Bounded independence plus noise fools products

Elad Haramaty* Chin Ho Lee* Emanuele Viola*

May 4, 2017

Abstract

Let D be a b-wise independent distribution over $\{0,1\}^m$. Let E be the "noise" distribution over $\{0,1\}^m$ where the bits are independent and each bit is 1 with probability $\eta/2$. We study which tests $f\colon \{0,1\}^m \to [-1,1]$ are ε-fooled by $D+E$, i.e., $|\operatorname{E}[f(D+E)] - \operatorname{E}[f(U)]| \le \varepsilon$ where U is the uniform distribution.

We show that $D+E$ ε-fools product tests $f : (\{0,1\}^n)^k \to [-1,1]$ given by the product of k bounded functions on disjoint n-bit inputs with error $\varepsilon = k(1-\eta)^{\Omega(b^2/m)}$, where $m = nk$ and $b \ge n$. This bound is tight when $b = \Omega(m)$ and $\eta \ge (\log k)/m$. For $b \ge m^{2/3}\log m$ and any constant η the distribution $D+E$ also 0.1-fools log-space algorithms.

We develop two applications of this type of results. First, we prove communication lower bounds for decoding noisy codewords of length m split among k parties. For Reed–Solomon codes of dimension m/k where $k = O(1)$, communication $\Omega(\eta m) - O(\log m)$ is required to decode one message symbol from a codeword with ηm errors, and communication $O(\eta m \log m)$ suffices. Second, we obtain pseudorandom generators. We can ε-fool product tests $f\colon (\{0,1\}^n)^k \to [-1,1]$ under any permutation of the bits with seed lengths $2n + \tilde{O}(k^2 \log(1/\varepsilon))$ and $O(n) + \tilde{O}(\sqrt{nk \log 1/\varepsilon})$. Previous generators have seed lengths $\ge nk/2$ or $\ge n\sqrt{nk}$. For the special case where the k bounded functions have range $\{0,1\}$ the previous generators have seed length $\ge (n + \log k)\log(1/\varepsilon)$.

*Supported by NSF grant CCF-1319206. Work done in part while at the College of Computer and Information Science, Northeastern University, in part while visiting Harvard University, with support from Salil Vadhan's Simons Investigator grant, and in part while at the Simons Institute for the Theory of Computing.

1 Introduction and our results

At least since the seminal work [CW79] the study of bounded independence has received a lot of attention in theoretical computer science. In particular, researchers have analyzed various classes of tests that cannot distinguish distributions with bounded independence from uniform. Such tests include (combinatorial) rectangles [EGL+98] (cf. [CRS00]), bounded-depth circuits [Baz09, Raz09, Bra10, Tal14], and halfspaces [DGJ+10, GOWZ10, DKN10], to name a few. We say that such tests are *fooled* by distributions with bounded independence.

In this work we consider fooling tests which are a product of several functions on disjoint inputs, and hence are called *product tests*.

Definition 1 (Product tests)**.** *A* product test *with k functions of input length n and alphabet size s is a function $f\colon ([s]^n)^k \to \mathbb{C}_1$ which is the product of k functions $f_1, f_2, \ldots, f_k$ on disjoint inputs, where each f_i maps $[s]^n$ to $\mathbb{C}_1$, the complex unit disk $\{z \in \mathbb{C} : |z| \leq 1\}$.*

We note that these tests make sense already for $n = 1$ and large s (and in fact as we will see have been considered for such parameters in the literature). But it is essential for our applications that the input set of the f_i has a product structure, so we think of n being large. We can choose $s = 2$ for almost all of our results. In this case, each f_i simply has domain $\{0,1\}^n$.

Product tests include as a special case several classes of tests which have been studied in the literature. Specifically, as mentioned in Definition 1, product tests include as a special case the important class of *combinatorial rectangles* [AKS87, Nis92, NZ96, INW94, EGL+98, ASWZ96, Lu02, Vio14, GMR+12, GY14, GKM15].

Definition 2 (Combinatorial rectangles)**.** *A* combinatorial rectangle *is a product test where each f_i has output in $\{0,1\}$.*

Product tests also include as a special case *combinatorial checkerboards* [Wat13], corresponding to functions f_i with range $\{-1,1\}$. More generally, the recent work [GKM15] highlights the unifying role of product tests (which are called Fourier shapes in [GKM15]) by showing that any distribution that fools product tests also fools a number of other tests considered in the literature, including generalized halfspaces [GOWZ10] and combinatorial shapes [GMRZ13, De15]. For the main points in this paper it suffices to consider combinatorial rectangles, but we get broader results working with products.

Bounded independence vs. products. A moment's thought reveals that bounded independence completely fails to fool product tests. Indeed, note that the parity function on $m := nk$ bits is a product test: set $s = 2$ and let each of the k functions compute the parity of their n-bit input, with output in $\{-1,1\}$. However, consider the distribution D which is uniform on $m-1$ bits and has the last bit equal to the parity of the first $m-1$ bits. D has independence $m-1$, which is just one short of maximum. And yet the expectation of parity under D is 1, whereas the expectation of parity under uniform is 0.

This parity counterexample is the simplest example of a general obstacle which has more manifestations. For another example define $g_i := (1 - f_i)/2$, where the f_i are as in the

previous example. Each g_i has range in $\{0,1\}$, and so $\prod g_i$ is a combinatorial rectangle. But the expectations of $\prod_i g_i$ under D and uniform differ by 2^{-k}. This error is too large for the applications in communication complexity and streaming where we have to sum over 2^k rectangles. Indeed, jumping ahead, having a much lower error is critical for our applications. Finally, the obstacle arises even if we consider distributions with small bias [NN93] instead of bounded independence. Indeed, the uniform distribution D over m bits whose inner product modulo 2 is one has bias $2^{-\Omega(m)}$, but inner product is a nearly balanced function which can be written as product, implying that the expectations under D and uniform differ by $1/2 - o(1)$.

The starting point of this work is the observation that all these examples break completely if we perturb just a few bits of D randomly. For parity, it suffices to perturb one bit and the expectation under D will be 0. For inner product, the distance between the expectations shrinks exponentially with the number of perturbed bits.

Our main result is that this is a general phenomenon: If we add a little noise to any distribution with bounded independence, or with small-bias, then we fool product tests with good error bounds. We first state the results for bounded independence. We begin with two definitions which are used extensively in this paper.

Definition 3. *A distribution D over $[s]^m$ is* b-wise independent*, or* b-uniform*, if any b symbols of D are uniformly distributed over $[s]^b$.*

Definition 4 (Noise)**.** *We denote by $E(s,m,\eta)$ the* noise distribution *over $[s]^m$ where the symbols are independent and each of them is set to uniform with probability η and is 0 otherwise. We simply write E when the parameters are clear from the context.*

Theorem 5 (Bounded independence plus noise fools products)**.** *Let $f_1,\ldots,f_k\colon [s]^n \to \mathbb{C}_1$ be k functions with $\mu_i = \mathrm{E}[f_i]$. Set $m := nk$ and let D be a b-uniform distribution over $[s]^m$. Let E be the noise distribution from Definition 4. Write $D = (D_1, D_2, \ldots, D_k)$ where each D_i is in $[s]^n$, and similarly for E. Then*

$$\left| \mathrm{E}\left[\prod_{i\le k} f_i(D_i + E_i) \right] - \prod_{i\le k} \mu_i \right| \le \varepsilon$$

for the following choices:

(1) if $b \ge n$ then $\varepsilon = k(1-\eta)^{\Omega(b^2/m)}$.

(2) if $b < n$ and each D_i is uniform over $[s]^n$ then $\varepsilon = k(1-\eta)^{\Omega(b/k)}$.

(3) if $b < n$ then $\varepsilon = ke^{-\Omega(\eta b/k)} + 2k\binom{n}{n-b}e^{-\Omega(\eta b)}$.

Moreover, there exist f_i and D such that $\varepsilon \ge (1-\eta)^b$. In particular, the bounds are tight up to the constants in the Ω when $b = \Omega(m)$ and $\eta \ge (\log k)/m$.

It is an interesting question whether the bounds are tight even for $b = o(m)$. We stress that the $D_i \in [s]^n$ in this theorem may not even be pairwise independent; only the m *symbols* of D are b-wise independent. We use (1) in most of our applications. Occasionally we use (3) with $b = n-1$, in which case the error bound is $O(nke^{-\Omega(\eta m/k)})$.

Also note that the theorem is meaningful for a wide range of the noise parameter η: we can have η constant, which means that we are perturbing a constant fraction of the symbols, or we can have $\eta = O(1/m)$ which means that we are only perturbing a constant number of symbols, just like in the observation mentioned above. To illustrate this setting, consider for example $k = O(1)$ and $b = n$. We can have an error bound of ε by setting $\eta = c/m$ for a c that depends only on ε.

We now move to our results for small-bias distributions. A distribution D over m bits has bias δ if any parity of the bits (with range $\{-1, 1\}$) has expectation at most δ in magnitude. The following definition extends this to larger alphabets.

Definition 6. *A distribution $D = (D_1, D_2, \ldots, D_m)$ over $[s]^m$ is (b, δ)-biased if, for every nonzero $\alpha \in [s]^m$ with at most b non-zero coordinates we have $|\mathrm{E}_D[\omega^{\sum_i \alpha_i D_i}]| \leq \delta$ where $\omega := e^{2\pi i/s}$. When $b = m$ we simply call D δ-biased.*

In the case of small-bias distribution our bound on the error is a bit more complicated. We state next one possible tradeoff and defer a more general statement to §2.

Theorem 7. *Let $f_1, \ldots, f_k \colon [s]^n \to \mathbb{C}_1$ be k functions with $\mu_i = \mathrm{E}[f_i]$. Assume $\delta \leq s^{-n}$. Let D be an (n, δ)-biased distribution over $[s]^m$. Let E be the noise distribution from Definition 4. Write $D = (D_1, D_2, \ldots, D_k)$ where each D_i is in $[s]^n$, and similarly for E. Then*

$$\left| \mathrm{E}\left[\prod_{i \leq k} f_i(D_i + E_i) \right] - \prod_{i \leq k} \mu_i \right| \leq 2k(1-\eta)^{\Omega(\log(1/\delta)/(k \log sk))} + \sqrt{\delta}.$$

Note that [AGM03] show that a (b, δ)-biased distribution over $\{0, 1\}^m$ is ε-close in statistical distance to a b-uniform distribution, for $\varepsilon = \delta \sum_{i=1}^{b} \binom{m}{i}$. (See [AAK+07] for a similar bound.) One can apply their results in conjunction with Theorem 5 to obtain a result for small-bias distribution, but only if $\delta \leq 1/\binom{m}{b}$. Via a direct proof we derive useful bounds already for $\delta = \Omega(2^{-b})$, and this will be used in §1.2.

