\in S}(Y_i - \tilde{Y}_i)\Big|\Big] \le 2^{O(d)} + 2\sum_{|S|<m} 2^{|S|}\varepsilon \le 2^{O(d)} + 2\cdot(2k)^m\varepsilon \le 2^{O(d)},$$

where the last inequality is because $\varepsilon \le k^{-Cd}$ for a sufficiently large $C$. So altogether the bound is $2^{O(d)}\cdot M \le 2^{O(d)}v^d + (k2^{\ell})^{O(d)}\varepsilon$, proving Inequality (4.6). This complete the proof of Claim 4.15. □

We now prove Claim 4.18 and 4.19. By Claim 4.16, $|\mathbb{E}[X_i] - \mathbb{E}[\tilde{X}_i]| \le 2v^{1/3}$. Also by assumption, $|\mathbb{E}[X_i]| \ge v^{1/6}$. So, we have $|\mathbb{E}[\tilde{X}_i]| \ge |\mathbb{E}[X_i]|/2 \ge v^{1/6}/2$.

*Proof of Claim 4.18.* As $|\mathbb{E}[\tilde{X}_i]| \ge v^{1/6}/2$, we have

$$\begin{aligned}
|Z_i| &= \frac{|\tilde{X}_i - \mathbb{E}[\tilde{X}_i]|}{|\mathbb{E}[\tilde{X}_i]|} \\
&\le \frac{|\tilde{X}_i - \mathbb{E}[X_i]| + |\mathbb{E}[\tilde{X}_i] - \mathbb{E}[X_i]|}{|\mathbb{E}[\tilde{X}_i]|} \\
&\le 6v^{1/3}/v^{1/6} \\
&\le 6v^{1/6},
\end{aligned}$$

and the same argument holds for $|Z_i'|$ because $|\tilde{Y}_i - \mathbb{E}[X_i]| \leq v^{1/3}$ by the definition of $\mathrm{rd}_i$. □

*Proof of Claim 4.19.* Since $h^* = \mathbb{E}[H]$ is the minimizer of $\mathbb{E}[|H-h|^2]$ for any random variable $H$, we have

$$\begin{aligned}
\mathrm{Var}[\tilde{X}_i] &= \mathbb{E}[|\tilde{X}_i - \mathbb{E}[\tilde{X}_i]|^2] \\
&\leq \mathbb{E}[|\tilde{X}_i - \mathbb{E}[X_i]|^2] \\
&\leq \mathbb{E}[|X_i - \mathbb{E}[X_i]|^2] && (\tilde{X}_i = \mathrm{rd}_i(X_i)) \\
&= \mathrm{Var}[X_i].
\end{aligned}$$

Therefore, $\mathrm{Var}[Z_i] = \mathrm{Var}[\tilde{X}_i]/|\mathbb{E}[\tilde{X}_i]|^2 \leq 4\,\mathrm{Var}[X_i]v^{-1/3}$ and thus $\sum_i \mathrm{Var}[Z_i] \leq 4\sigma^2 v^{-1/3}$. □

### 4.2.4 The general case

We first prove Lemma 2.1 by assuming $\mathrm{Var}[X_j] \leq \sigma^2/d$ for all $j$ and the $Y_i$'s are only $(\varepsilon, 15d)$-close to the $X_i$'s. Later we will handle the general case. We will again assume $\sigma^2/d$ is less than a sufficiently small constant and $\varepsilon \leq (k2^\ell)^{-Cd}$ for a sufficiently large constant $C$.

**Lemma 4.20.** *Let $X_1, X_2, \ldots, X_k$ be $k$ independent random variables over $\mathbb{C}_{\leq 1}$ such that for every $i$ and $z_i \in \mathrm{Supp}(X_i)$ we have $\Pr[X_i = z_i] \geq 2^{-\ell}$. Let $Y_1, Y_2, \ldots, Y_k$ be $k$ random variables over $\mathbb{C}_{\leq 1}$. Suppose $\mathrm{Var}[X_i] \leq v = \sigma^2/d$ for every $i \in [k]$ and the $Y_i$ are $(\varepsilon, 15d)$-close to the $X_i$. Then for every $\sigma \geq \sqrt{\sum_{i=1}^k \mathrm{Var}[X_i]}$,*

$$\left| \mathbb{E}\left[\prod_{i=1}^k Y_i\right] - \mathbb{E}\left[\prod_{i=1}^k X_i\right] \right| \leq 2^{O(d)} \left(\frac{\sigma^2}{d}\right)^{d/2} + (k2^\ell)^{O(d)}\varepsilon.$$

*Proof.* Let $m$ be the number of $i$ such that $|\mathbb{E}[X_i]| \leq v^{1/6}$. If $m \leq 6d$, let $J$ be the set of indices for which $|\mathbb{E}[X_i]| \leq v^{1/6}$. We can write

$$\prod_i X_i - \prod_i Y_i = \Big(\prod_{j\in J} X_j - \prod_{j\in J} Y_j\Big) \prod_{j\notin J} X_j + \prod_{j \in J} Y_j \Big(\prod_{j\notin J} X_j - \prod_{j\notin J} Y_j\Big).$$

It suffices to show that

$$\left|\mathbb{E}\Big[\Big(\prod_{j\in J} X_j - \prod_{j\in J} Y_j\Big) \prod_{j\notin J} X_j\Big]\right| \leq \varepsilon \tag{4.7}$$

$$\left|\mathbb{E}\Big[\prod_{j\in J} Y_j \Big(\prod_{j\notin J} X_j - \prod_{j\notin J} Y_j\Big)\Big]\right| \leq 2^{O(d)} v^d + (k2^\ell)^{O(d)}\varepsilon. \tag{4.8}$$

We first show Inequality (4.7). Since the $X_i$ are independent and the $Y_i$ are $(\varepsilon, 15d)$-close to the $X_i$, the left hand side of (4.7) is

$$\begin{aligned}
\left|\Big(\mathbb{E}\Big[\prod_{j\in J} X_j\Big] - \mathbb{E}\Big[\prod_{j\in J} Y_j\Big]\Big) \mathbb{E}\Big[\prod_{j\notin J} X_j\Big]\right| &\leq \left|\mathbb{E}\Big[\prod_{j\in J} X_j\Big] - \mathbb{E}\Big[\prod_{j\in J} Y_j\Big]\right| \\
&\leq \varepsilon.
\end{aligned}$$

To prove Inequality (4.8), note that conditioned on the values of the $Y_i$'s for which $i \in J$, by Claim 4.9, the rest of the $Y_i$'s are still $(2^{O(d\ell)}\varepsilon, 9d)$-close to the corresponding $X_i$'s with $|\mathbb{E}[X_i]| \geq v^{1/6}$. (Recall that we can assume $\varepsilon = (k2^{\ell})^{-Cd}$ for a sufficiently large $C$.) So the bound follows from Lemma 4.13.

If $m \geq 6d$, then note that

$$\left|\mathbb{E}\Big[\prod_{i=1}^{k} X_i\Big]\right| = \prod_{i=1}^{k} |\mathbb{E}[X_i]| \leq v^{m/6} \leq v^d.$$

So it suffices to show that

$$\left|\mathbb{E}\left[\prod_{i=1}^{k} Y_i\right]\right| \leq 2^{O(d)} v^{d/2} + k^{O(d)}\varepsilon.$$

Consider the event $E$ that at least $3d$ of the $Y_i$ for $i \in J$ have absolute value less than $2v^{1/6}$. Conditioned on $E$, we have that

$$\left|\prod_{i=1}^{k} Y_i\right| \leq 2^{3d} \cdot v^{d/2}.$$

We will show that $E$ happens except with probability at most $v^{2d} + k^{3d}\varepsilon$. Let $N \in \{0,1,2,\ldots,m\}$ be the number of $i \in J$ such that $|Y_i| \geq 2v^{1/6}$. Note that

$$\begin{aligned}\Pr[N \geq 3d] &\leq \sum_{S \subseteq J : |S| = 3d} \Pr\Big[\bigwedge_{i \in S} (|Y_i| \geq 2v^{1/6})\Big] \\ &\leq \sum_{S \subseteq J : |S| = 3d} \prod_{i \in S} \Pr\left[|X_i| \geq 2v^{1/6}\right] + k^{3d}\varepsilon.\end{aligned}$$

By Chebyshev's inequality,

$$\Pr[|X_i| \geq 2v^{1/6}] \leq \Pr[|X_i - \mathbb{E}[X_i]| \geq v^{1/6}] \leq \mathrm{Var}[X_i] v^{-1/3}.$$

Hence, by Maclaurin's inequality,

$$\begin{aligned}\sum_{S \subseteq J : |S| = 3d} \prod_{i \in S} \Pr\left[|X_i| \geq 2v^{1/6}\right] &\leq \left(\frac{\mathrm{e} \cdot \sum_{i=1}^{m} \Pr[|X_i| \geq 2v^{1/6}]}{3d}\right)^{3d} \\ &\leq \left(\frac{\mathrm{e} \cdot \sum_{i=1}^{m} \mathrm{Var}[X_i] v^{-1/3}}{3d}\right)^{3d} \\ &\leq \left(\frac{\mathrm{e} \cdot \sigma^2 v^{-1/3}}{3d}\right)^{3d} \\ &\leq v^{2d}.\end{aligned}$$

So,

$$\Pr[N \geq 3d] \leq v^{2d} + k^{3d}\varepsilon.$$

Therefore,

$$\begin{aligned}\left|\mathbb{E}\left[\prod_i Y_i\right]\right| &\le 2^{3d} v^{d/2} + v^{2d} + k^{3d}\varepsilon \\ &\le 2^{O(d)} v^{d/2} + k^{O(d)}\varepsilon.\end{aligned}$$

□

We now use Lemma 4.20 to prove Lemma 2.1 by paying a small overhead in the $(\varepsilon, d)$-closeness between the random variables $X_i$ and $Y_i$ to remove the assumption $\mathrm{Var}[X_i] \le v = \sigma^2/d$ for each $i$.