We note that summing a noise vector to a string x is equivalent to taking a *random restriction* of x. With this interpretation our results show that on average a random restriction of a product test is a function f' that is simpler in the sense that f' is fooled by any (n, δ)-biased distribution, for certain values of δ. (The latter property has equivalent formulations in terms of the Fourier coefficients of f', see [Baz09].) Thus, our results fall in the general theme "restrictions simplify functions" that has been mainstream in complexity theory since at least the work of Subbotovskaya [Sub61]. For an early example falling in this theme, consider AC^0 circuits. There are distributions with super-constant independence which do not fool AC^0 circuits of bounded depth and polynomial size. (Take the uniform distribution conditioned on the parity of the first log many bits equal to 1, and use the fact that such circuits can compute parity on log many bits.) On the other hand, the *switching lemma* [FSS84, Ajt83, Yao85, Hås87, Hås14, IMP12] shows that randomly restricting all but a 1/polylog fraction of the variables collapses the circuit to a function that depends only on $c = O(1)$ variables, and such a function is fooled by any c-wise independent distribution.

Thus, adding noise dramatically reduces the amount of independence that is required to fool AC^0 circuits. For a more recent example, Lemma 7.2 in [GMR+12] shows that for a special case of AC^0 circuits – read-once CNF – one can restrict all but a constant fraction of the variables and then the resulting function is fooled by any ε-bias distribution for a certain $\varepsilon = 1/n^{\omega(1)}$ which is seen to be larger than the bias that would be required had we not applied a restriction.

We are not aware of prior work which applies to arbitrary functions as in our theorems. Another difference between our results and all the previous works that we are aware of lies in the parameter η. In previous works η is large, in particular $\eta = \Omega(1)$, which corresponds to restricting many variables. We can instead set η arbitrarily, and this flexibility is used in both of our applications.

1.1 Application: The complexity of decoding

Error-correcting codes are a fundamental concept with myriad applications in computer science. It is relevant to several of these, and perhaps also natural, to ask what is the complexity of basic procedures related to error-correcting codes. In this paper we focus on *decoding*. The question of the complexity of decoding has already been addressed in [BYRST02, BM05, Gro06]. However, all previous lower bounds that we are aware of are perhaps not as strong as one may hope. First, they provide no better negative results for decoding than for *encoding*. But common experience shows that decoding is much harder! Second, they do not apply to decision problems, but only to multi-output problems such as computing the entire message. Third, they apply to small-space algorithms but not to stronger models such as communication protocols.

In this work we obtain new lower bounds for decoding which overcome these limitations. First, we obtain much stronger bounds for decoding than for encoding. For example, we prove below that decoding a message symbol from Reed–Solomon codeword of length q with $\Omega(q)$ errors requires $\Omega(q)$ communication. On the other hand, encoding is a linear map, and so one can compute any symbol with just $O(\log q)$ communication (or space). This exponential gap may provide a theoretical justification for the common experience that decoding is harder than encoding. Second, our results apply to decision problems. Third, our results apply to stronger models than space-bounded algorithms. Specifically, our lower bounds are proved in the k-party "number-in-hand" communication complexity model, where each of k collaborating parties receives a disjoint portion of the input. The parties communicate by broadcast (a.k.a. writing on a blackboard). For completeness we give next a definition. Although we only define deterministic protocols, our lower bounds in fact bound the *correlation* between such protocols and the hard problem, and so also hold for distributions of protocols (a.k.a. allowing the parties to share a random string).

Definition 8 (Number-in-hand protocols)**.** *A k-party number-in-hand, best-partition, communication protocol for a function $f\colon [s]^m \to Y$, where k divides m, is given by a partition of m into k sets $S_1, S_2, \ldots, S_k$ of equal size m/k and a binary tree. Each internal node v of the tree is labeled with a set $S_v \in \{S_1, S_2, \ldots, S_k\}$ and a function $f_v\colon [s]^{m/k} \to \{0,1\}$, and*

has two outgoing edges labeled 0 *and* 1*. The leaves are labeled with elements from* Y*. On input* $x \in [s]^m$ *the protocol computes* $y \in Y$ *following the root-to-leaf path where from node* v *we follow the edge labeled with the value of* f_v *on the* m/k *symbols of* x *corresponding to* S_v*. The communication cost of the protocol is the depth of the tree.*

Note that we insisted that k divides m, but all the results can be generalized to the case when this does not hold. However this small additional generality makes the statements slightly more cumbersome, so we prefer to avoid it. Jumping ahead, for Reed–Solomon codes this will mean that the claims do not apply as stated to prime fields (but again can be modified to apply to such fields).

Again for completeness, we give next a definition of space-bounded algorithms. For simplicity we think of the input as being encoded in bits.

Definition 9 (One-way, bounded-space algorithm)**.** *A* width-W (a.k.a. space-$\log W$) one-way algorithm *(or* branching program *or* streaming algorithm*) on* m *bits consists of a layered directed graph with* $m+1$ *layers. Each layer has* W *nodes, except the first layer, which has* 1 *node, and the last layer, which has* $2W$*. Each node in layer* $i \leq m$ *has two edges, labeled with* 0 *and* 1*, connecting to nodes in layer* $i+1$*. Each node on layer* $m+1$ *is labeled with an output element. On an* m*-bit input, the algorithm follows the path corresponding to the input, reading the input in a one-way fashion (so layer* i *reads the* i*-th input bit), and then outputs the label of the last node.*

We note that a space-s one-way algorithm can be simulated by a k-party protocol with communication sk. Thus our negative results apply to space-bounded algorithms as well. In fact, this simulation only uses one-way communication and a fixed partition (corresponding to the order in which the algorithm reads the input). But our communication lower bounds hold even for two-way communication and for any partition of the input into k parties, as in Definition 8.

Our lower bound holds when the uniform distribution over the code is b-uniform.

Definition 10. *A code* $C \subseteq \mathbb{F}_q^m$ *is* b*-uniform if the uniform distribution over* C *is* b*-uniform.*

The following standard fact relates the above definition to the *dual distance* of the code.

Fact 11. *Let* X *be the uniform distribution over a linear code* $C \subseteq \mathbb{F}_q^m$*. Then* X *is* d*-wise independent if and only if the dual of* C *has minimum distance* $\geq d+1$*.*

We state next a lower bound for distinguishing a noisy codeword from uniform. The "-1" in the assumption on b will be useful later.

Theorem 12 (Distinguishing noisy codewords from uniform is hard)**.** *Let* $C \subseteq \mathbb{F}_q^m$ *be a* b*-uniform code. Let* E *be the noise distribution from Definition 4. Let* k *be an integer dividing* m *such that* $b \geq m/k - 1$*. Let* $P \colon \mathbb{F}_q^m \to \{0,1\}$ *be a* k*-party protocol using* c *bits of communication. Then*

$$|\Pr[P(C+E)=1] - \Pr[P(U)=1]| \leq \varepsilon \quad \text{for} \quad \varepsilon = 2^{c+\log(m)+O(1)-\Omega(\eta b^2/m)},$$

where C *and* U *denote the uniform distributions over the code* C *and* $\mathbb{F}_q^m$ *respectively.*

We now make some remarks on this theorem. First, we note that a (ck)-party protocol can be simulated by a k-party protocol, so in this sense the lower the number of parties the stronger the lower bound. Also, the smallest number of parties to which the theorem can apply is $k = m/b$, because for $k = m/b - 1$ one can design b-uniform codes such that the distribution $C + E$ can be distinguished well from uniform by just one party, cf. §A. And our lower bound applies for that number. The theorem is non-trivial whenever $b = \omega(\sqrt{m})$, but we illustrate it in the setting of $b = \Omega(m)$ which is typical in coding theory as we are also going to discuss. In this setting we can also set $k = m/b = O(1)$. Hence the communication lower bound is

$$c \geq \Omega(\eta m)$$

when $\eta \geq C \log m/m$ for a universal constant C. When $\eta = \Omega(1)$ this becomes $\Omega(m)$. Note that this bound is within an $O(\log q)$ factor of the bit-length of the input, which is $O(m \log q)$, and within a constant factor if $q = O(1)$.

We prove an essentially matching upper bound in terms of η, stated next. The corresponding distinguisher is a simple variant of *syndrome decoding* which we call "truncated syndrome decoding." It can be implemented as a small-space algorithm with one-sided error, and works even against adversarial noise. So the theorems can be interpreted as saying that syndrome decoding uses an optimal amount of space. We denote by $V(t)$ the volume of the q-ary Hamming ball in m dimensions of radius t, i.e., the number of $x \in \mathbb{F}_q^m$ with at most t non-zero coordinates.

Theorem 13 (Truncated syndrome decoding)**.** *Let $C \subseteq \mathbb{F}_q^m$ be a linear code with dimension d. Given t and $\delta > 0$ define $s := \lceil \log_q(V(t)/\delta) \rceil$. If $d \leq m - s$ there is a one-way algorithm A that runs in space $s \log q$ such that*

(1) for every $x \in C$ and for every e of Hamming weight $\leq t$, $A(x + e) = 1$, and

(2) $\Pr[A(U) = 1] \leq \delta$, where U is uniform in $\mathbb{F}_q^m$.

Moreover, the space bound $s \log q$ is at most $O(t \log(mq/t)) + \log 1/\delta$.

Note that when $t = O(\eta m)$ and δ is constant the space bound is $O(\eta m \log(q/\eta))$, which matches our $\Omega(\eta m)$ lower bound up to the $O(\log(q/\eta))$ factor.

These results in particular apply to Reed–Solomon codes. Recall that a Reed–Solomon code of dimension b is the linear code where a message in $\mathbb{F}_q^b$ is interpreted as a polynomial p of degree $b - 1$ and encoded as the q evaluations of p over any element in the field. (In some presentations, the element 0 is excluded.) Such a code is b-uniform because for any b points (x_i, y_i) where the x_i's are different, there is exactly one polynomial p of degree $b - 1$ such that $p(x_i) = y_i$ for every i.

For several binary codes $C \subseteq \mathbb{F}_2^m$ and constant η we can obtain a communication lower bound of $\Omega(m)$ which is tight up to constant factors. This is true for example for random, linear codes (with bounded rate). The complexity of decoding such codes is intensely studied, also because the assumed intractability of their decoding is a basis for several cryptographic applications. See for example [BJMM12], a slight improvement on the running time which already has more than 100 citations. We also obtain a tight lower bound of $\Omega(m)$ for several explicitly-defined binary codes. For example, we can pick an explicit binary code $C \subseteq \mathbb{F}_2^m$

which is $\Omega(m)$-uniform and that can be decoded in polynomial time for a certain constant noise parameter η (with high probability), see [Shp09] for a construction.

Lower bounds for decoding one symbol. The lower bound in Theorem 12 is for the problem of distinguishing noisy codewords from uniform. Intuitively, this is a strong lower bound saying that no bit of information can be obtained from a noisy codeword. We next use this result to obtain lower bounds for decoding one symbol of the message given a noisy codeword. Some care is needed because some message symbols may just be copied in the codeword. This would allow one party to decode those symbols with no communication, even though the noisy codeword may be indistinguishable from uniform. The lower bound applies to codes that remain b-uniform even after fixing some input symbol. For such codes, a low-communication protocol cannot decode that symbol significantly better than by guessing at random.