*Proof of Lemma 2.1.* Note that there can be at most $d$ different indices $i$ for which $\mathrm{Var}[X_i] > v$. Let $J$ be the set of these indices. We have

$$\begin{aligned}\prod_i X_i - \prod_i Y_i &= \prod_{j\in J} X_j \prod_{i\notin J} X_i - \prod_{j\in J} Y_j \prod_{i\notin J} Y_i \\ &= \Big(\prod_{j\in J} X_j - \prod_{j\in J} Y_j\Big)\prod_{i\notin J} X_j + \prod_{j\in J} Y_j \Big(\prod_{i\notin J} X_j - \prod_{i\notin J} Y_j\Big).\end{aligned}$$

We first bound the expectation of the first term. Since the $X_i$'s are independent,

$$\begin{aligned}\left|\mathop{\mathbb{E}}_{X,Y}\left[\Big(\prod_{j\in J} X_j - \prod_{j\in J} Y_j\Big)\prod_{i\notin J} X_j\right]\right| &= \left|\mathbb{E}\Big[\prod_{j\in J} X_j\Big] - \mathbb{E}\Big[\prod_{j\in J} Y_j\Big]\right| \cdot \left|\mathbb{E}\Big[\prod_{i\notin J} X_j\Big]\right| \\ &\le \left|\mathbb{E}\Big[\prod_{j\in J} X_j\Big] - \mathbb{E}\Big[\prod_{j\in J} Y_j\Big]\right| \\ &\le \varepsilon.\end{aligned}$$

For the second term, note that conditioned on the values of the $Y_j$ for which $j \in J$, by Claim 4.9, the remaining variables are $(2^{d\ell}\varepsilon, 15d)$-close to the corresponding $X_j$ and $\mathrm{Var}[X_i] \le v$ for each of them. So we can apply Lemma 4.20 and this completes the proof. □

## 5 Improved bound for bounded independence plus noise fools products

In this section we prove Theorem 1.13, which improves the error bound in Theorem 1.11 from $2^{-\Omega(d)}$ to $\ell^{-\Omega(d)}$, and Theorem 1.15, which gives the optimal error bound for nice product tests. The proof of Theorem 1.13 requires developing a few additional technical tools. We first outline the high-level idea on how to obtain the improvement.

For simplicity, we will assume $d = O(1)$ and show how to obtain an error bound of $\ell^{-\Omega(1)}$. Recall in the proof of Theorem 1.11 (see also Table 1) that we used a win-win argument on the total variance: we applied two different arguments depending on whether the total variance of a product test $f$ is above or below a certain threshold. Suppose now the total variance of $f$ is guaranteed to lie outside the interval $[\ell^{-0.1}, \ell^{0.1}]$. Then applying the same arguments as before would already give us an error of $\ell^{-\Omega(1)}$. So it suffices to handle the additional case, where the total variance is in the range of $[\ell^{-0.1}, \ell^{0.1}]$. Our goal is to use noise to reduce the total variance down to $\ell^{-0.1}$, which can then be handled by the low total variance argument. To achieve this, as a first step we will handle the functions $f_i$ with (individual) variance above and below $\ell^{-0.6}$ separately, and show that $O(\ell)$-wise independence plus noise fools the product of the $f_i$ in each case.

For the former, note that since the total variance is $\leq \ell^{0.1}$, there can be at most $\ell^{0.7}$ functions with variances above $\ell^{-0.6}$. In this case we can simply apply the result in [20] (Theorem 1.10). To prove the latter case, we use noise to reduce the variance of each function. Specifically, we use the hypercontractivity theorem to show that applying the noise operator to a function reduces its variance from $\sigma^2$ to $(\sigma^2)^{(4/3)}$. This is proved in Section 5.1 below. Hence, on average over the noise, the variance $\sigma_i^2$ of each $f_i$ is reduced to at most $(\ell^{-0.6})^{1/3}\sigma_i^2$, and so the total variance of the $f_i$ is at most $(\ell^{-0.6})^{1/3} \cdot \ell^{0.1} = \ell^{-0.1}$ and we can argue as before. To combine the two cases, we prove a new XOR Lemma for bounded independent distributions, inspired by a similar lemma for small-bias distributions which is proved in [16], and the theorem follows.

## 5.1 Noise reduces variance of bounded complex-valued functions

In this section, we show that on average, noise reduces the variance of bounded complex-valued functions. We will use the hypercontractivity theorem for complex-valued functions (cf. [21, Theorem 6.1.8]).

Let $f\colon \{0,1\}^n \to \mathbb{C}$ be any function. For every $\rho \in [0,1]$, define the noise operator $T_\rho$ to be $T_\rho f(x) := \mathbb{E}_N[f(x+N)]$, where $N = (N_1,\dots,N_n)$ is a random variable over $\{0,1\}^n$, and the variables $N_i$ are independent and $\mathbb{E}[N_i] = 1-\rho$ for all $i$.

**Theorem 5.1** (Hypercontractivity Theorem). *Let $q \in [2,\infty)$. Then for any $\rho \in [0, \sqrt{1/(q-1)}]$,*

$$\mathbb{E}\big[|T_\rho f(x)|^q\big]^{1/q} \leq |\mathbb{E}[f(x)^2]|^{1/2}.$$

We will use the following well-known corollary.

**Corollary 5.2.** *Let $f\colon \{0,1\}^n \to \mathbb{C}$. Then*

$$\mathbb{E}\big[|T_\rho f(x)|^2\big] \leq \mathbb{E}\big[|f(x)|^{1+\rho^2}\big]^{\frac{2}{1+\rho^2}}.$$

*Proof.*

$$\begin{aligned}
\mathbb{E}\big[|T_\rho f(x)|^2\big] &= \mathop{\mathbb{E}}_x\big[\mathop{\mathbb{E}}_{N,N'}[f(x+N)\overline{f(x+N')}]\big] \\
&= \mathop{\mathbb{E}}_x\big[\mathop{\mathbb{E}}_{N,N'}[f(x)\overline{f(x+N+N')}]\big] \\
&= \mathop{\mathbb{E}}_x\big[f(x)\mathop{\mathbb{E}}_{N,N'}[\overline{f(x+N+N')}]\big] \\
&= \mathop{\mathbb{E}}_x\big[f(x)\overline{T_\rho T_\rho f(x)}\big] \\
&\leq \mathbb{E}\big[|f(x)|^{1+\rho^2}\big]^{\frac{1}{1+\rho^2}}\,\mathbb{E}\big[|T_\rho T_\rho f(x)|^{1+\frac{1}{\rho^2}}\big]^{\frac{1}{1+1/\rho^2}} \\
&\leq \mathbb{E}[|f(x)|^{1+\rho^2}]^{\frac{1}{1+\rho^2}}\,\mathbb{E}[|T_\rho f(x)|^2]^{1/2}.
\end{aligned}$$

The first inequality follows from Hölder's inequality because $\frac{1}{1+\rho^2} + \frac{1}{1+1/\rho^2} = 1$, and the second inequality follows from Theorem 5.1 with $q = 1+1/\rho^2$. □

Let $T = (T_1, \dots, T_m)$ be an $m$-bit random variable, where the variables $T_i$ are independent and $\mathbb{E}[T_i] = 1-\rho$ for all $i$.

**Claim 5.3.** $\mathbb{E}_{T,U}\left[|\mathbb{E}_{U'}[f(U+T\wedge U')]|^2\right] = \mathbb{E}\left[|T_{\sqrt{\rho}}f(x)|^2\right]$.

*Proof.*

$$\begin{aligned}
\mathop{\mathbb{E}}_{T,U}\left[|\mathop{\mathbb{E}}_{U'}[f(U+T\wedge U')]|^2\right] &= \mathop{\mathbb{E}}_{T}\Big[\sum_{\alpha,\alpha'} \hat{f}_\alpha \overline{\hat{f}_{\alpha'}} \mathop{\mathbb{E}}_{U}[\chi_{\alpha+\alpha'}(U)] \mathop{\mathbb{E}}_{U'}[\chi_\alpha(T\wedge U')] \mathop{\mathbb{E}}_{U''}[\chi_{\alpha'}(T\wedge U'')]\Big] \\
&= \sum_\alpha |\hat{f}_\alpha|^2 \mathop{\mathbb{E}}_{T,U',U''}\left[\chi_\alpha\big(T\wedge(U'+U'')\big)\right] \\
&= \sum_\alpha |\hat{f}_\alpha|^2 \rho^{|\alpha|} = \mathbb{E}\left[|T_{\sqrt{\rho}}f(x)|^2\right],
\end{aligned}$$

where the last inequality follows from Parseval's identity because the Fourier expansion of $T_\rho f(x)$ is $\sum_\alpha \hat{f}_\alpha \rho^{|\alpha|}\chi_\alpha(x)$. □

We are now ready to prove that noise reduces the variance of a function. The main idea is to translate the function to a point close to its mean so that its variance is close to its second moment, and then apply Corollary 5.2 to it.

**Lemma 5.4.** *Let $f\colon \{0,1\}^n \to \mathbb{C}_{\le 1}$ be any function. Let $\delta := \min\{|f(x)-f(x')| : f(x) \ne f(x')\}$. Then*

$$\mathop{\mathbb{E}}_{T}\Big[\mathop{\mathrm{Var}}_{x}\big[\mathop{\mathbb{E}}_{U}[f(x+T\wedge U)]\big]\Big] \le 4\left(\frac{2\,\mathrm{Var}[f]}{\delta^2}\right)^{\frac{2}{1+\rho}}.$$

*Proof.* We can assume $\mathrm{Var}[f] \le \delta^2/2$; otherwise the conclusion is trivial. Let $S$ be the support of $f$. For every $y \in S$, let $p_y := \Pr[f(x) = y]$. Let $\mu = \mathbb{E}[f]$ and $\sigma^2 = \mathrm{Var}[f]$. Since $\sigma^2 = \mathbb{E}[|f(x)-\mu|^2]$, there is a point $z \in S$ such that $|z-\mu|^2 \le \sigma^2$. We have

$$\sigma^2 = \sum_{y\in S} p_y|y-\mu|^2 \ge \sum_{y\in S: y\ne z} p_y|y-\mu|^2 \ge \min_{y\in S:y\ne z}|y-\mu|^2\Big(\sum_{y\in S:y\ne z} p_y\Big).$$

Define $g(x) := \frac{f(x)-z}{2}$. We have for every $t$,

$$\mathop{\mathrm{Var}}_{x}\big[\mathop{\mathbb{E}}_{U}[f(x+t\wedge U)]\big] = 4\mathop{\mathrm{Var}}_{x}\big[\mathop{\mathbb{E}}_{U}[g(x+t\wedge U)]\big] \le 4\mathop{\mathbb{E}}_{x}\Big[\big|\mathop{\mathbb{E}}_{U}[g(x+t\wedge U)]\big|^2\Big].$$

By Corollary 5.2,

$$\mathbb{E}\left[|T_\rho g|^2\right] \le \mathbb{E}\left[|g|^{1+\rho^2}\right]^{\frac{2}{1+\rho^2}} = \left(\sum_{y\in S:y\ne z} p_y\left|\frac{y-z}{2}\right|^{1+\rho^2}\right)^{\frac{2}{1+\rho^2}} \le \Big(\sum_{y\in S:y\ne z} p_y\Big)^{\frac{2}{1+\rho^2}}$$

because $|y-y'| \le 2$ for every $y, y' \in \mathbb{C}_{\le 1}$. So by Claim 5.3, we have

$$\mathop{\mathbb{E}}_{T,x}\Big[\big|\mathop{\mathbb{E}}_{U}[g(x+T\wedge U)]\big|^2\Big] = \mathbb{E}\left[|T_\rho g|^2\right] \le \Big(\sum_{y\in S:y\ne z} p_y\Big)^{\frac{2}{1+\rho}}.$$

It follows from above that

$$\mathop{\mathbb{E}}_{T}\Big[\mathop{\mathrm{Var}}_{x}\big[\mathop{\mathbb{E}}_{U}[f(x+T\wedge U)]\big]\Big] \le 4\mathop{\mathbb{E}}_{T,x}\big[\big|\mathop{\mathbb{E}}_{U}[g(x+T\wedge U)]\big|^2\big] \le 4\Big(\sum_{y\in S:y\neq z} p_y\Big)^{\frac{2}{1+\rho}} \le 4\left(\frac{\mathrm{Var}[f]}{\min_{y\in S:y\neq z}|y-\mu|^2}\right)^{\frac{2}{1+\rho}}.$$

Now we bound below $\min_{y\in S:y\neq z}|y-\mu|^2$. For every $y\neq z$,

$$\delta^2 \le |y-z|^2 \le |y-\mu|^2 + |\mu - z|^2 \le |y-\mu|^2+\sigma^2.$$

Because $\sigma^2 \le \delta^2/2$, we have

$$\mathop{\mathbb{E}}_{T}\Big[\mathop{\mathrm{Var}}_{x}\big[\mathop{\mathbb{E}}_{U}[f(x+T\wedge U)]\big]\Big] \le 4\left(\frac{\mathrm{Var}[f]}{\delta^2-\sigma^2}\right)^{\frac{2}{1+\rho}} \le 4\left(\frac{2\,\mathrm{Var}[f]}{\delta^2}\right)^{\frac{2}{1+\rho}}. \qquad \square$$

**Remark 5.5.** The dependence on $\delta$ is necessary. Consider a function $f$ with support $\{0,\varepsilon\}$. Then $f=\varepsilon g$, where $g$ has support $\{0,1\}$. We have $\mathrm{Var}[f]=\varepsilon^2\,\mathrm{Var}[g]$. Applying noise to $f$ is the same as applying noise to $g$, but $g$ has no dependence on $\varepsilon$.

### 5.2 XOR Lemma for bounded independence

We now prove a version of XOR lemma for bounded independent distributions that is similar to the one in [16], which proves the lemma for small-bias distributions.

**Lemma 5.6.** *Let $f_1,\dots,f_k\colon \{0,1\}^m \to [0,1]$ be $k$ functions on disjoint inputs. Let $H\colon [0,1]^k \to [0,1]$ be a multilinear function in its input. If each $f_i$ is fooled by any $d_i$-wise independent distribution with error $\varepsilon$, then the function $h\colon \{0,1\}^m\to[0,1]$ defined by $h(x) := H(f_1(x), f_2(x),\dots,f_k(x))$ is fooled by any $(\sum_{i\le k} d_i)$-wise independent distribution with error $8^k\varepsilon$.*

We will use the following dual equivalence between bounded independence and sandwiching polynomials that was introduced by Bazzi [4]. This equivalence was stated for Boolean functions in [4], but it is clear from the proof that it holds for $[0,1]$-valued functions as well.

**Fact 5.7** ([4]). *A function $f\colon \{0,1\}^m\to[0,1]$ is fooled by every $d$-wise independent distribution if and only if there exist two multivariate polynomials $p_\ell$ and $p_u$ of degree $d$ such that*

1. *For every $x\in\{0,1\}^m$, we have $p_\ell(x)\le f(x)\le p_u(x)$, and*
2. *$\mathbb{E}[p_u(U)-f(U)]\le\varepsilon$ and $\mathbb{E}[f(U)-p_\ell(U)]\le\varepsilon$.*

*Proof of Lemma 5.6.* By Fact 5.7, for each $i\in\{1,\dots,k\}$, there exist two degree-$d_i$ polynomials $f_i^u$ and $f_i^\ell$ for $f_i$ which satisfy the conditions in Fact 5.7. Hence, we have

$$f_i^u(x)\ge f_i(x)\ge 0 \quad\text{and}\quad 1-f_i^\ell(x)\ge 1-f_i(x)\ge 0.$$

For every $\alpha\in\{0,1\}^k$, define

$$M_\alpha^u(x) := \prod_{i:\alpha_i=1} f_i^u(x)\prod_{j:\alpha_j=0}\big(1-f_j^\ell(x)\big) \quad\text{and}\quad M_\alpha(x) := \prod_{i:\alpha_i=1} f_i(x)\prod_{j:\alpha_j=0}\big(1-f_j(x)\big).$$

Clearly, $M^u_\alpha(x) \geq M_\alpha(x)$, and $M^u_\alpha(x)$ has degree $\sum_{i\leq k} d_i$. We claim that for every $\alpha \in \{0,1\}^k$,

$$\mathbb{E}[M^u_\alpha(x) - M_\alpha(x)] \leq 2^k \varepsilon.$$

Fix a string $\alpha \in \{0,1\}^k$. Define the hybrids $M_0 = M^u_\alpha(x), M_1, \dots, M_k = M_\alpha(x)$, where

$$M_i(x) := M_i^{(1)}(x) \cdot M_i^{(2)}(x),$$

where

$$M_i^{(1)}(x) := \prod_{j\leq i, \alpha_j=1} f_j(x) \prod_{j\leq i:\alpha_j=0} \big(1 - f_j(x)\big),$$

and

$$M_i^{(2)}(x) := \prod_{j>i:\alpha_j=1} f^u_j(x) \prod_{j>i:\alpha_j=0} \big(1 - f^\ell_j(x)\big).$$

Note that

$$\mathbb{E}[M_i^{(2)}(x)] = \prod_{j>i:\alpha_j=1} \mathbb{E}\big[f^u_j(x)\big] \prod_{j>i:\alpha_j=0} \mathbb{E}\big[\big(1 - f^\ell_j(x)\big)\big] \leq (1+\varepsilon)^{k-i},$$

and $M_i^{(1)}(x) \leq 1$. So, if $\alpha_i = 1$, then

$$\mathbb{E}\big[M_{i-1}(x) - M_i(x)\big] = \mathbb{E}\Big[\big(f^u_i(x) - f_i(x)\big) \cdot M^{(1)}_{i-1}(x) \cdot M_i^{(2)}(x)\Big] \leq \varepsilon \cdot (1+\varepsilon)^{k-i}.$$

Likewise, if $\alpha_i = 0$, we have

$$\mathbb{E}[M_{i-1}(x) - M_i(x)] = \mathbb{E}\Big[\big((1 - f^\ell_i(x)) - (1 - f_i(x))\big) \cdot M^{(1)}_{i-1}(x) \cdot M_i^{(2)}(x)\Big] \leq \varepsilon \cdot (1+\varepsilon)^{k-i}.$$

Hence,

$$\mathbb{E}[M^u_\alpha(x) - M_\alpha(x)] \leq \sum_{1\leq i\leq k} \mathbb{E}[M_{i-1}(x) - M_i(x)] \leq \varepsilon \sum_{0\leq i\leq k-1} (1+\varepsilon)^i \leq 2^k\varepsilon.$$

Now we define $M^\ell_\alpha(x) := 1 - \sum_{\beta:\beta\neq\alpha} M^u_\beta(x)$. Note that $M^\ell_\alpha(x)$ also has degree $\sum_{i\leq k} d_i$. Since

$$\sum_{\alpha\in\{0,1\}^k} M_\alpha(x) = \prod_{i\leq k} \big(f_i(x) + (1 - f_i(x))\big) = 1,$$

we have

$$M^\ell_\alpha(x) = 1 - \sum_{\beta:\beta\neq\alpha} M^u_\beta(x) \leq 1 - \sum_{\beta:\beta\neq\alpha} M_\beta(x) = M_\alpha(x).$$

Hence,

$$\mathbb{E}[M_\alpha(x) - M^\ell_\alpha(x)] = \sum_{\beta:\beta\neq\alpha} \big(M^u_\beta(x) - M_\beta(x)\big) \leq \sum_{\beta:\beta\neq\alpha} 2^k\varepsilon \leq 2^k 2^k \varepsilon = 4^k\varepsilon.$$