Theorem 14. *Let $C' \subseteq \mathbb{F}_q^m$ be a linear code with an $m \times r$ generator matrix G. Let $i \in \{1, 2, \ldots, r\}$ be an index, and let C be the code defined as $C := \{Gx \mid x_i = 0\}$. Let E be the noise distribution from Definition 4. Let k be an integer. Suppose that C is b-uniform for $b \geq m/k - 1$. Let $P \colon \mathbb{F}_q^m \to \mathbb{F}_q$ be a k-party protocol using c bits of communication. Then*

$$\Pr[P(GU + E) = U_i] \leq 1/q + \varepsilon,$$

where $U = (U_1, U_2, \ldots, U_r)$ is the uniform distribution and ε is as in Theorem 12.

We remark that whether C is b-uniform in general depends on both G and i. For example, let C' be a Reed–Solomon code of dimension $b = m/k$. Recall that C' is b-uniform. Note that if we choose $i = 0$ (corresponding to the evaluation of the polynomial at the point $0 \in \mathbb{F}_q$, which as we remarked earlier is a point we consider) then C has a fixed symbol and so is not even 1-uniform. On the other hand, if $i = b - 1$ then we obtain a Reed–Solomon code with dimension $b - 1$, which is $(b - 1)$-uniform, and the lower bound in Theorem 14 applies.

We again obtain an almost matching upper bound. In fact, the corresponding protocol recovers the entire message.

Theorem 15 (Recovering messages from noisy codewords)**.** *Let $C \subseteq \mathbb{F}_q^m$ be a code with distance d. Let t be an integer such that $2t < d$, and let k be an integer dividing q.*

There is a k-party protocol $P \colon \mathbb{F}_q^m \to \mathbb{F}_q^b$ communicating $\max\{m - d + 2t + 1 - m/k, 0\} \lceil \log_2 q \rceil$ bits such that for every $x \in C$ and every e of Hamming weight $\leq t$, $P(x + e) = x$.

A Reed–Solomon code with dimension b has distance $d = m - b + 1$. Hence we obtain communication $\max\{b + 2t - m/k, 0\} \lceil \log_2 q \rceil$, for any t such that $2t < m - b + 1$. This upper bound matches the lower bound in Theorem 14 up to a $\log q$ factor. For example, when $k = O(1)$ and $b = q/k$ our upper bound is $O(\eta q \log q)$ and our lower bound is $\Omega(\eta q) - O(\log q)$.

1.2 Application: Pseudorandomness

The construction of explicit *pseudorandom generators* against restricted classes of tests is a fundamental challenge that has received a lot of attention at least since the 80's, cf. [AW89, AKS87]. One class of tests extensively considered in the literature is concerned with algorithms that read the input bits in a *one-way fashion in a fixed order*. A leading goal is to prove RL=L by constructing generators with logarithmic seed length that fool one-way, space-bounded algorithms, but here the seminal papers [Nis92, INW94, NZ96] remain the state of the art and have larger seed lengths. However, somewhat better generators have been obtained for several special cases, including for example combinatorial rectangles [AKS87, Nis92, NZ96, INW94, EGL+98, ASWZ96, Lu02, Vio14, GMR+12, GY14], combinatorial shapes [GMRZ13, De15, GKM15], and product tests [GKM15]. In particular, for combinatorial rectangles $f\colon (\{0,1\}^n)^k \to \{0,1\}$ two incomparable results are known. For context, the minimal seed length up to constant factors is $O(n + \log(k/\varepsilon))$. One line of research culminating in [Lu02] gives generators with seed length $O(n + \log k + \log^{3/2}(1/\varepsilon))$. More recently, [GMR+12] (cf. [GY14]) improve the dependence on ε while making the dependence on the other parameters a bit worse: they achieve seed length $O((\log n)(n + \log(k/\varepsilon))) + O(\log(1/\varepsilon)\log\log(1/\varepsilon)\log\log\log(1/\varepsilon))$. The latter result is extended to products in [GKM15] (with some other lower-order losses).

Recently there has been considerable interest in extending tests by allowing them to read the bits in *any order*: [BPW11, BPW12, IMZ12, RSV13, SVW14]. This extension is significantly more challenging, and certain instantiations of generators against one-way tests are known to fail [BPW11].

We contribute new pseudorandom generators that fool product tests in any order.

Definition 16 (Fooling)**.** *A generator $G\colon \{0,1\}^\ell \to \{0,1\}^m$* ε-fools *(or* fools with error ε*) a class T of tests on m bits if for every function $f \in T$ we have $|\mathrm{E}[f(G(U_\ell)) - \mathrm{E}[f(U_m)]| \le \varepsilon$, where U_ℓ and U_m are the uniform distributions on ℓ and m bits respectively. We call ℓ the seed length of G. We call G explicit if it is computable in time polynomial in m.*

Definition 17 (Any order)**.** *We say that a generator $G\colon \{0,1\}^\ell \to \{0,1\}^m$ ε-fools a class T of tests* in any order *if for every permutation π on m bits the generator $\pi \circ G\colon \{0,1\}^\ell \to \{0,1\}^m$ ε-fools T.*

The next theorem gives some of our generators. The notation $\tilde{O}()$ hides logarithmic factors in k and n. In this section we only consider alphabet size $s = 2$. We write the range $\{0,1\}^{nk}$ of the generators as $(\{0,1\}^n)^k$ to indicate the parameters of the product tests.

Theorem 18 (PRG for any-order products, I)**.** *There exist explicit pseudorandom generators $G\colon \{0,1\}^\ell \to (\{0,1\}^n)^k$ that ε-fool product tests in any order, with the following seed lengths:*
(1) $\ell = 2n + O(k^2 \log k \log(k/\varepsilon)\log n) = 2n + \tilde{O}(k^2\log(1/\varepsilon))$, and
(2) $\ell = O(n) + O(n^{2/3}(k^2\log k\log(k/\varepsilon)\log n)^{1/3}) = O(n) + \tilde{O}((nk)^{2/3}\log^{1/3}(1/\varepsilon))$.
Moreover, the generators' output has the form $D+E'$, where D is a small-bias distribution and E' is statistically close to a noise vector.

One advantage of these generators is their simplicity. Constructions in the literature tend to be somewhat more involved. In terms of parameters, we note that when $k = O(1)$ we achieve in (1) seed length $\ell = 2n + O(\log 1/\varepsilon) \log n$, which is close to the value of $n + O(\log 1/\varepsilon)$, which is optimal even for the case of fixed order and $k = 2$. Our result is significant already for $k = 3$, but not for $k = 2$. In the latter case the seed length of $(2 - \Omega(1))n$ obtained in [BPW11] remains the best known. For $k \geq \sqrt{n}$ our generator in (2) has polynomial stretch, using a seed length $\tilde{O}(m^{2/3})$ for output length m.

We note that for the special case of combinatorial rectangles $f\colon (\{0,1\}^n)^k \to \{0,1\}$ a pseudorandom generator with seed length $O((n + \log k) \log(1/\varepsilon))$ follows from previous work. The generator simply outputs m bits such that any $d \cdot n$ of them are $1/k^d$ close to uniform in statistical distance, where $d = c \log(1/\varepsilon)$ for an appropriate constant c. Theorem 3 in [AGHP92] shows how to generate these bits from a seed of length $O(n \log(1/\varepsilon) + \log \log m + \log(1/\varepsilon) \log k) = O((n + \log k) \log(1/\varepsilon))$. The analysis of this generator is as follows. The induced distribution on the outputs of the f_i is a distribution on $\{0,1\}^k$ such that any d bits are $1/k^d$ close to the distribution of independent variables whose expectations are equal to the $\mathrm{E}[f_i]$. Now Lemma 5.2 in [CRS00] (cf. [EGL+92]) shows that the probability that the And of the output is 1 equals the product of the expectations of the f_i plus an error which is $\leq 2^{-\Omega(d)} + d\binom{k}{d}/k^d \leq \varepsilon$. However this generator breaks down if the output of the functions is $\{-1, 1\}$ instead of $\{0, 1\}$. Moreover, its parameters are incomparable with those in Theorem 18.(I). In particular, for $k = O(1)$ its seed length is $\geq n \log(1/\varepsilon)$, while as remarked above we achieve $O(n + \log(n) \log(1/\varepsilon))$.

We are able to improve the seed length of (2) in Theorem 18 to $\tilde{O}(\sqrt{m})$, but then the resulting generator is more complicated and in particular it does not output a distribution of the form $D + E'$. For this improvement we "derandomize" our theorems 5 and 7 and then combine them with a recursive technique originating in [GMR+12] (cf. [AW89]) and used in several subsequent works including [RSV13, SVW14, CSV15]. Our context and language are somewhat different from previous work, and this fact may make this paper useful to readers who wish to learn the technique.

Theorem 19 (PRG for any-order products, II)**.** *There exists an explicit pseudorandom generator $G\colon \{0,1\}^\ell \to (\{0,1\}^n)^k$ that ε-fools product tests in any order and seed length $\ell = O(n + \sqrt{nk \log k \log(k/\varepsilon)}) = O(n) + \tilde{O}(\sqrt{nk \log 1/\varepsilon})$.*

Recall that for $b = n$ the error bound in our Theorem 5 is $k(1 - \eta)^{\Omega(b/k)}$, and that it is open whether the exponent can be improved to $\Omega(b)$. We show that if such an improvement is achieved for the derandomized version of the theorem (stated later in Theorem 37) then one would get much better seed length: $\ell = O((n + \log k \log(m/\varepsilon)) \log m)$.

Reingold, Steinke, and Vadhan [RSV13] give a generator that ε-fools width-W space algorithms on m bits in any order, with seed length $\ell = \tilde{O}(\sqrt{m} \log(W/\varepsilon))$. Every combinatorial rectangle $f\colon (\{0,1\}^n)^k \to \{0,1\}$ can be computed by a one-way algorithm with width $2^{n-1} + 1$ on $m = nk$ bits. Hence they also get seed length $\tilde{O}(\sqrt{nk}(n + \log 1/\varepsilon))$ for combinatorial rectangles. Our Theorem 19 improves upon this by removing a factor of n.

Going in the other direction, if D is a distribution on $(\{0,1\}^n)^k$ bits that ε-fools combinatorial rectangles, then D also fools width-W one-way algorithms on $m = nk$ bits with error

$W^k\varepsilon$. Using this we obtain from Theorem 5 a new class of distributions that fools space, namely any distribution that is the sum of a distribution with high-enough independence (or small enough bias) and suitable noise. We state one representative result.

Corollary 20 (Bounded independence plus noise fools space)**.** *Let D be a b-uniform distribution on m bits. Let E be the noise distribution from Definition 4. If $b \geq m^{2/3} \log m$ and η is any constant then $D + E$ fools $O(\log m)$-space algorithms in any order with error $o(1)$.*

As mentioned earlier, [GKM15] show that if a generator fools products then it also fools several other computational models, with some loss in parameters. As a result, we obtain generators for the following two models, extended to read bits in any order.