As $H$ is multilinear, we can write $H$ as

$$H(y_1, \dots, y_k) = \sum_{\alpha\in\{0,1\}^k} H(\alpha) \prod_{i:\alpha_i=1} y_i \prod_{i:\alpha_i=0} (1 - y_i),$$

where $H(\alpha) \in [0,1]$ for every $\alpha$. So

$$h(x) = \sum_{\alpha \in \{0,1\}^k} H(\alpha) \prod_{i:\alpha_i=1} f_i(x) \prod_{i:\alpha_i=0} (1-f_i(x)) = \sum_{\alpha \in \{0,1\}^k} H(\alpha) M_\alpha(x).$$

Now if we define

$$h^u(x) := \sum_{\alpha \in \{0,1\}^k} H(\alpha) M^u_\alpha(x) \quad \text{and} \quad h^\ell(x) := \sum_{\alpha \in \{0,1\}^k} H(\alpha) M^\ell_\alpha(x).$$

Clearly $h^u$ and $h^\ell$ both have degree $\sum_{i \le k} d_i$. We also have $h^u(x) \ge h(x) \ge h^\ell(x)$, and

$$\mathbb{E}[h^u(x) - h(x)] \le \sum_{\alpha \in \{0,1\}^k} H(\alpha)\, \mathbb{E}[M^u_\alpha(x) - M_\alpha(x)] \le \sum_{\alpha \in \{0,1\}^k} 2^k \varepsilon \le 4^k \varepsilon.$$

A similar calculation shows that $\mathbb{E}[h(x) - h^\ell(x)] \le 8^k \varepsilon$. Therefore, since $h^u$ and $h^\ell$ are two polynomials that satisfy the conditions in Fact 5.7 with error $8^k\varepsilon$, the lemma follows. □

## 5.3 Proof of Theorem 1.13

Armed with Lemma 5.4 and Lemma 5.6, we are now ready to prove Theorem 1.13. We first need the following useful fact to handle the case when $S$ is the $M$-th roots of unity.

**Fact 5.8.** *Let $X$ and $Y$ be two random variables on $\{0,1\}^m$. Suppose for every $m$-bit product test $g\colon \{0,1\}^m \to S$, where $S$ is the set of all $M$-th roots of unity, we have $\left|\mathbb{E}[g(X)] - \mathbb{E}[g(Y)]\right| \le \varepsilon$. Then for every $m$-bit product test $g\colon \{0,1\}^m \to S$ and every $z \in S$, we have $\left|\Pr[g(X) = z] - \Pr[g(Y) = z]\right| \le \varepsilon$.*

*Proof.* Let $\omega$ be a primitive $M$-th root of unity. Note that for every integer $j$, the function $g^j$ is also a product test with the same range. So for every $j,k \in \{0,\dots,M-1\}$,

$$\left|\mathbb{E}[(\omega^{-k} g(X))^j] - \mathbb{E}[(\omega^{-k} g(Y))^j]\right| \le \left|\omega^{-kj}\right| \cdot \left|\mathbb{E}[g(X)^j] - \mathbb{E}[g(Y)^j]\right| \le \varepsilon.$$

Using the identity

$$\frac{1}{M}\sum_{j=0}^{M-1} \omega^{(i-k)j} = \begin{cases} 1 & \text{if } i = k \\ 0 & \text{otherwise,} \end{cases}$$

we have for every $k \in \{0,\dots,M-1\}$,

$$\left|\Pr[g(X) = \omega^k] - \Pr[g(Y) = \omega^k]\right| \le \frac{1}{M}\sum_{j=0}^{M-1}\left|\mathbb{E}\left[(\omega^{-k} g(X))^j\right] - \mathbb{E}\left[(\omega^{-k} g(Y))^j\right]\right| \le \varepsilon. \qquad \square$$

*Proof of Theorem 1.13.* Write $f = \prod_{i=1}^k f_i$, where $f_i\colon \{0,1\}^{I_i} \to \mathbb{C}_{\le 1}$. Let $\sigma$ denote $(\sum_{i \le k} \mathrm{Var}[f_i])^{1/2}$. We will consider two cases: $\sigma^2 \ge d\ell^{0.1}$ and $\sigma^2 \le d\ell^{0.1}$.

If $\sigma^2 \ge d\ell^{0.1}$, then the expectation of $f$ under the uniform distribution is small. Specifically, we have

$$\left|\prod_{i \le k} \mathop{\mathbb{E}}_U[f_i(U)]\right| \le \prod_{i \le k} (1 - \mathrm{Var}[f_i])^{1/2} \le \mathrm{e}^{-\frac{1}{2}\sigma^2} \le 2^{-\Omega(d\ell^{0.1})} \le \ell^{-\Omega(d)}. \tag{5.1}$$

Thus, it suffices to show that its expectation under $D+T\wedge U$ is at most $\ell^{-\Omega(d)}$. We now use Claim 2.2 to show that

$$\left|\mathop{\mathbb{E}}_{D,T,U}\left[\prod_{i=1}^{k} f_i(D+T\wedge U)\right]\right| \le \ell^{-\Omega(d)}.$$

For each $t,x\in\{0,1\}^m$, and each $i\in\{1,2,\ldots,k\}$, let $\sigma^2_{t,x,i}$ denote $\mathrm{Var}_{U'}[f_i(x+t\wedge U')]$. Let $T'$ be a uniform $m$-bit random variable. By Claim 2.2 with $\eta=1/2$, we have $\mathbb{E}_{T',U}[\sigma^2_{T',U,i}]\ge \mathrm{Var}[f_i]/2$. So by linearity of expectation,

$$\mathop{\mathbb{E}}_{T',U}\left[\sum_{i\le k}\sigma^2_{T',U,i}\right]\ge \sigma^2/2\ge d\ell^{0.1}/2.$$

Since $T$ and $D$ are both $d\ell$-wise independent, the random variables $\sigma^2_{T,D,1},\ldots,\sigma^2_{T,D,k}$ are $(0,d)$-close to $\sigma^2_{T',U,1},\ldots,\sigma^2_{T',U,k}$. Let $\mu=\mathbb{E}_{T',U}\left[\sum_{i\le k}\sigma^2_{T',U,i}\right]\ge d\ell^{0.1}/2$. By Lemma 2.3,

$$\Pr_{T,D}\left[\sum_{i\le k}\sigma^2_{T,D,i}\le \mu/2\right]\le 4\cdot 2^d\left(\frac{\sqrt{\mu d}+d}{\mu/2}\right)^d=\ell^{-\Omega(d)}.$$

Hence, except with probability $\ell^{-\Omega(d)}$ over $t\in T$ and $x\in D$, we have

$$\sum_{i\le k}\sigma^2_{t,x,i}=\sum_{i\le k}\mathop{\mathrm{Var}}_{U'}[f_i(x+t\wedge U')]\ge d\ell^{0.1}/4.$$

By a similar calculation to (5.1), for every such $t$ and $x$,

$$\begin{aligned}\left|\prod_{i\le k}\mathop{\mathbb{E}}_{U}[f_i(x+t\wedge U)]\right| &\le \prod_{i\le k}\left|\mathop{\mathbb{E}}_{U}[f_i(x+t\wedge U)]\right|\\ &=\prod_{i\le k}(1-\sigma^2_{t,x,i})^{1/2}\\ &\le \mathrm{e}^{-\frac{1}{2}\sum_{i\le k}\sigma^2_{t,x,i}}\le 2^{-\Omega(d\ell^{0.1})}\le \ell^{-\Omega(d)}.\end{aligned}$$

In addition, we always have $|f|\le 1$. Hence,

$$\left|\mathop{\mathbb{E}}_{D,T,U}\left[\prod_{i\le k}f_i(D+T\wedge U)\right]\right|\le \mathop{\mathbb{E}}_{D,T}\left[\left|\prod_{i\le k}\mathop{\mathbb{E}}_{U}[f_i(D+T\wedge U)]\right|\right]\le \ell^{-\Omega(d)}.$$

Suppose $\sigma^2\le d\ell^{0.1}$. Let $\sigma_1^2\ge\sigma_2^2\ge\cdots\ge\sigma_k^2$ be the variances of $f_1,f_2,\ldots,f_k$ respectively. Let $k'=d\ell^{0.7}$. We have $\sigma^2_{k'}\le d\ell^{0.1}/k'=\ell^{-0.6}$; for otherwise $\sigma^2\ge\sum_{i=1}^{k'}\sigma_i^2\ge k'\sigma^2_{k'}>d\ell^{0.1}$, a contradiction. Let $T'$ be a uniform $m$-bit random variable. Let $\tilde{\sigma}_i^2$ denote

$$\mathop{\mathrm{Var}}_{T',U}\left[\mathop{\mathbb{E}}_{U'}[f_i(U+T'\wedge U')]\right].$$

We now show that $\tilde{\sigma}_i^2 \le O(\sigma_i^2)^{4/3}$. For every $i \in \{1,\dots,k\}$, define $g_i \colon \{0,1\}^m \to \mathbb{C}_{\le 1}$ to be $g_i(x) = (f_i(x) - \mathbb{E}[f_i])/2$ so that $\mathbb{E}[g_i] = 0$ and $\operatorname{Var}[g_i] = \operatorname{Var}[f_i]/4$. We apply Lemma 5.4 with $\rho = 1/2$. Notice that since $M$ is fixed, we have $|g(x) - g(x')| = \Omega(1)$ whenever $g(x) \ne g(x')$. Hence,

$$\begin{aligned}\tilde{\sigma}_i^2 &= 4 \operatorname*{Var}_{T',U}\Big[\mathop{\mathbb{E}}_{U'}[g_i(U + T' \wedge U')]\Big]\\ &= 4 \mathop{\mathbb{E}}_{T'}\Big[\operatorname*{Var}_{U}\big[\mathop{\mathbb{E}}_{U'}[g_i(U + T' \wedge U')]\big]\Big]\\ &= O(\sigma_i^2)^{4/3}.\end{aligned}$$

It follows that

$$\sum_{i>k'} \tilde{\sigma}_i^2 = O\Big(\sum_{i>k'} (\sigma_i^2)^{4/3}\Big) \le O\big((\sigma_{k'}^2)^{1/3}\big) \sum_{i>k'} \sigma_i^2 \le O(\ell^{-0.2}) \cdot d\ell^{0.1} = d\ell^{-\Omega(1)}.$$

Now, if we let $F_2 := \prod_{i>k'} f_i$, then by Lemma 2.1,

$$\Big|\mathop{\mathbb{E}}_{D,T,U}[F_2(D + T \wedge U)] - \mathop{\mathbb{E}}_{U}[F_2(U)]\Big| \le \ell^{-\Omega(d)}. \tag{5.2}$$

On the other hand, if we define $F_1$ to be $\prod_{i=1}^{k'} f_i$, then it follows from Theorem 1.10 that

$$\Big|\mathop{\mathbb{E}}_{D,T,U}[F_1(D + T \wedge U)] - \mathop{\mathbb{E}}_{U}[F_1(U)]\Big| \le k' 2^{-\Omega(d^2\ell/k')} = 2^{-\Omega(d\ell^{0.3})}. \tag{5.3}$$

We now combine (5.2) and (5.3) using Lemma 5.6. To begin, define $g_1(x) := \mathbb{E}_{T,U}[F_1(x + T \wedge U)]$ and $g_2(x) := \mathbb{E}_{T,U}[F_2(x + T \wedge U)]$.

Let $S$ be the range of the functions $f_i$. If $S = [0,1]$, then the theorem follows immediately by applying Lemma 5.6 to $g_1$ and $g_2$. However, if $S$ is the set of $M$-th roots of unity, then we cannot apply Lemma 5.6 directly because it only applies to functions with range $[0,1]$. Nevertheless we can use Fact 5.8 to reduce from $S$ to $\{0,1\}$.

We now reduce $S$ to $\{0,1\}$ and apply Lemma 5.6. For every $z \in S$, we define the point function $\mathbb{1}_z \colon S \to \{0,1\}$ by $\mathbb{1}_z(x) = 1$ if and only if $x = z$. Then for every random variable $Z$ on $S$,

$$\mathbb{E}[Z] = \sum_{z \in S} z \Pr[Z = z] = \sum_{z \in S} z\, \mathbb{E}[\mathbb{1}_z(Z)].$$

Hence,

$$\begin{aligned}\mathbb{E}[g_1(X) g_2(X)] &= \sum_{z \in S} z\, \mathbb{E}\big[\mathbb{1}_z\big(g_1(X) g_2(X)\big)\big]\\ &= \sum_{z \in S} z\, \mathbb{E}\Big[\sum_{u,v \in S : uv = z} \mathbb{1}_u\big(g_1(X)\big) \mathbb{1}_v\big(g_2(X)\big)\Big]\\ &= \sum_{z \in S} z \sum_{u,v \in S : uv = z} \mathbb{E}\Big[\mathbb{1}_u\big(g_1(X)\big) \mathbb{1}_v\big(g_2(X)\big)\Big].\end{aligned}$$

Hence, by Fact 5.8, for every $u, v \in S$, the functions $\mathbb{1}_u \circ g_1$ and $\mathbb{1}_v \circ g_2$ are fooled by $d$-wise independence with error $\ell^{-\Omega(d)}$. So by Lemma 5.6, $(\mathbb{1}_u \circ f)(\mathbb{1}_v \circ g)$ are fooled by $2d$-wise independence with error $\ell^{-\Omega(d)}$. Hence,

$$\begin{aligned}
&\left|\mathbb{E}[f(D+T\wedge U)] - \mathbb{E}[f(U)]\right| \\
&= \left|\mathbb{E}[(g_1 g_2)(D)] - \mathbb{E}[(g_1 g_2)(U)]\right| \\
&\leq \sum_{z\in S}|z| \sum_{u,v\in S: uv=z} \left|\mathbb{E}\big[(\mathbb{1}_u \circ g_1)(\mathbb{1}_v \circ g_2)\big(D\big)\big] - \mathbb{E}\big[(\mathbb{1}_u \circ g_1)(\mathbb{1}_v \circ g_2)\big(U\big)\big]\right| \\
&\leq M^2 \cdot \ell^{-\Omega(d)} = \ell^{-\Omega(d)}
\end{aligned}$$

because $M$ is fixed, proving the theorem. □

## 5.4 Proof of Theorem 1.15

We now prove Theorem 1.15. We will need the following theorem that is implicit in [13]. The original theorem was stated for read-once branching programs. Below we sketch how to modify their proof to handle product tests. Combining the theorem with Claim 2.4 proves Theorem 1.15.

**Theorem 5.9** ([13] (Implicit))**.** *Let $f : \{0,1\}^m \to \mathbb{C}_{\leq 1}$ be an $m$-bit product test with $k$ functions of input length $\ell$. Let $D, T$ and $U$ be three independent $m$-bit random variables, where $D$ and $T$ are $2t$-wise independent and $U$ is uniform. Then*

$$\left|\mathop{\mathbb{E}}_{D,T,U}[f(D+T\wedge U)] - \mathop{\mathbb{E}}_{U}[f(U)]\right| \leq k\cdot 2^{-(t-\ell+1)/2}.$$

*Proof.* We can assume $t \geq \ell$ for otherwise the conclusion is trivial. Let $t' := t-\ell+1 \geq 1$. We slightly modify the decomposition in [13, Proposition 6.1] as follows. Let $f$ be an $m$-bit product test and write $f = \prod_{i=1}^k f_i$. As the random variable $D+T\wedge U$ is symmetric, i. e., invariant under permutations of its coordinates, we can assume the function $f_i$ is defined on the $i$-th block of $\ell$ bits. For every $i \in \{1,\dots,k\}$, let $f^{\leq i} = \prod_{j\leq i} f_j$ and $f^{>i} = \prod_{j>i} f_j$. We decompose $f$ into

$$f = \hat{f}_\emptyset + L + \sum_{i=1}^k H_i f^{>i}, \tag{5.4}$$

where

$$L := \sum_{\substack{\alpha\in\{0,1\}^{\ell k}\\ 0<|\alpha|<t'}} \hat{f}_\alpha \chi_\alpha$$

and

$$H_i := \sum_{\substack{\alpha=(\alpha_1,\dots,\alpha_i)\in\{0,1\}^{\ell i}:\\ \text{the } t'\text{-th 1 in } \alpha \text{ appears in } \alpha_i}} \hat{f}^{\leq i}_\alpha \chi_\alpha.$$

We now show that the expressions on both sides of Equation (5.4) 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 $\alpha = (\alpha_1, \dots, \alpha_k) \in \{0,1\}^{\ell k}$ and $\hat{f}_\alpha = \prod_{i=1}^{k} \hat{f}_i(\alpha_i)$ be a coefficient of $f$. If $|\alpha| < t'$, then $\hat{f}_\alpha$ appears in $\hat{f}_\emptyset$ or $L$. Otherwise, $|\alpha| \geq t'$. Then the $t'$-th 1 in $\alpha$ must appear in one of $\alpha_1, \dots, \alpha_k$. Say it appears in $\alpha_i$. Then we claim that $\alpha$ appears in $H_i f^{>i}$. This is because the coefficient indexed by $(\alpha_1, \dots, \alpha_i)$ appears in $H_i$, and the coefficient indexed by $(\alpha_{i+1}, \dots, \alpha_k)$ appears in $f^{>i}$. Note that all the coefficients in each function $H_i$ have weights between $t' = t - \ell + 1$ and $t' + \ell - 1 = t$, and because our random variables $D$ and $T$ are both $2t$-wise independent, we get an error of $2^{-t'} = 2^{-(t-\ell+1)}$ in Lemma 6.2 in [13]. The rest of the analysis follows from [13] or [20]. □

Theorem 1.15 easily follows from Theorem 5.9 and Claim 2.4.

*Proof of Theorem 1.15.* We may assume $t \geq 8\ell$, otherwise the conclusion is trivial. If $k \geq 2^{3\ell+1}\lceil t/\ell \rceil$, then the theorem follows from Claim 2.4. Otherwise, $k \leq 2^{3\ell+1}\lceil t/\ell \rceil$ and the theorem follows from Theorem 5.9. □

# 6 Small-bias plus noise fools degree-2 polynomials

In this section we show that small-bias distributions plus noise fool non-read-once $\mathbb{F}_2$-polynomials of degree 2. We first state a structural theorem about degree-2 polynomials over $\mathbb{F}_2$ which will be used in our proof.