Definition 21 (Generalized halfspaces and combinatorial shapes)**.** *A* generalized halfspace *is a function $h\colon (\{0,1\}^n)^k \to \{0,1\}$ defined by $h(x) := 1$ if and only if $\sum_{i \leq k} g_i(x_i) \geq \theta$, where $g_1, \ldots, g_k\colon \{0,1\}^n \to \mathbb{R}$ are arbitrary functions and $\theta \in \mathbb{R}$.*

A combinatorial shape *is a function $f\colon (\{0,1\}^n)^k \to \{0,1\}$ defined by $f(x) := g(\sum_{i \leq k} g_i(x_i))$, where $g_1, \ldots, g_k\colon \{0,1\}^n \to \{0,1\}$ and $g\colon \{0, \ldots, k\} \to \{0,1\}$ are arbitrary functions.*

Theorem 22 (PRG for generalized halfspaces and combinatorial shapes, in any-order)**.** *There exists an explicit pseudorandom generator $G\colon \{0,1\}^\ell \to (\{0,1\}^n)^k$ that ε-fools both generalized halfspaces and combinatorial shapes in any order with seed length $\ell = \tilde{O}(n\sqrt{k} + \sqrt{nk \log(1/\varepsilon)})$.*

Note that for $\varepsilon = 2^{-O(n)}$ the seed length simplifies to $\tilde{O}(n\sqrt{k})$.

An original motivation for this work is the study of the sum of small-bias distributions [LV]. However the relationship between the results in this work and [LV] is somewhat technical, applying only to certain settings of parameters. Hence we defer the discussion to §6.

1.3 Techniques

We now give an overview of the proof of Theorem 5. The natural high-level idea, which our proof adopts as well, is to apply Fourier analysis and use noise to bound high-degree terms and independence to bound low-degree terms. Part of the difficulty is finding the right way to decompose the product $\prod_{i \leq k} f_i$. We proceed as follows. For a function f let f^H be its "high-degree" Fourier part and f_L be its "low-degree" Fourier part, so that $f = f^H + f^L$. Our goal is to go from $\prod f_i$ to $\prod f_i^L$. The latter is a product of low-degree functions and hence has low degree. Therefore, its expectation will be close to $\prod_i \mu_i$ by the properties of the distribution D; here we do not use the noise E.

To move from $\prod f_i$ to $\prod f_i^L$ we pick one f_j and we decompose it as $f_j^H + f_j^L$. Iterating this process we indeed arrive to $\prod f_i^L$, but we also obtain k extra terms of the form

$$f_1 f_2 \ldots f_{j-1} f_j^H f_{j+1}^L f_{j+2}^L \ldots f_k^L$$

for $j = 1, \ldots, k$. We show that each of these terms is close to 0 thanks to the presence of the high-degree factor f_j^H. Here we use both D and E.

We conclude this section with a brief technical comparison with the recent papers [GMR$^+$12, GY14, GKM15] which give generators for combinatorial rectangles (and product tests). We note that the generators in those papers only fool tests $f = f_1 \cdot f_2 \cdots f_k$ that read the input in a fixed order (whereas our results allow for any order). Also, they do not use noise, but rather hash the functions f_i in a different way. Finally, a common technique in those papers is, roughly speaking, to use hashing to *reduce the variance of the functions*, and then show that bounded independence fools functions with small variance. We note that the noise parameters we consider in this work are too small to be used to reduce the variance. Specifically, for a product test f those papers define a new function $g = g_1 \cdot g_2 \cdots g_k$ which is the average of f over t independent inputs. While g has the same expectation as f, the variance of each g_i is less than that of f_i by a factor of t. Their goal is to make the variance of each g_i less than $1/k$ so that the sum of the variances is less than 1. In order to achieve this reduction with noise we would have to set $\eta \geq 1 - 1/\sqrt{k}$. This is because if f_i simply is $(-1)^x$ where x is one bit, then the variance of f_i perturbed by noise is $\mathrm{E}_x[\mathrm{E}_E^2[(-1)^{x+E}]] - \mathrm{E}_{x,E}[(-1)^{x+E}] = \mathrm{E}_{x,E,E'}[(-1)^{E+E'}] = (1-\eta)^2$.

Organization. In §2 we prove our main theorems, 5 and 7. In §3 we give the proof details for the results in §1.1. The details for the results in §1.2 are spread over three sections. In §4 we prove Theorem 18. In §5 we prove Theorem 19, and discuss the potential improvement. In §6 we prove Theorem 22, and discuss the relationship between this paper and an original motivation [LV]. We conclude in §7. In §A we include for completeness a lower bound on the values of b and η for which Theorem 5 can apply.

2 Bounded independence plus noise fools products

In this section we prove Theorem 5 and Theorem 7. They both follow easily from the next theorem which is the main result in this section.

Theorem 23. *Let $t \in [0, n]$. Let $f_1, \ldots, f_k \colon [s]^n \to \mathbb{C}_1$ be k functions with $\mu_i = \mathrm{E}[f_i]$. Let D be a (b, δ)-biased distribution over $[s]^m$ for $b \geq \max\{n, 2(k-1)t\}$. Let E be the noise distribution from Definition 4. Write $D = (D_1, D_2, \ldots, D_k)$ where each D_i is in $[s]^n$, and similarly for E. Then*

$$\left| \mathrm{E}\left[\prod_{i \leq k} f_i(D_i + E_i) \right] - \prod_{i \leq k} \mu_i \right| \leq k(1-\eta)^t \sqrt{(1 + s^n\delta)(1 + V(t)^{k-1}\delta)} + V(t)^{k/2}\delta.$$

Let us quickly derive Theorem 5 and 7 in the introduction.

Proof of Theorem 5. Setting $\delta = 0$ and $t = b/2(k-1)$ in Theorem 23 gives the bound

$$k(1-\eta)^{b/2(k-1)} \qquad (\star)$$

which proves the theorem in the case $n \le b = O(n)$.

To prove (1) we need to handle larger b. For this, let $c := \lfloor b/n \rfloor$, and group the k functions into $k' \le k/c + 1$ functions on input length $n' := cn$. Note that $b \ge n'$, and so we can apply $(\star)$ to

$$k'(1-\eta)^{\Omega(b/k')} \le k(1-\eta)^{\Omega(b^2/kn)}.$$

To prove (2) one can observe that in the proof of $(\star)$ the condition $b \ge n$ is only used to guarantee that each D_i is uniform. The latter is now part of our assumption.

To prove (3) view the noise vector E as the sum of two noise vectors E' and E'' with parameter α such that $1-\eta = (1-\alpha)^2$. Note this implies $\alpha = \Omega(\eta)$. If E' sets to uniform at least $n-b$ coordinates in each function then we can apply $(\star)$ to functions on $\le b$ symbols with η replaced by α. The probability that E' does not set to uniform that many coordinates is at most

$$k\binom{n}{n-b}(1-\alpha)^b \le k\binom{n}{n-b}e^{-\Omega(\eta b)},$$

and in that case the distance between the expectations is at most two.

To show the "moreover" part let the f_i compute parity on the first $b+1$ bits, and let D be the b-wise independent distribution which is uniform on strings whose parity of the $b+1$ bits is 0. The other bits are irrelevant. The expectation of parity under uniform is 0. The expectation of parity under D is 1 if no symbol is perturbed with noise, and is 0 otherwise. Hence the error is $\ge (1-\eta)^{b+1}$. In particular, if $b = \Omega(m)$ then an upper bound on the error of the form $k(1-\eta)^{cm}$ is false for sufficiently large c, using that $\eta \ge (\log k)/m$. □

Proof of Theorem 7. Let $c := \lfloor \sqrt{\log(1/\delta)/(n \log s)} \rfloor$. Note that $c \ge 1$ because $\delta \le s^{-n}$. We group the k functions into $k' = \lceil k/c \rceil$ functions on input length $n' := cn$. The goal is to make $s^{n'} \approx 1/\delta$. By Claim 33, $V_{n'}(t) \le (en's/t)^t$. Hence $V_{n'}(t)^{k'/2} \le V_{n'}(t)^{k'-1} \le (en's/t)^{k't}$. Now let $t = \alpha n' \log s/(k' \log sk')$ for a small constant $\alpha > 0$ so that the latter bound is $\le s^{n'/2} \approx 1/\sqrt{\delta}$.

The error bound in Theorem 23 now becomes at most

$$k(1-\eta)^t(1+s^{n'}\delta) + s^{n'/2}\delta.$$

And so the bound is at most

$$2k(1-\eta)^{\Omega(\log(1/\delta)/(k \log sk))} + \sqrt{\delta}.$$ □

We now turn to the proof of Theorem 23. We begin with some preliminaries.

2.1 Preliminaries

Denote by U the uniform distribution. Let s be any positive integer. We write $[s]$ for $\{0, 1, 2, \ldots, s-1\}$. Let $\omega := e^{2\pi i/s}$ be a primitive s-th root of unity. For any $\alpha \in [s]^u$, we define $\chi_\alpha(x)\colon [s]^u \to \mathbb{C}$ to be

$$\chi_\alpha(x) := \omega^{\langle \alpha, x \rangle},$$

where α and x are viewed as vectors in $\mathbb{Z}_s^u$ and $\langle \alpha, x \rangle := \sum_i \alpha_i x_i$.

For any function $f\colon [s]^u \to \mathbb{C}$, its Fourier expansion is

$$f(x) := \sum_{\alpha \in [s]^u} \hat{f}_\alpha \chi_\alpha(x),$$

where $\hat{f}_\alpha \in \mathbb{C}$ is given by

$$\hat{f}_\alpha := \mathop{\mathrm{E}}_{x \sim [s]^u}[f(x)\overline{\chi_\alpha(x)}].$$

Here and elsewhere, random variables are uniformly distributed unless specified otherwise.

The Fourier L_1*-norm* of f is defined as $\sum_\alpha |\hat{f}_\alpha|$, and is denoted by $L_1[f]$. The *degree* of f is defined as $\max\{|\alpha| : \hat{f}_\alpha \neq 0\}$, where $|\alpha|$ is the number of nonzero coordinates of α, and is denoted by $\deg(f)$. Note that we have $L_1[f] = L_1[\overline{f}]$. The following fact bounds the L_1-norm and degree of product functions.

Fact 24. *For any two functions $f, g\colon [s]^u \to \mathbb{C}$, we have*
(1) $\deg(fg) \leq \deg(f) + \deg(g)$, and
(2) $L_1[fg] \leq L_1[f]L_1[g]$.

Proof. We have

$$f(x)g(x) = \left(\sum_{\alpha \in [s]^n} \hat{f}_\alpha \chi_\alpha(x)\right)\left(\sum_{\beta \in [s]^n} \hat{g}_\beta \chi_\beta(x)\right) = \sum_{\alpha,\beta} \hat{f}_\alpha \hat{g}_\beta \chi_{\alpha+\beta}(x) = \sum_\alpha \left(\sum_\beta \hat{f}_{\alpha-\beta}\hat{g}_\beta\right)\chi_\alpha(x).$$

Hence the α-th Fourier coefficient of $f \cdot g$ is $\sum_\beta \hat{f}_{\alpha-\beta}\hat{g}_\beta$.

To see (1), note that in the latter expression the sum over β can be restricted to those β with $|\beta| \leq \deg(g)$. Now note that if $|\alpha| > \deg(f) + \deg(g)$ then $|\alpha - \beta| > \deg(f)$ and hence $\hat{f}_{\alpha-\beta}$ will be zero for every β.