**Theorem 6.1** (Theorem 6.30 in [24]). *For every $\mathbb{F}_2$-polynomial $p\colon \{0,1\}^m \to \{0,1\}$ of degree 2, there exists an invertible matrix $A \in \mathbb{F}_2^{m \times m}$, an integer $k \leq \lfloor m/2 \rfloor$, and a subset $L \subseteq [m]$ such that $p(Ax) := \sum_{i=1}^{k} x_{2i-1}x_{2i} + \sum_{i \in L} x_i$.*

*Proof of Claim 1.14.* Let $p$ be a degree-2 polynomial. It suffices to fool $q(x) := (-1)^{p(x)}$. By Theorem 6.1, there exists an invertible matrix $A$, an integer $k$ and a subset $L \subseteq [m]$ such that $q(Ax) = r(x) \cdot \chi_L(x)$, where $r(x) := (-1)^{\sum_{i=1}^{k} x_{2i-1}x_{2i}}$, and $\chi_L(x) = (-1)^{\sum_{i \in L} x_i}$. By writing $r(x)$ in its Fourier expansion, $q(Ax)$ has the Fourier expansion

$$q(Ax) = \Big( \sum_{S \subseteq [2k]} \hat{r}_S \chi_S(x) \Big) \chi_L(x),$$

where $|\hat{r}_S| = 2^{-k}$. Note that $L$ is a subset of $[m]$. Viewing the sets $S$ and $L$ as vectors in $\{0,1\}^m$, we have

$$\begin{aligned} \big|\mathbb{E}[q(D + T \wedge U)] - \mathbb{E}[q(U)]\big| &\leq \sum_{\emptyset \neq S \subseteq [2k]} 2^{-k} \big|\mathbb{E}[\chi_{S+L}(A^{-1}(D))]\big| \cdot \big|\mathbb{E}[\chi_{S+L}(A^{-1}(T \wedge U))]\big| \\ &\leq 2^{-k}\delta \sum_{\emptyset \neq S \subseteq [2k]} \big|\mathbb{E}[\chi_{S+L}(A^{-1}(T \wedge U))]\big| \\ &= 2^{-k}\delta \sum_{\emptyset \neq S \subseteq [2k]} \big|\mathbb{E}[\chi_{A(S+L)}(T \wedge U)]\big| \\ &= 2^{-k}\delta \sum_{\emptyset \neq S \subseteq [2k]} (1/3)^{|A(S+L)|}, \end{aligned}$$

where the second inequality follows because small-bias distributions are closed under linear transformations. We now bound above the summation. We claim that

$$\sum_{S\subseteq[2k]} (1/3)^{|A(S+L)|} \le \sum_{S\subseteq[2k]} (1/3)^{|S|} = (4/3)^{2k}.$$

The equality is clear. To see the inequality, notice that since $S \subseteq [2k]$, when viewed as a vector in $\{0,1\}^m$ its last $m-2k$ positions must be 0. So we will instead think of $S$ as a vector in $\{0,1\}^{2k}$, and rewrite $A(S+L)$ as $A'S+AL$, where $A'$ is the first $2k$ columns of the full rank matrix $A$. In particular, $A'$ is a full rank $m \times 2k$ matrix. As we are only concerned with the Hamming weight of $A'S+AL$, we can permute its coordinates and rewrite $A'$ as $[I_{2k}|A'']^T$ for some $2k \times (m-2k)$ matrix $A''$. (Readers who are familiar with linear codes should think of the standard form of a generator matrix.) Moreover, for a lower bound on the Hamming weight, we can restrict our attention to the first $2k$ bits of $A'S+AL$. Hence, we can think of first $2k$ bits of $A'S+AL$ as $S$ shifted by the first $2k$ bits of the fixed vector $AL$. Since we are summing over all $S$ in $\{0,1\}^{2k}$, the shift does not affect the sum, and the inequality follows. Therefore, we have

$$\left|\mathbb{E}[q(D+T\wedge U)] - \mathbb{E}[q(U)]\right| \le 2^{-k}\delta \cdot (4/3)^{2k} \le (8/9)^k \delta,$$

and proving the claim. □

## 7 Proof of Claim 1.12

In this section, we more generally exhibit a distribution $D$ that is $(d^2/10k, d)$-close to uniform for any $d$. One can obtain Claim 1.12 by setting $d = k^{1/3}$. To simplify notation we will switch from $\{0,1\}$ to $\{-1,1\}$, and replace $k$ with $2k$.

We define $D$ to be a random variable distributed uniformly over strings in $\{-1,1\}^{2k}$ with equal number of $-1$'s and 1's.

**Claim 7.1.** *$D$ is $(10d^2/k, d)$-close to uniform for every integer $d$.*

*Proof.* We can assume $d^2 \le k/10$, for otherwise the conclusion is trivial. Let $I \subseteq [k]$ be a subset of size $d$. For every $x \in \{-1,1\}^d$, we have

$$\Pr[D_I = x] = \frac{\binom{2k-d}{k-\mathrm{wt}(x)}}{\binom{2k}{k}},$$

where $\mathrm{wt}(x)$ is the number of $-1$'s in $x$. We bound below the right hand side by

$$\begin{aligned}\frac{\binom{2k-d}{k-d}}{\binom{2k}{k}} &= \frac{k(k-1)\cdots(k-d+1)}{2k(2k-1)\cdots(2k-d+1)}\\ &\geq \left(\frac{k-d+1}{2k}\right)^d\\ &= 2^{-d}\left(1-\frac{d-1}{k}\right)^d\\ &\geq 2^{-d}\left(1-\frac{d(d-1)}{k}\right)\\ &\geq 2^{-d}\cdot(1-d^2/k),\end{aligned}$$

and bound it above by

$$\begin{aligned}\frac{\binom{2k-d}{k-d/2}}{\binom{2k}{k}} &= \frac{(k(k-1)\cdots(k-d/2+1))^2}{2k(2k-1)\cdots(2k-d+1)}\\ &\leq \left(\frac{k}{2k-d+1}\right)^d\\ &= 2^{-d}\left(1+\frac{d-1}{2k-d+1}\right)^d\\ &\leq 2^{-d}\left(1+\sum_{i=1}^{d}\left(\frac{d(d-1)}{2k-d+1}\right)^i\right)\\ &\leq 2^{-d}\left(1+2\cdot\frac{d(d-1)}{2k-d+1}\right)\\ &\leq 2^{-d}\cdot(1+2d^2/k).\end{aligned}$$

The second inequality follows from $(1+a)^d = 1+\sum_{i=1}^{d}\binom{d}{i}a^i \leq 1+\sum_{i=1}^{d}(da)^i$, where $a=(d-1)/(2k-d+1)$. The third inequality is because the geometric sum has ratio $\leq 1/2$ as $d^2\leq k/10$, and so is bounded by twice the first term. Hence, we have $|\Pr[D_I = x]-2^{-d}|\leq 2^{-d}\cdot 2d^2/k$ for every $x\in\{-1,1\}^d$. The claim then follows from summing the inequality over every $x\in\{-1,1\}^d$. □

We now define our product test $f$. For each $j\in\{1,\ldots,2k\}$, define $f_j\colon\{-1,1\}^{2k}\to\mathbb{C}_{\leq 1}$ to be $f_j(x)=\omega^{x_j}$, where $\omega := \mathrm{e}^{-i/\sqrt{2k}}$. Let $f=\prod_{j\leq 2k} f_j$. We now show that for every large enough $k$ we have

$$\Big|\mathbb{E}[f(D+T\wedge U)]-\mathbb{E}[f(U)]\Big| \geq 1/100.$$

We now bound above and below the expectation of $f$ under the distributions $D+T\wedge U$ and $U$ We will use the fact that $1-\theta^2/2\leq\cos\theta\leq 1-2\theta^2/5$ for $\theta\in[-1,1]$. First, we have

$$\mathbb{E}[f(U)] = \prod_{j\leq 2k}\mathop{\mathbb{E}}_{x\sim\{-1,1\}}[\omega^x] = \prod_{j\leq 2k}(\omega+\omega^{-1})/2 = \left(\cos(1/\sqrt{2k})\right)^{2k} \leq (1-1/(5k))^{2k}.$$

Next for every $j \in \{1,2,\ldots,2k\}$, we have

$$\mathop{\mathbb{E}}_{T,U}[f_j(x+T\wedge U)] = \frac{3}{4}\omega^{x_j} + \frac{1}{4}\omega^{-x_j}.$$

Define $\beta\colon \{-1,1\} \to \mathbb{C}_{\le 1}$ to be $\beta(x) := \frac{3}{4}\omega^{x} + \frac{1}{4}\omega^{-x}$. Since $D$ has the same number of $-1$'s and 1's,

$$\begin{aligned}\mathop{\mathbb{E}}_{D}\Big[\prod_{j\le 2k}\beta_j(D)\Big] &= \beta(1)^k\beta(-1)^k\\ &= (10/16+3/16\cdot(\omega^2+\omega^{-2}))^k\\ &= (5/8+3/8\cdot\cos(2/\sqrt{2k}))^k\\ &\ge (5/8+3/8\cdot(1-1/k))^k\\ &= (1-3/(8k))^k,\end{aligned}$$

Therefore $|\mathbb{E}[f(D+T\wedge U]-\mathbb{E}[f(U)]| \ge (1-3/(8k))^k-(1-1/(5k))^{2k} \ge 1/100$, for every sufficiently large $k$, concluding the proof.

The $f_i$ in this proof have variance $\Theta(1/k)$. So this counterexample gives a product test with total variance $O(1)$, and is relevant also to Lemma 2.1. Specifically it shows that for $\ell = 1$ and say $d = O(1)$, the error term $(k2^{\ell})^{O(d)}\varepsilon$ in Lemma 2.1 cannot be replaced with $k^c\varepsilon$ for a certain constant $c$. Moreover, it cannot be replaced even if any $k^{\Omega(1)}$ of the $Y_i$ are close to the $X_i$ (as opposed to just $O(1)$).

# 8 Moment bounds for sum of almost $d$-wise independent variables

In this section we prove some moment bounds and tail bounds for sum of almost $d$-wise independent complex variables.

**Lemma 8.1.** *Let $Z_1,Z_2,\ldots,Z_k \in \mathbb{C}$ be independent random variables with $\mathbb{E}[Z_i]=0$, $|Z_i|<B$. Let $d$ be an even positive integer. Let $W_1,W_2,\ldots,W_k \in \mathbb{C}$ be random variables that are $(\varepsilon,d)$-close to $Z_1,\ldots,Z_k$ with $|W_i|<B$. Then,*

$$\mathbb{E}\left[\Big|\sum_{i=1}^{k}W_i\Big|^d\right] \le 4\cdot 2^d\left(\Big(\sum_i \mathrm{Var}[Z_i]\cdot d\Big)^{1/2}+dB\right)^d + 4(2kB)^d\varepsilon.$$

*Proof.* Note that for any random variable $W \in \mathbb{C}$ we have

$$\begin{aligned}\mathbb{E}\Big[|W|^d\Big] &= \mathbb{E}\Big[\Big(|\Re(W)|^2+|\Im(W)|^2\Big)^{d/2}\Big]\\ &\le \mathbb{E}\Big[\Big(2\max\{|\Re(W)|^2,|\Im(W)|^2\}\Big)^{d/2}\Big]\\ &\le 2^{d/2}\cdot\mathbb{E}\Big[|\Re(W)|^d+|\Im(W)|^d\Big], \qquad (8.1)\end{aligned}$$

and $\mathrm{Var}[W] = \mathrm{Var}[\Re(W)]+\mathrm{Var}[\Im(W)]$. We first prove a bound assuming the $W_i$ are real-valued.

Since $W_1,\dots,W_k$ are $(\varepsilon,d)$-close to $Z_1,\dots,Z_k$, and $d$ is even, we have

$$\begin{aligned}\mathbb{E}\left[\left|\sum_{i=1}^{k} W_i\right|^d\right] &= \mathbb{E}\left[\Big(\sum_i W_i\Big)^d\right]\\ &= \sum_{i_1,\dots,i_d} \mathbb{E}\left[\prod_{j=1}^{d} W_{i_j}\right]\\ &\le \sum_{i_1,\dots,i_d} \mathbb{E}\left[\prod_{j=1}^{d} Z_{i_j}\right] + 2k^dB^d\varepsilon,\end{aligned}$$

because there are $k^d$ products in the sum and we can apply Claim 4.7 to each product, which is bounded by $B^d$.

We now estimate the quantity $\sum_{i_1,\dots,i_d}\mathbb{E}\left[\prod_{j=1}^{d} Z_{i_j}\right]$. We have

$$\sum_{i_1,\dots,i_d} \mathbb{E}\left[\prod_{j=1}^{d} Z_{i_j}\right] = \sum_{m=1}^{d}\sum_{|S|=m}\sum_{\substack{i_1,\dots,i_d\in S:\\ \{i_j\}_j=S}} \mathbb{E}\left[\prod_{j=1}^{d} Z_{i_j}\right].$$

The expectation is zero whenever $Z_{i_j}$ appears only once for some $i_j\in S$. So each $Z_{i_j}$ must appear at least twice. So the expectation is 0 whenever $m>d/2$. As each $Z_i$ is bounded by $B$, each product is bounded by $B^{d-2m}\prod_{j\in S}\mathbb{E}[Z_j^2] = B^{d-2m}\prod_{j\in S}\mathrm{Var}[Z_j]$. For each $S\subseteq[k]$ of size $m$, there are at most $m^d$ such terms. Let $\sigma$ denote $(\sum_{i=1}^{k}\mathrm{Var}[Z_i])^{1/2}$. Then,

$$\begin{aligned}\sum_{i_1,\dots,i_d} \mathbb{E}\left[\prod_{j=1}^{d} Z_{i_j}\right] &\le \sum_{m=1}^{d/2} B^{d-2m}m^d\sum_{|S|=m}\prod_{j\in S}\mathrm{Var}[Z_j]\\ &\le \sum_{m=1}^{d/2} B^{d-2m}m^{d-m}\mathrm{e}^m\sigma^{2m} && \text{(Maclaurin's inequality, see Claim 4.8)}\\ &\le \mathrm{e}^{d/2}\sum_{m=1}^{d/2} B^{d-2m}(d/2)^{d-m}\sigma^{2m}\\ &\le \mathrm{e}^{d/2}(d/2)^dB^d\sum_{m=0}^{d/2}\left(\frac{\sigma^2}{(d/2)B^2}\right)^m\\ &\le \mathrm{e}^{d/2}(d/2)^dB^d\cdot\left(d\left(1+\frac{\sigma^d}{(d/2)^{d/2}B^d}\right)\right) && \Big(\sum_{m=0}^{d/2}\alpha^m \le d(\alpha^0+\alpha^{d/2}), \forall\alpha>0\Big)\\ &\le d\mathrm{e}^{d/2}\Big((d/2)^dB^d + (d/2)^{d/2}\sigma^d\Big)\\ &\le 2\cdot 2^{d/2}(dB+\sigma\sqrt{d})^d.\end{aligned}$$

Hence,

$$\mathbb{E}\left[\left|\sum_{i=1}^{k} W_i\right|^d\right] \le 2\cdot 2^{d/2}\left(dB + \Big(\sum_{i=1}^{k}\mathrm{Var}[Z_i]\cdot d\Big)^{1/2}\right)^d + 2(kB)^d\varepsilon.$$

To handle complex-valued $W_i$, we apply the bound above to $|\sum_{i=1}^k \Re(W_i)|^d$ and $|\sum_{i=1}^k \Im(W_i)|^d$, and plug both bounds in (8.1), giving us

$$\begin{aligned}
&\mathbb{E}\left[\left|\sum_{i=1}^k W_i\right|^d\right] \\
&\le 2\cdot 2^d\left(\left(\sum_{i=1}^k \mathrm{Var}[\Re(Z_i)]\cdot d\right)^{1/2}+dB\right)^d + 2\cdot 2^d\left(\left(\sum_{i=1}^k \mathrm{Var}[\Im(Z_i)]\cdot d\right)^{1/2}+dB\right)^d + 4\cdot 2^{d/2}(kB)^d\varepsilon \\
&\le 4\cdot 2^d\left(\left(\sum_{i=1}^k \mathrm{Var}[Z_i]\cdot d\right)^{1/2}+dB\right)^d + 4(2kB)^d\varepsilon. \qquad \square
\end{aligned}$$

**Lemma 8.2.** *Let $X_1, X_2, \ldots, X_k \in [0,1]$ be independent random variables. Let $d$ be an even positive integer. Let $Y_1, Y_2, \ldots, Y_k \in [0,1]$ be random variables that are $(\varepsilon, d)$-close to $X_1, \ldots, X_k$. Let $Y = \sum_{i\le k} Y_i$ and $\mu = \mathbb{E}[\sum_i X_i]$. Then,*

$$\Pr[|Y-\mu| \ge \delta\mu] \le 4\cdot 2^d \left(\frac{\sqrt{\mu d}+d}{\delta\mu}\right)^d + 4\left(\frac{2k}{\delta\mu}\right)^d \varepsilon.$$

*In particular, if $\mu \ge 25d$, we have* $\Pr[|Y-\mu| \ge \mu/2] \le 2^{-\Omega(d)} + k^d\varepsilon$.

*Proof.* Let $X_i' = X_i - \mathbb{E}[X_i]$, $Y_i' = Y_i - \mathbb{E}[X_i]$ and $Y' = \sum_i Y_i'$. Note that $X_i' \in [-1,1]$ and $\mathbb{E}[X_i'] = 0$. Since $X_i \in [0,1]$, we have

$$\mathbb{E}[X_i] \ge \mathbb{E}[X_i^2] \ge \mathrm{Var}[X_i] = \mathrm{Var}[X_i - \mathbb{E}[X_i]] = \mathrm{Var}[X_i'].$$

By Lemma 8.1 and Markov's inequality,

$$\begin{aligned}
\Pr[|Y-\mu| \ge \delta\mu] &= \Pr[|Y'|^d \ge (\delta\mu)^d] \\
&\le 4\cdot 2^d\left(\frac{(\sum_i \mathrm{Var}[X_i']\cdot d)^{1/2}+d}{\delta\mu}\right)^d + 4\left(\frac{2k}{\delta\mu}\right)^d\varepsilon \\
&\le 4\cdot 2^d\left(\frac{\sqrt{\mu d}+d}{\delta\mu}\right)^d + 4\left(\frac{2k}{\delta\mu}\right)^d\varepsilon,
\end{aligned}$$

where in the last inequality we used $\mu \ge \sum_i \mathrm{Var}[X_i']$. □

**Acknowledgments.** We thank Daniel Kane for answering some questions about [15], and Kavi Rama Murthy for providing a proof of Lemma 4.4. An earlier version of this paper had the bound $\tilde{O}(\ell + \log k)\log(k/\varepsilon)\log k$ in Theorem 1.5. We thank Dean Doron, Pooya Hatami and William Hoza for pointing out an improvement of this bound. We are also very grateful to the anonymous reviewers for their detailed comments which, in addition to correcting several mistakes, also strengthened some of the intermediate claims and simplified their proofs.

## References

[1] MIKLÓS AJTAI, JÁNOS KOMLÓS, AND ENDRE SZEMERÉDI: Deterministic simulation in LOGSPACE. In *Proc. 19th STOC*, pp. 132–140. ACM Press, 1987. [doi:10.1145/28395.28410] 3

[2] MIKLÓS AJTAI AND AVI WIGDERSON: Deterministic simulation of probabilistic constant depth circuits. *Advances in Computing Research*, 5(1):199–222, 1989. Preliminary version in FOCS'85. 4

[3] ROY ARMONI, MICHAEL SAKS, AVI WIGDERSON, AND SHIYU ZHOU: Discrepancy sets and pseudorandom generators for combinatorial rectangles. In *Proc. 37th FOCS*, pp. 412–421. IEEE Comp. Soc. Press, 1996. [doi:10.1109/SFCS.1996.548500] 3

[4] LOUAY M. J. BAZZI: Polylogarithmic independence can fool DNF formulas. *SIAM J. Comput.*, 38(6):2220–2272, 2009. Preliminary version in FOCS'07. [doi:10.1137/070691954] 35

[5] ANDREJ BOGDANOV, PERIKLIS A. PAPAKONSTANTINOU, AND ANDREW WAN: Pseudorandomness for read-once formulas. In *Proc. 52nd FOCS*, pp. 240–246. IEEE Comp. Soc. Press, 2011. [doi:10.1109/FOCS.2011.57] 3

[6] ANDREJ BOGDANOV AND EMANUELE VIOLA: Pseudorandom bits for polynomials. *SIAM J. Comput.*, 39(6):2464–2486, 2010. Preliminary version in FOCS'07. [doi:10.1137/070712109] 2

[7] RAVI BOPPANA, JOHAN HÅSTAD, CHIN HO LEE, AND EMANUELE VIOLA: Bounded independence versus symmetric tests. *ACM Trans. Computation Theory*, 11(4):21:1–27, 2019. Preliminary version in RANDOM'16. [doi:10.1145/3337783] 9