To show (2) write $L_1[fg] = \sum_\alpha |\sum_\beta \hat{f}_{\alpha-\beta}\hat{g}_\beta| \leq \sum_{\alpha,\beta} |\hat{f}_{\alpha-\beta}||\hat{g}_\beta| = (\sum_\alpha |\hat{f}_\alpha|)(\sum_\beta |\hat{g}_\beta|) = L_1[f]L_1[g]$. □

Fact 25 (Parseval's identity)**.** $\sum_{\alpha \in [s]^n} |\hat{f}_\alpha|^2 = \mathrm{E}_{x \sim [s]^n}[|f(x)|^2]$. *In the case of $f \in \mathbb{C}_1$, this quantity is at most* 1.

Proof.

$$\mathop{\mathrm{E}}_{x \sim [s]^n}[f(x)\overline{f(x)}] = \mathop{\mathrm{E}}_{x \sim [s]^n}[\sum_{\alpha \in [s]^n} \hat{f}_\alpha \chi_\alpha(x) \cdot \overline{\sum_{\alpha' \in [s]^n} \hat{f}_{\alpha'} \chi_{\alpha'}(x)}] = \sum_{\alpha,\alpha' \in [s]^n} \hat{f}_\alpha \overline{\hat{f}_{\alpha'}} \mathop{\mathrm{E}}_{x \sim [s]^n}[\chi_{\alpha-\alpha'}(x)] = \sum_{\alpha \in [s]^n} |\hat{f}_\alpha|^2.$$

where the last equality holds because we have $\mathrm{E}_{x \sim [s]^n}[\chi_{\alpha-\alpha'}(x)]$ equals 0 if $\alpha \neq \alpha'$ and equals 1 otherwise. □

Fact 26. *Let $E = (E_1, \ldots, E_k)$ be the distribution over $[s]^k$, where the symbols are independent and each of them is set to uniform with probability η and is* 0 *otherwise. Then for every $\alpha \in [s]^n$, $\mathrm{E}[\chi_\alpha(E)] = (1-\eta)^{|\alpha|}$.*

Proof. The expectation conditioned on the event "E sets none of the nonzero positions of α to uniform" is 1. This event happens with probability $(1-\eta)^{|\alpha|}$. Conditioned on its complement, the expectation is 0. To see this, assume that the noise vector sets to uniform position i of α, and that $\alpha_i \neq 0$. Let $\beta := \omega^{\alpha_i}$. Then the expectation can be written as a product where a factor is

$$\mathop{\mathrm{E}}_{x\sim\{0,1,\ldots,s-1\}}[\beta^x] = \frac{1}{s}\cdot\frac{\beta^s-1}{\beta-1} = 0,$$

using the fact that $\beta \neq 1$ because $\alpha_i \in \{1,2,\ldots,s-1\}$ and that $\beta^s = (\omega^s)^{\alpha_i} = 1$. Therefore the total expectation is $(1-\eta)^{|\alpha|}$. □

Note that this lemma includes the uniform $\eta = 1$ case, with the convention $0^0 = 1$.

We will use the following facts multiple times.

Fact 27. *Let $f\colon [s]^n \to \mathbb{C}$ be a function with degree b. We have:*

(1) For any (b,δ)-biased distribution D over $[s]^n$, $|\mathrm{E}[f(D)] - \mathrm{E}[f(U)]| \leq L_1[f]\delta$,

(2) For any $(2b,\delta)$-biased distribution D over $[s]^n$, $|\mathrm{E}[|f(D)|^2] - \mathrm{E}[|f(U)|^2]| \leq L_1[f]^2\delta$, and

(3) the bound in (2) holds even if D is (n,δ) biased.

Proof. For (1), note that $|\mathrm{E}[f(D)]-\mathrm{E}[f(U)]| = \left|\sum_{0<|\alpha|\leq b}\hat{f}_\alpha \mathrm{E}[\chi_\alpha(D)]\right| \leq \sum_{0<|\alpha|\leq b}|\hat{f}_\alpha||\mathrm{E}[\chi_\alpha(D)]| \leq L_1[f]\delta$.

For (2), recall that $|f(x)|^2 = f(x)\overline{f(x)}$. By Fact 24, the function $|f(x)|^2$ has degree $\leq 2b$. Also, again by Fact 24 the L_1-norm of that function is at most $L_1[f]\cdot L_1[\bar{f}] = L_1[f]^2$. Now the result follows by (1).

Finally, (3) is proved like (2), noting that a function on $[s]^n$ always has degree $\leq n$. □

Actually the bounds hold with $\sum_{\alpha\neq 0}|\hat{f}_\alpha|$ instead of $L_1[f]$, but we will not use that.

2.2 Proof of Theorem 23

For a function $f\colon [s]^n \to \mathbb{C}_1$, consider its Fourier expansion $f(x) := \sum_\alpha \hat{f}_\alpha\chi_\alpha(x)$, and let $f^L(x) := \sum_{\alpha:|\alpha|\leq t}\hat{f}_\alpha\chi_\alpha(x)$ and $f^H(x) := \sum_{\alpha:|\alpha|>t}\hat{f}_\alpha\chi_\alpha(x)$. Define $F_i\colon ([s]^n)^k \to \mathbb{C}$ to be

$$F_i(x_1,\ldots,x_k) := \left(\prod_{j<i} f_j(x_j)\right)\cdot f_i^H(x_i)\cdot\left(\prod_{\ell>i} f_\ell^L(x_\ell)\right).$$

Pick f_k and write it as $f_k^L + f_k^H$. We can then rewrite

$$\prod_{1\leq i\leq k} f_i = F_k + \left(\prod_{1\leq i\leq k-1} f_i\right)\cdot f_k^L.$$

We can reapply the process to $(\prod_{1\le i\le k-1} f_i)$. Continuing this way, we eventually have what we want to bound, i.e. $|\mathrm{E}[\prod_{i\le k} f_i(D_i+E_i)] - \prod_{i\le k}\mu_i|$, is at most

$$\left|\sum_{i\le k} \mathrm{E}[F_i(D+E)]\right| + \left|\mathrm{E}[\prod_{i\le k} f_i^L(D_i+E_i)] - \prod_{i\le k}\mu_i\right|.$$

The theorem follows readily from the next two lemmas, the second of which has a longer proof.

Lemma 28. $|\mathrm{E}[\prod_{i\le k} f_i^L(D_i+E_i)] - \prod_{i\le k}\mu_i| \le V(t)^{k/2}\delta$.

Proof. Fix E arbitrarily. Each f_i^L has degree at most t, and by the Cauchy–Schwarz inequality, it has L_1-norm $\sum_{|\alpha|\le t}|\hat{f}_\alpha| \le V(t)^{1/2}(\sum_\alpha |\hat{f}_\alpha|^2)^{1/2} \le V(t)^{1/2}$. Here we use the fact that f maps to $\mathbb{C}_1$ and Fact 25. Hence, by Fact 24, $\prod_{0<i\le k} f_i^L$ has degree at most kt and L_1-norm at most $V(t)^{k/2}$. By hypothesis, D is (b,δ)-biased, and this also holds for $D+E$ for any fixed E. Moreover, $b \ge 2(k-1)t \ge kt$, and so by (1) in Fact 27 we have

$$\left|\mathop{\mathrm{E}}_D[\prod_{i\le k} f_i^L(D_i+E_i)] - \prod_{i\le k}\mu_i\right| \le V(t)^{k/2}\delta.$$

Averaging over E proves the claim. □

Lemma 29. *For every $i \in \{1,2,\dots,k\}$, we have* $|\mathrm{E}[F_i(D+E)]| \le (1-\eta)^t\sqrt{(1+s^n\delta)(1+V(t)^{k-1}\delta)}$.

Proof. We have

$$\begin{aligned}
|\mathrm{E}[F_i(D+E)]| &= \left|\mathrm{E}\left[\prod_{j<i} f_j(D_j+E_j)\cdot f_i^H(D_i+E_i)\cdot\prod_{\ell>i} f_\ell^L(D_\ell+E_\ell)\right]\right| \\
&\le \mathop{\mathrm{E}}_D\left[\prod_{j<i}\left|\mathop{\mathrm{E}}_{E_j}[f_j(D_j+E_j)]\right|\cdot\left|\mathop{\mathrm{E}}_{E_i}[f_i^H(D_i+E_i)]\right|\cdot\prod_{\ell>i}\left|\mathop{\mathrm{E}}_{E_\ell}[f_\ell^L(D_\ell+E_\ell)]\right|\right] \\
&\le \mathop{\mathrm{E}}_D\left[\left|\mathop{\mathrm{E}}_{E_i}[f_i^H(D_i+E_i)]\right|\cdot\prod_{\ell>i}\left|\mathop{\mathrm{E}}_{E_\ell}[f_\ell^L(D_\ell+E_\ell)]\right|\right],
\end{aligned}$$

where the last inequality holds because $|\mathrm{E}_{E_j}[f_j(D_j+E_j)]| \le \mathrm{E}_{E_j}[|f_j(D_j+E_j)|] \le 1$ for every $j<i$, by Jensen's inequality, convexity of norms, and the fact that the range of f_j is $\mathbb{C}_1$.

By the Cauchy–Schwarz inequality, we get

$$|\mathrm{E}[F_i(D+E)]| \le \mathop{\mathrm{E}}_D\left[\left|\mathop{\mathrm{E}}_{E_i}[f_i^H(D_i+E_i)]\right|^2\right]^{1/2}\cdot\mathop{\mathrm{E}}_D\left[\prod_{\ell>i}\left|\mathop{\mathrm{E}}_{E_\ell}[f_\ell^L(D_\ell+E_\ell)]\right|^2\right]^{1/2}.$$

In claims 31 and 32 below we bound above the square of the two terms on the right-hand side. In both cases, we view our task as bounding $|\mathrm{E}_D[g(D)]|$ for a certain function g, and we proceed by computing the L_1-norm, average over uniform, and degree of g, and then we apply Fact 27.

We start with a claim that is useful in both cases.

Claim 30. *Let $f\colon [s]^n \to \mathbb{C}$ be a function. Then:*
(1) for every x, $\mathrm{E}_E[f(x+E)] = \sum_\alpha \hat{f}_\alpha \chi_\alpha(x)(1-\eta)^{|\alpha|}$, and
(2) $\mathrm{E}_U\left[\left|\mathrm{E}_E[f(U+E)]\right|^2\right] = \sum_\alpha |\hat{f}_\alpha|^2(1-\eta)^{2|\alpha|}$.

Proof. For (1), write $\mathrm{E}_E[f(x+E)] = \mathrm{E}_E[\sum_\alpha \hat{f}_\alpha \chi_\alpha(x+E)] = \sum_\alpha \hat{f}_\alpha \chi_\alpha(x)\,\mathrm{E}_E[\chi_\alpha(E)]$. Then apply Fact 26.