[8] SURESH CHARI, PANKAJ ROHATGI, AND ARAVIND SRINIVASAN: Improved algorithms via approximations of probability distributions. *J. Comput. System Sci.*, 61(1):81–107, 2000. Preliminary version in STOC'94. [doi:10.1006/jcss.1999.1695] 4, 6

[9] ESHAN CHATTOPADHYAY, POOYA HATAMI, OMER REINGOLD, AND AVISHAY TAL: Improved pseudorandomness for unordered branching programs through local monotonicity. In *Proc. 50th STOC*, pp. 363–375. ACM Press, 2018. [doi:10.1145/3188745.3188800] 3

[10] ANINDYA DE: Beyond the central limit theorem: asymptotic expansions and pseudorandomness for combinatorial sums. In *Proc. 56th FOCS*, pp. 883–902. IEEE Comp. Soc. Press, 2015. [doi:10.1109/FOCS.2015.59] 3

[11] ANINDYA DE, OMID ETESAMI, LUCA TREVISAN, AND MADHUR TULSIANI: Improved pseudorandom generators for depth 2 circuits. In *Proc. 14th Internat. Workshop on Randomization and Computation (RANDOM'10)*, pp. 504–517. Springer, 2010. [doi:10.1007/978-3-642-15369-3_38] 4

[12] GUY EVEN, ODED GOLDREICH, MICHAEL LUBY, NOAM NISAN, AND BOBAN VELIČKOVIĆ: Efficient approximation of product distributions. *Random Structures Algorithms*, 13(1):1–16, 1998. Preliminary version in STOC'92. [doi:10.1002/(SICI)1098-2418(199808)13:1<1::AID-RSA1>3.0.CO;2-W] 3, 6

[13] MICHAEL A. FORBES AND ZANDER KELLEY: Pseudorandom generators for read-once branching programs, in any order. In *Proc. 59th FOCS*, pp. 946–955. IEEE Comp. Soc. Press, 2018. [doi:10.1109/FOCS.2018.00093, arXiv:1808.06265] 9, 40, 41

[14] DMITRY GAVINSKY, SHACHAR LOVETT, AND SRIKANTH SRINIVASAN: Pseudorandom generators for read-once ACC$^0$. In *Proc. 27th IEEE Conf. on Computational Complexity (CCC'12)*, pp. 287–297. IEEE Comp. Soc. Press, 2012. [doi:10.1109/CCC.2012.37] 2, 7

[15] PARIKSHIT GOPALAN, DANIEL M. KANE, AND RAGHU MEKA: Pseudorandomness via the discrete Fourier transform. *SIAM J. Comput.*, 47(6):2451–2487, 2018. Preliminary version in FOCS'15. [doi:10.1137/16M1062132, arXiv:1506.04350] 3, 7, 18, 46

[16] PARIKSHIT GOPALAN, RAGHU MEKA, OMER REINGOLD, LUCA TREVISAN, AND SALIL VADHAN: Better pseudorandom generators from milder pseudorandom restrictions. In *Proc. 53rd FOCS*, pp. 120–129. IEEE Comp. Soc. Press, 2012. [doi:10.1109/FOCS.2012.77] 3, 4, 6, 7, 8, 33, 35

[17] PARIKSHIT GOPALAN, RAGHU MEKA, OMER REINGOLD, AND DAVID ZUCKERMAN: Pseudorandom generators for combinatorial shapes. *SIAM J. Comput.*, 42(3):1051–1076, 2013. Preliminary version in STOC'11. [doi:10.1137/110854990] 3, 6

[18] PARIKSHIT GOPALAN, RYAN O'DONNELL, YI WU, AND DAVID ZUCKERMAN: Fooling functions of halfspaces under product distributions. In *Proc. 25th IEEE Conf. on Computational Complexity (CCC'10)*, pp. 223–234. IEEE Comp. Soc. Press, 2010. [doi:10.1109/CCC.2010.29, arXiv:1001.1593] 3

[19] PARIKSHIT GOPALAN AND AMIR YEHUDAYOFF: Inequalities and tail bounds for elementary symmetric polynomial with applications, 2014. [arXiv:1402.3543] 3, 6, 7

[20] ELAD HARAMATY, CHIN HO LEE, AND EMANUELE VIOLA: Bounded independence plus noise fools products. *SIAM J. Comput.*, 47(2):493–523, 2018. Preliminary version in CCC'17. [doi:10.1137/17M1129088] 3, 4, 5, 9, 33, 41

[21] HAMED HATAMI: Lecture notes on Harmonic Analysis of Boolean functions, 2014. Available at author's home page. 33

[22] RUSSELL IMPAGLIAZZO, NOAM NISAN, AND AVI WIGDERSON: Pseudorandomness for network algorithms. In *Proc. 26th STOC*, pp. 356–364. ACM Press, 1994. [doi:10.1145/195058.195190] 3

[23] CHIN HO LEE: Fourier bounds and pseudorandom generators for product tests. In *34th Computational Complexity Conference (CCC'19)*, volume 137, pp. 7:1–25. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2019. [doi:10.4230/LIPIcs.CCC.2019.7] 6

[24] RUDOLF LIDL AND HARALD NIEDERREITER: *Finite Fields*. Cambridge Univ. Press, 1997. [doi:10.1017/CBO9780511525926] 41

[25] SHACHAR LOVETT: Unconditional pseudorandom generators for low-degree polynomials. *Theory of Computing*, 5(3):69–82, 2009. Preliminary version in STOC'08. [doi:10.4086/toc.2009.v005a003] 2

[26] CHI-JEN LU: Improved pseudorandom generators for combinatorial rectangles. *Combinatorica*, 22(3):417–433, 2002. Preliminary version in ICALP'98. [doi:10.1007/s004930200021] 3

[27] MICHAEL LUBY, BOBAN VELIČKOVIĆ, AND AVI WIGDERSON: Deterministic approximate counting of depth-2 circuits. In *2nd Israel Symp. on Theory and Computing Systems (ISTCS'93)*, pp. 18–24. IEEE Comp. Soc. Press, 1993. [doi:10.1109/ISTCS.1993.253488] 2

[28] RAGHU MEKA, OMER REINGOLD, AND AVISHAY TAL: Pseudorandom generators for width-3 branching programs. In *Proc. 51st STOC*. ACM Press, 2019. [doi:10.1145/3313276.3316319, arXiv:1806.04256] 6

[29] JOSEPH NAOR AND MONI NAOR: Small-bias probability spaces: efficient constructions and applications. *SIAM J. Comput.*, 22(4):838–856, 1993. Preliminary version in STOC'90. [doi:10.1137/0222053] 4, 6, 9, 14, 16, 17

[30] NOAM NISAN: Pseudorandom bits for constant depth circuits. *Combinatorica*, 11(1):63–70, 1991. [doi:10.1007/BF01375474] 4

[31] NOAM NISAN: Pseudorandom generators for space-bounded computation. *Combinatorica*, 12(4):449–461, 1992. Preliminary version in STOC'90. [doi:10.1007/BF01305237] 3, 4

[32] NOAM NISAN AND DAVID ZUCKERMAN: Randomness is linear in space. *J. Comput. System Sci.*, 52(1):43–52, 1996. [doi:10.1006/jcss.1996.0004] 3

[33] MICHAEL SAKS AND SHIYU ZHOU: $BP_HSPACE(S) \subseteq DSPACE(S^{3/2})$. *J. Comput. System Sci.*, 58(2):376–403, 1999. Preliminary version in FOCS'95. [doi:10.1006/jcss.1998.1616] 3

[34] ROCCO A. SERVEDIO AND LI-YANG TAN: Improved pseudorandom generators from pseudorandom multi-switching lemmas. In *Proc. 23rd Internat. Workshop on Randomization and Computation (RANDOM'19)*, pp. 45:1–45:23. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019. [doi:10.4230/LIPIcs.APPROX-RANDOM.2019.45] 2

[35] J. MICHAEL STEELE: *The Cauchy-Schwarz Master Class*. Cambridge Univ. Press, 2004. [doi:10.1017/CBO9780511817106] 21

[36] LUCA TREVISAN: Open problems in unconditional derandomization. Slides presented at China Theory Week (CTW'10), 2010. 3, 6

[37] YOAV TZUR: Notions of weak pseudorandomness and $GF(2^n)$-polynomials. Master's thesis, Weizmann Institute of Science, 2009. Link at ECCC. 3

[38] EMANUELE VIOLA: Pseudorandom bits for constant-depth circuits with few arbitrary symmetric gates. *SIAM J. Comput.*, 36(5):1387–1403, 2007. Preliminary version in CCC'05. [doi:10.1137/050640941] 2

[39] EMANUELE VIOLA: The sum of $D$ small-bias generators fools polynomials of degree $D$. *Comput. Complexity*, 18(2):209–217, 2009. Preliminary version in CCC'08. [doi:10.1007/s00037-009-0273-5] 2

[40] EMANUELE VIOLA: Randomness buys depth for approximate counting. *Comput. Complexity*, 23(3):479–508, 2014. Preliminary version in FOCS'11. [doi:10.1007/s00037-013-0076-6] 3

[41] EMANUELE VIOLA: Special topics in complexity theory. Lecture notes, Northeastern Univ., 2017. ECCC. 6

## AUTHORS

Chin Ho Lee
Postdoctoral research fellow
Department of Computer Science
Columbia University
New York, NY
c.h.lee@columbia.edu
https://cs.columbia.edu/~chlee

Emanuele Viola
Associate professor
Khoury College of Computer Sciences
Northeastern University
Boston, MA
viola@ccs.neu.edu
http://www.ccs.neu.edu/home/viola/

## ABOUT THE AUTHORS

CHIN HO LEE just completed his Ph. D. at Northeastern University under the guidance of Emanuele Viola. He is now doing his postdoc in Columbia Unviersity. He recently got married! The picture in the HTML page was taken in Jeju Island, Korea.

EMANUELE VIOLA is contemplating replacing his humidifier with an areca palm.