For (2), write $\left|\mathrm{E}_E[f(x+E)]\right|^2$ as $\mathrm{E}_E[f(x+E)]\overline{\mathrm{E}_E[f(x+E)]}$. Then apply (1) twice to further write it as

$$\mathop{\mathrm{E}}_U\left[\sum_{\alpha,\alpha'} \hat{f}_\alpha \overline{\hat{f}_{\alpha'}} \chi_{\alpha-\alpha'}(U)(1-\eta)^{|\alpha|+|\alpha'|}\right] = \sum_{\alpha,\alpha'} \hat{f}_\alpha \overline{\hat{f}_{\alpha'}} \mathop{\mathrm{E}}_U[\chi_{\alpha-\alpha'}(U)](1-\eta)^{|\alpha|+|\alpha'|}.$$

The claim then follows because U is uniform. □

We can now bound our terms.

Claim 31. *For every i, $\mathrm{E}_D\left[\left|\mathrm{E}_{E_i}[f_i^H(D_i+E_i)]\right|^2\right] \le (1-\eta)^{2t}(1+s^n\delta)$.*

Proof. Let $g(x)$ be the function $g(x) = \mathrm{E}_{E_i}[f_i^H(x+E_i)]$. By (1) in Claim 30, the L_1-norm of g is at most $\sum_{\alpha:|\alpha|>t} |\hat{f}_\alpha|(1-\eta)^{|\alpha|} \le (1-\eta)^t \sum_\alpha |\hat{f}_\alpha| \le (1-\eta)^t s^{n/2}$, where we used Cauchy–Schwarz and Fact 25.

By (2) in Claim 30 and Fact 25, $\mathrm{E}_U[|g(U)|^2]$ under uniform is at most $(1-\eta)^{2t}$.

Because $b \ge n$ we can apply (3) in Fact 27 to obtain that $\mathrm{E}_D[|g(D)|^2] \le (1-\eta)^{2t} + (1-\eta)^{2t}s^n\delta$ as claimed. □

Claim 32. $\mathrm{E}_D\left[\prod_{\ell>i}\left|\mathrm{E}_{E_\ell}[f_\ell^L(D_\ell+E_\ell)]\right|^2\right] \le 1 + V(t)^{k-1}\delta$.

Proof. Pick any $\ell > i$ and let $g_\ell(x) := \mathrm{E}_E[f_\ell^L(x+E_\ell)]$.

The L_1-norm of g_ℓ is at most $V(t)^{1/2}$ by (1) in Claim 30 and Cauchy–Schwarz. Also by (2) in the same claim we have $\mathrm{E}_U[|g_\ell(U)|^2] \le 1$. Moreover, g_ℓ has degree at most t by (1) in the same claim.

Now define $g\colon ([s]^n)^{k-i} \to \mathbb{C}$ as $g(x_{i+1}, x_{i+2}, \dots, x_k) := g_{i+1}(x_{i+1}) \cdot g_{i+2}(x_{i+2}) \cdots g_k(x_k)$. Note that g has L_1-norm at most $V(t)^{(k-i)/2} \le V(t)^{(k-1)/2}$ and degree $(k-i)t \le (k-1)t$, by Fact 24 applied with $u = n(k-i)$. Moreover, $\mathrm{E}_{U_{i+1},U_{i+2},\dots,U_k}[|g(U_{i+1},U_{i+2},\dots,U_k)|^2] = \mathrm{E}_{U_{i+1}}[|g_{i+1}|^2] \cdot \mathrm{E}_{U_{i+2}}[|g_{i+2}|^2] \cdots \mathrm{E}_{U_k}[|g_k|^2] \le 1$.

Because $b \ge 2(k-1)t$, we can apply (2) in Fact 27 to obtain

$$\mathop{\mathrm{E}}_D[|g(D)|^2] \le 1 + V(t)^{k-1}\delta$$

as desired. □

Lemma 29 follows by combining claims 31 and 32. □

3 Proofs for §1.1

In this section we provide the proofs for the claims made in §1.1.

Proof of Theorem 12. Let L be the set of the 2^c leaves of the protocol tree. For $\ell \in L$, note that the set of inputs that lead to ℓ forms a rectangle, denoted R_ℓ. Moreover, these rectangles are disjoint.

Hence, applying Theorem 5 to each R_ℓ we can write

$$\begin{aligned}|\Pr[P(C+E)=1] - \Pr[P(U)=1]| &= |\sum_\ell \Pr[C+E \in R_\ell] - \sum_\ell \Pr[U \in R_\ell]| \\ &\leq \sum_\ell |\Pr[C+E \in R_\ell] - \Pr[U \in R_\ell]|,\end{aligned}$$

from which the result follows. □

Recall that we denote by $V(t)$ the number of $x \in \mathbb{F}_q^m$ with at most t non-zero coordinates.

Claim 33. *The following two inequalities hold:* $V(t) \leq \binom{m}{t} q^t \leq (emq/t)^t$.

Proof. The second is standard. To see the first, note that to specify a string with Hamming weight $\leq t$ we can specify a super-set of size t of the non-zero positions, and then values for those positions, including 0. □

Proof of Theorem 13. Let $H \in \mathbb{F}_q^{(m-d)\times m}$ be the parity-check matrix of C. Let H' be the matrix consisting of the first s rows of H. Note that we do have at least this many rows by our hypothesis on d. Also note that H' has full rank.

On input $x \in \mathbb{F}_q^m$, the algorithm computes $H'x$, and accepts if and only if $H'x$ equals to $H'e$ for any $e \in \mathbb{F}_q^m$ of Hamming weight at most t.

To analyze the correctness, let y be a codeword with at most t errors. Then $H(y-e) = 0$ for some $e \in \mathbb{F}_q^m$ with Hamming weight at most t, and so the algorithm always accepts. On the other hand if U is uniform, then as H' has full rank, $H'U$ is uniform in $\mathbb{F}_q^s$. Since there are $V(t)$ vectors in $\mathbb{F}_q^m$ with Hamming weight at most t, the algorithm accepts with probability $\leq V(t)/q^s \leq \delta$.

Now we show how to compute $H'x$ using s symbols of space (and so $s \log q$ bits). For $i \leq s$, let h_i be the i-th row of H'. Note that the i-th symbol of $H'x$ equals $\sum_{j \leq n} h_{i,j} x_j$, which can be computed with one symbol of space by keeping the partial sum. The result follows.

The "moreover" part follows from Claim 33. □

Proof of Theorem 14. Suppose

$$\Pr\left[P(GU+E) = U_i\right] \geq 1/q + \varepsilon.$$

Let D_a be the uniform distribution over $\{Gx \mid x_i = a\}$. We can rewrite the inequality as

$$\mathop{\mathrm{E}}_{a \in \mathbb{F}_q} [\Pr[P(D_a+E) = a] - \Pr[P(U) = a]] \geq \varepsilon.$$

Therefore, there exists an a such that $\Pr[P(D_a + E) = a] - \Pr[P(U) = a] \geq \varepsilon$.

We now use P to construct a protocol P' that distinguishes $D_0 + E$ from uniform. Given $y \in \mathbb{F}_q^m$, the parties add to y the ith column G_i of G multiplied by a. This can be done without communication. Then they run the protocol P on $y + aG_i$ and accept if and only if the output is a. We have

$$\begin{aligned}\Pr[P'(D_0 + E) = 1] - \Pr[P'(U) = 1] &= \Pr[P(D_0 + aG_i + E) = a] - \Pr[P(U) = a] \\ &= \Pr[P(D_a + E) = a] - \Pr[P(U) = a] \\ &\geq \varepsilon.\end{aligned}$$

So the result follows from Theorem 12.

□

Proof of Theorem 15. Let $n := q/k$ be the input length to a party. The parties communicate $m-d+2t+1-n$ symbols that the first does not have, and no symbol if $m-d+2t+1-n \leq 0$. The first party then outputs the unique message whose encoding is at distance $\leq t$ with the $m - d + 2t + 1$ symbols z they have, i.e., the symbols they received plus the n they already have. The message corresponding to x clearly is such a message. Also no other such message exists, because if two encodings are at distance $\leq t$ with z then they agree with each other in $\geq m - d + 1$ symbols, and so they cannot differ in d positions and must be the same. □

4 Pseudorandomness: I

In this section we prove our first theorem on pseudorandom generators, Theorem 18.

First, we shall need the following lemma to sample our noise vectors, which is also used in the next section. We write SD for *statistical distance*.

Lemma 34. *There is a polynomial-time computable function f mapping $O(\eta \log(1/\eta)m)$ bits to $\{0,1\}^m$ such that $\mathrm{SD}(f(U), E) \leq e^{-\Omega(\eta m)}$.*

In turn, that will use the following lemma to sample arbitrary distributions through discretization. A version of the lemma appears in [Vio12], Lemma 5.2. That version only bounds the number of bits of the sampler. Here we also need that the sampler is efficient.

Lemma 35. *Let D be a distribution on $S := \{1, 2, \ldots, n\}$. Suppose that given $i \in S$ we can compute in time polynomial in $|i| = O(\log n)$ the cumulative distribution $\Pr[D \leq i]$.*

Then there is a polynomial-time computable function f such that given any $t \geq 1$ uses $\lceil \log_2 nt \rceil$ bits to sample a string in the support of D such that $\mathrm{SD}(f(U), D) \leq 1/t$.

Proof. Following [Vio12, Lemma 5.2], partition the interval $[0, 1]$ into n intervals I_i of lengths $\Pr[D = i]$, $i = 1, \ldots, n$. Also partition $[0, 1]$ in $\ell := 2^{\lceil \log_2 nt \rceil} \geq nt$ intervals of size $1/\ell$ each, which we call blocks. The function f interprets an input as a choice of a block b, and outputs i if $b \subseteq I_i$ and, say, outputs 1 if b is not contained in any interval.

For any i we have $|\Pr[D = i] - \Pr[f(U) = i]| \le 2/\ell$. Hence the statistical distance is $\le (1/2)\sum_i |\Pr[D = i] - \Pr[f(U) = i]| \le (1/2)n2/\ell \le 1/t$.

To show efficiency we have to explain how given b we determine the i such that $b \subseteq I_i$. We perform binary search. This requires $O(\log n)$ steps, and in each step we compute the cumulative distribution function of D, which by assumption is in polynomial time. □

Proof of Lemma 34. Our function f first samples a weight distribution W on $\{0, \dots, m\}$ so that $\mathrm{SD}(W, |E|) \le e^{-\Omega(\eta m)}$. By Lemma 35, this uses a seed of length $O(\eta m + \log(m+1))$ and runs in polynomial time. Given a sample $w \sim W$. If $w \ge 2\eta m$, we output the all-zero string. Otherwise we sample a string in $\{0,1\}^m$ with Hamming weight w almost uniformly. To do this, first we index the $\binom{m}{w}$ strings in lexicographical order. We then use Lemma 35 again to sample an index in $\{1, \dots, \binom{m}{w}\}$ from a distribution that is $e^{-\Omega(\eta m)}$-close to uniform. This takes another seed of length at most $O(\eta m + \log \binom{m}{2\eta m}) = O(\eta m + \eta \log(1/\eta) m)$ and can be computed in polynomial time.

Given an index i, we output the corresponding string efficiently using the following recurrence. Let $s(m, k, i)$ denote the i-th m-bit string with Hamming weight k, in lexicographical order. We have

$$s(m,k,i) = \begin{cases} 0 \circ s(m-1,k,i) & \text{if } i \le \binom{m-1}{k} \\ 1 \circ s(m-1,k-1,i-\binom{m-1}{k}) & \text{otherwise.} \end{cases}$$

Note that $s(m, k, i)$ outputs the string by m comparisons of $\lceil \binom{m}{2\eta m} \rceil$-bit strings, and thus can be computed in polynomial time.

Therefore f has input length $O(\eta m + \eta \log(1/\eta) m + \log(m+1)) = O(\eta \log(1/\eta) m)$. Let $D := f(U)$. We now bound above the statistical distance between D and E. Denote D_w as the distribution D conditioned on $|D| = w$ and denote E_w analoguously. We have

$$\begin{aligned} \sum_{x \in \{0,1\}^m} |\Pr[D(x)] - \Pr[E(x)]| &= \sum_{w=0}^{m} \sum_{|x|=w} |\Pr[D(x)] - \Pr[E(x)]| \\ &= \sum_{w=0}^{m} \sum_{|x|=w} |\Pr[D_w(x)] \Pr[|D| = w] - \Pr[E_w(x)] \Pr[|E| = w]|, \end{aligned}$$

Adding $-\Pr[D_w(x)] \Pr[|E| = w] + \Pr[D_w(x)] \Pr[|E| = w] = 0$ in each summand, this is at most

$$\sum_{w=0}^{m} \sum_{|x|=w} \Pr[D_w(x)] \cdot |\Pr[|D| = w] - \Pr[|E| = w]| + \sum_{w=0}^{m} \sum_{|x|=w} |\Pr[D_w(x)] - \Pr[E_w(x)]| \cdot \Pr[|E| = w].$$

The first double summation is at most $2\,\mathrm{SD}(|D|, |E|) = 2\,\mathrm{SD}(W, |E|)$. We now bound above the second summation as follows. We separate the outer sum into $w > 2\eta m$ and $w \le 2\eta m$. For the first case, we have

$$\sum_{w > 2\eta m} \sum_{|x|=w} |\Pr[D_w(x)] - \Pr[E_w(x)]| \cdot \Pr[|E| = w] \le 2 \Pr[|E| > 2\eta m].$$

By the Chernoff Bound, this is at most $2e^{-\Omega(\eta m)}$. For the other case, we have

$$\sum_{w \le 2\eta m} \sum_{|x|=w} |\Pr[D_w(x)] - \Pr[E_w(x)]| \cdot \Pr[|E| = w] \le 2 \max_{w \le 2\eta m} \mathrm{SD}(D_w, E_w).$$

Therefore,

$$\mathrm{SD}(D, E) \le \mathrm{SD}(W, |E|) + \max_{w \le 2\eta m} \mathrm{SD}(D_w, E_w) + e^{-\Omega(\eta m)} \le 3e^{-\Omega(\eta m)}. \qquad \square$$

We can now prove our first theorem on pseudorandom generators.

Proof of Theorem 18. (1) We apply Theorem 23. Known constructions [AGHP92, Theorem 2] (see also [NN93]) produce a δ-biased distribution over m bits using $2\log(1/\delta) + O(\log m)$ bits. We set $\delta = O(2^{-n}\varepsilon)$, resulting in a seed length of $2n + 2\log(1/\varepsilon) + O(\log m)$ bits.

For the noise we set $\eta = O(k \log k \log(k/\varepsilon)/n)$. Note that $\eta \le 1$ because we can assume $k^2 \log k \log(k/\varepsilon) \log n \le n$, for else (2) gives a better bound.

By Lemma 34, the seed length to generate the noise vector is $O(k^2 \log k \log(k/\varepsilon) \log(n/k))$.

In Theorem 23 set $t = cn/(k \log k)$ for a small enough constant c. Then we can bound $V(t)^{k/2} \le V(t)^{k-1} \le 2^n$. Thus the error bound from Theorem 23 is at most $k(1-\eta)^{cn/(k \log k)}(1 + 2^n\delta) + 2^n\delta \le 2k(1-\eta)^{cn/(k \log k)} + \varepsilon/4 \le \varepsilon/2$.

The error from Lemma 34 is $e^{-\Omega(\eta m)} \le \varepsilon/2$. Thus overall the error is at most ε.

The fact that we can apply any permutation π follows from the fact that applying such a permutation does not change the noise distribution, and preserves the property of being b-wise independent.

(2) Let $c := \lfloor (k^2 \log k \log(k/\varepsilon) \log n/n)^{1/3} \rfloor$. We can assume $c \ge 1$ for else (1) gives a better bound. Group the k functions into $k' = \lceil k/c \rceil$ functions on input length $n' := cn$. We can now apply (1) to n' and k' to get the desired seed length.

$\square$

5 Pseudorandomness, II

We now move to our second theorem on pseudorandom generators, Theorem 19. We begin by modifying Theorem 23 to allow us to sample the noise in a certain pseudorandom way. Specifically, we can write our noise vector E in the previous sections as $E = T \wedge U$, where U is uniform, T is a distribution of i.i.d. bits where each comes 1 with probability η, and $\wedge$ denotes bit-wise And. In the derandomized way, we keep U uniform but select T using an almost n-wise independent distribution. The analogue of Theorem 23 with this derandomization is proved below as Theorem 37. Finally, we show how to recurse on U in §5.2.

At the end of the section we show that a certain improvement in the error bound of Theorem 37 would yield much better pseudorandom generators.

Definition 36. *A distribution T on m bits is* γ-almost d-wise independent *if for every d indices $i_1, \ldots, i_d$ and any $S \subseteq \{0,1\}^d$ we have*

$$\left| \sum_{x \in S} \left(\Pr[\bigwedge_{j \le d} T_{i_j} = x_j] - \prod_{j \le d} \Pr[T_{i_j} = x_j] \right) \right| \le \gamma.$$

Theorem 37 (Bounded independence plus derandomized noise fools products)**.** *Let $t \in [0,n]$. Let $f_1, \ldots, f_k \colon \{0,1\}^n \to \mathbb{C}_1$ be k functions with $\mu_i = \mathrm{E}[f_i]$. Let D be an δ-biased distribution over $(\{0,1\}^n)^k$. Let T be a γ-almost n-wise distribution over $(\{0,1\}^n)^k$ which sets each bit to 1 with probability η and 0 otherwise. Assume $\gamma \le \eta$. Let U be the uniform distribution over $(\{0,1\}^n)^k$. Write $D = (D_1, D_2, \ldots, D_k)$ where each D_i is in $\{0,1\}^n$, and similarly for T and U. Then*

$$\left| \mathrm{E}\left[\prod_{i \le k} f_i(D_i + T_i \wedge U_i) \right] - \prod_{i \le k} \mu_i \right| \le k((1-\eta)^t + \gamma)^{1/2} \sqrt{(1+2^n\delta)(1+V(t)^{k-1}\delta)} + V(t)^{k/2}\delta.$$

5.1 Proof of Theorem 37

We begin exactly as in the proof of Theorem 23. For a function $f \colon \{0,1\}^n \to \mathbb{C}_1$, consider its Fourier expansion $f(x) := \sum_\alpha \hat{f}_\alpha \chi_\alpha(x)$, and let $f^L(x) := \sum_{\alpha : |\alpha| \le t} \hat{f}_\alpha \chi_\alpha(x)$ and $f^H(x) := \sum_{\alpha : |\alpha| > t} \hat{f}_\alpha \chi_\alpha(x)$. Define $F_i \colon (\{0,1\}^n)^k \to \mathbb{C}$ to be

$$F_i(x_1, \ldots, x_k) := \left(\prod_{j<i} f_j(x_j) \right) \cdot f_i^H(x_i) \cdot \left(\prod_{\ell > i} f_\ell^L(x_\ell) \right).$$

Pick f_k and write it as $f_k^L + f_k^H$. We can then rewrite

$$\prod_{1 \le i \le k} f_i = F_k + \left(\prod_{1 \le i \le k-1} f_i \right) \cdot f_k^L.$$

We can reapply the process to $(\prod_{1 \le i \le k-1} f_i)$. Continuing this way, we eventually have what we want to bound, i.e. $|\mathrm{E}[\prod_{i \le k} f_i(D_i + T_i \wedge U_i)] - \prod_{i \le k} \mu_i|$, is at most

$$\left| \sum_{i \le k} \mathrm{E}[F_i(D + T \wedge U)] \right| + \left| \mathrm{E}[\prod_{i \le k} f_i^L(D_i + T_i \wedge U_i)] - \prod_{i \le k} \mu_i \right|.$$

The theorem follows readily from the next two lemmas, the second of which has a longer proof. The first one has the same proof as Lemma 28.

Lemma 38. $|\mathrm{E}_{D,T,U}[\prod_{i \le k} f_i^L(D_i + T_i \wedge U_i)] - \prod_{i \le k} \mu_i| \le V(t)^{k/2}\delta$.

Lemma 39. *For every $i \in \{1, 2, \ldots, k\}$, we have*

$$|\mathop{\mathrm{E}}_{D,T,U}[F_i(D + T \wedge U)]| \le ((1-\eta)^t + \gamma)^{1/2} \sqrt{(1+2^n\delta)(1+V(t)^{k-1}\delta)}.$$

Proof. We have

$$\begin{aligned}
|\mathrm{E}[F_i(D+T\wedge U)]| &= \left|\mathrm{E}\left[\prod_{j<i} f_j(D_j+T_j\wedge U_j)\cdot f_i^H(D_i+T_i\wedge U_i)\cdot\prod_{\ell>i} f_\ell^L(D_\ell+T_\ell\wedge U_\ell)\right]\right| \\
&\le \mathop{\mathrm{E}}_{D,T}\left[\prod_{j<i}\left|\mathop{\mathrm{E}}_{U_j}[f_j(D_j+T_j\wedge U_j)]\right|\cdot\left|\mathop{\mathrm{E}}_{U_i}[f_i^H(D_i+T_i\wedge U_i)]\right|\cdot\prod_{\ell>i}\left|\mathop{\mathrm{E}}_{U_\ell}[f_\ell^L(D_\ell+T_\ell\wedge U_\ell)]\right|\right] \\
&\le \mathop{\mathrm{E}}_{D,T}\left[\left|\mathop{\mathrm{E}}_{U_i}[f_i^H(D_i+T_i\wedge U_i)]\right|\cdot\prod_{\ell>i}\left|\mathop{\mathrm{E}}_{U_\ell}[f_\ell^L(D_\ell+T_\ell\wedge U_\ell)]\right|\right],
\end{aligned}$$

where the last inequality holds because $|\mathrm{E}_{U_j}[f_j(D_j+T_j\wedge U_j)]| \le \mathrm{E}_{U_j}[|f_j(D_j+T_j\wedge U_j)|] \le 1$ for every $j<i$, by Jensen's inequality, convexity of norms, and the fact that the range of f_j is $\mathbb{C}_1$.

By the Cauchy–Schwarz inequality, we get

$$|\mathrm{E}[F_i(D+T\wedge U)]| \le \mathop{\mathrm{E}}_{D,T}\left[\left|\mathop{\mathrm{E}}_{U_i}[f_i^H(D_i+T_i\wedge U_i)]\right|^2\right]^{1/2}\cdot\mathop{\mathrm{E}}_{D,T}\left[\prod_{\ell>i}\left|\mathop{\mathrm{E}}_{U_\ell}[f_\ell^L(D_\ell+T_\ell\wedge U_\ell)]\right|^2\right]^{1/2}.$$

In claims 41 and 42 below we bound from above the square of the two terms on the right-hand side. In both cases, we view our task as bounding $|\mathrm{E}_D[g(D)]|$ for a certain function g, and we proceed by computing the L_1-norm, average over uniform, and degree of g, and then we apply Fact 27.

We start with a claim that is useful in both cases.

Claim 40 (Replacing Claim 30)**.** *Let $f\colon\{0,1\}^n\to\mathbb{C}$ be a function. Let T be a γ-almost n-wise independent distribution which sets each bit to 1 with probability η and 0 otherwise. Let U and U' be two independent uniform distributions over n bits. Then:*

(1) for every x, $\mathrm{E}_{T,U}[f(x+T\wedge U)] = \sum_\alpha \hat{f}_\alpha\chi_\alpha(x)((1-\eta)^{|\alpha|}+\gamma)$, and

(2) $\mathrm{E}_{U,T}\left[|\mathrm{E}_{U'}[f(U+T\wedge U')]|^2\right] = \sum_\alpha|\hat{f}_\alpha|^2((1-\eta)^{|\alpha|}+\gamma)$.

Proof. For (1), write $\mathrm{E}_{T,U}[f(x+T\wedge U)] = \mathrm{E}_{T,U}[\sum_\alpha \hat{f}_\alpha\chi_\alpha(x+T\wedge U)] = \sum_\alpha \hat{f}_\alpha\chi_\alpha(x)\,\mathrm{E}_{T,U}[\chi_\alpha(T\wedge U)]$. If T does not intersect α then the expectation is one, and this happens with probability at most $(1-\eta)^{|\alpha|}+\gamma$. Otherwise, the expectation is 0.

For (2), write $\mathrm{E}_{U,T}[|\mathrm{E}_{U'}[f(x+T\wedge U')]|^2]$ as

$$\mathop{\mathrm{E}}_{U,T}\left[\sum_{\alpha,\alpha'}\hat{f}_\alpha\overline{\hat{f}_{\alpha'}}\chi_{\alpha-\alpha'}(U)\mathop{\mathrm{E}}_{U',U''}[\chi_\alpha(T\wedge U')\overline{\chi_{\alpha'}(T\wedge U'')}]\right].$$

Since U is uniform this becomes $\sum_\alpha|\hat{f}_\alpha|^2\,\mathrm{E}_{T,U',U''}[\chi_\alpha(T\wedge(U'-U''))] = \sum_\alpha|f_\alpha|^2\,\mathrm{E}[\chi_\alpha(T\wedge U)]$. The claim then follows as in (1). □

We can now bound our terms.

Claim 41 (Replacing Claim 31)**.** *For every i,* $\mathrm{E}_{D,T}[|\mathrm{E}_U[f_i^H(D_i + T_i \wedge U_i)]|^2] \le ((1-\eta)^t + \gamma)(1 + 2^n\delta)$.

Proof. Let $g(x)$ be the function $g(x) = \mathrm{E}_{T_i,U_i}[f_i^H(x + T_i \wedge U_i)]$. By (1) in Claim 40, the L_1-norm of g is at most $\sum_{\alpha:|\alpha|>t} |\hat{f}_\alpha|((1-\eta)^{|\alpha|} + \gamma) \le ((1-\eta)^t + \gamma) \sum_\alpha |\hat{f}_\alpha| \le ((1-\eta)^t + \gamma)2^{n/2}$, where we used Cauchy–Schwarz and Fact 25.

Also, by (2) in Claim 40 and Fact 25, $\mathrm{E}_U[|g(U)|^2]$ under uniform is at most $(1-\eta)^t + \gamma$.

Because $nk \ge n$ we can apply (3) in Fact 27 to obtain that $\mathrm{E}_D[|g(D)|^2] \le ((1-\eta)^t + \gamma) + ((1-\eta)^t + \gamma)^2 2^n\delta \le ((1-\eta)^t + \gamma)(1 + 2^n\delta)$ as claimed. □

Claim 42. $\mathrm{E}_{D,T}\left[\prod_{\ell>i} \left|\mathrm{E}_{U_\ell}[f_\ell^L(D_\ell + T_\ell \wedge U_\ell)]\right|^2\right] \le 1 + V(t)^{k-1}\delta$.

Proof. Pick any $\ell > i$ and let $g_\ell(x) := \mathrm{E}_E[f_\ell^L(x + E_\ell)]$.

The L_1-norm of g_ℓ is at most $V(t)^{1/2}$ by (1) in Claim 30 and Cauchy–Schwarz. Also by (2) in the same claim we have $\mathrm{E}_U[|g_\ell(U)|^2] \le 1$. Moreover, g_ℓ has degree at most t by (1) in the same claim.

Now define $g\colon ([s]^n)^{k-i} \to \mathbb{C}$ as $g(x_{i+1}, x_{i+2}, \ldots, x_k) := g_{i+1}(x_{i+1}) \cdot g_{i+2}(x_{i+2}) \cdots g_k(x_k)$. Note that g has L_1-norm at most $V(t)^{(k-i)/2} \le V(t)^{(k-1)/2}$ and degree $(k-i)t \le (k-1)t$, by Fact 24 applied with $u = n(k-i)$. Moreover, $\mathrm{E}_{U_{i+1},U_{i+2},\ldots,U_k}[|g(U_{i+1}, U_{i+2}, \ldots, U_k)|^2] = \mathrm{E}_{U_{i+1}}[|g_{i+1}|^2] \cdot \mathrm{E}_{U_{i+2}}[|g_{i+2}|^2] \cdots \mathrm{E}_{U_k}[|g_k|^2] \le 1$.

Because $b \ge 2(k-1)t$, we can apply (2) in Fact 27 to obtain

$$\mathrm{E}_D[|g(D)|^2] \le 1 + V(t)^{k-1}\delta$$

as desired. □

Proof. Pick any $\ell > i$ and let $g_\ell(x) := \mathrm{E}_{T,U_\ell}[f_\ell^L(x + T_\ell \wedge U_\ell)]$.

The L_1-norm of g_ℓ is at most $V(t)^{1/2}$ by (1) in Claim 40, Cauchy–Schwarz, and the assumption that $\gamma \le \eta$. Also by (2) in the same claim we have $\mathrm{E}_U[|g_\ell(U)|^2] \le 1$. Moreover, g_ℓ has degree at most t by (1) in the same claim.

Now define $g\colon ([s]^n)^{k-i} \to \mathbb{C}$ as $g(x_{i+1}, x_{i+2}, \ldots, x_k) := g_{i+1}(x_{i+1}) \cdot g_{i+2}(x_{i+2}) \cdots g_k(x_k)$. Note that g has L_1-norm at most $V(t)^{(k-i)/2} \le V(t)^{(k-1)/2}$ and degree $(k-i)t \le (k-1)t$, by Fact 24 applied with $u = n(k-i)$. Moreover, $\mathrm{E}_{U_{i+1},U_{i+2},\ldots,U_k}[|g(U_{i+1}, U_{i+2}, \ldots, U_k)|^2] = \mathrm{E}_{U_{i+1}}[|g_{i+1}|^2] \cdot \mathrm{E}_{U_{i+2}}[|g_{i+2}|^2] \cdots \mathrm{E}_{U_k}[|g_k|^2] \le 1$.

Because $nk \ge 2(k-1)t$, we can apply (2) in Fact 27 to obtain

$$\mathrm{E}_D[|g(D)|^2] \le 1 + V(t)^{k-1}\delta$$

as desired. □

Lemma 39 follows by combining claims 41 and 42. □

5.2 A recursive generator

Lemma 43. *Suppose $n \geq Ck \log k \log(k/\varepsilon)$ for a universal constant C. Let c be an integer. If there is an explicit generator $G_{cn/4,k/c}\colon \{0,1\}^{\ell} \to (\{0,1\}^{cn/4})^{k/c}$ that ε-fools product tests that read bits in any order and uses a seed of length ℓ, then there is an explicit generator $G_{n,k}\colon \{0,1\}^{\ell'} \to (\{0,1\}^n)^k$ that fools product tests in any order with error $\varepsilon/k + \varepsilon$ and uses a seed of length $\ell' = O(n) + \ell$.*

Proof. Our generator $G_{n,k}\colon \{0,1\}^{\ell'} \to (\{0,1\}^n)^k$ samples a 2^{-2n}-biased distribution D on m bits, and a 2^{-2n}-almost n-wise independent distribution $T = (T_1, \ldots, T_k)$ on m bits which sets each bit to 1 with probability $1/8$ and 0 otherwise. If $|T_i| > n/4$ for some $1 \leq i \leq k$, let G output the all-zero m-bit string. Otherwise, output $D + T \wedge \mathrm{PAD}_T(G_{cn/4,k/c}(U_\ell))$, where $\mathrm{PAD}_T(x)_j$ is defined as follows: If $j \in T$, $\mathrm{PAD}_T(x)_j$ equals the first bit of x that has not appeared in the first $j-1$ bits of $\mathrm{PAD}_T(x)$. Otherwise $\mathrm{PAD}_T(x)_j = 0$. Note that $|T| \leq m/4$.

Now we analyze the seed length of $G_{n,k}$. Standard constructions [NN93, AGHP92] use a seed of $O(n)$ bits to sample D. To sample T, we will use the following lemma from [RSV13].

Lemma 44 (Lemma B.2 in [RSV13])**.** *There is an explicit sampler that samples a γ-almost n-wise independent distribution T on m bits which sets each bit to 1 with probability η and 0 otherwise and uses a seed of length $O(n \log(1/\eta) + \log((\log m)/\gamma))$.*

Applying the lemma with $\gamma = 2^{-2n}$ and $\eta = 1/8$, we can sample T with $O(n)$ bits. So the total seed length of $G_{n,k}$ is $\ell' = O(n) + \ell$.

We now analyze the error of $G_{n,k}$. Let $f\colon (\{0,1\}^n)^k \to \mathbb{C}_1$ be a product test. We bound above $|\mathrm{E}[f(U_m)] - \mathrm{E}[f(G_{n,k}(U_{\ell'}))]|$ by

$$|\mathrm{E}[f(U_m)] - \mathrm{E}[f(D + T \wedge U_m)]| + |\mathrm{E}[f(D + T \wedge U_m)] - \mathrm{E}[f(G_{n,k}(U_{\ell'})]|.$$

The first term is at most $\varepsilon/2k$ by Theorem 37 with the following choice of parameters. We set $t = cn/(k \log k)$ for a small enough constant c. Then we can bound $V(t)^{k/2} \leq V(t)^{k-1} \leq 2^n$. We set $\eta = 1/8$. Thus, by the condition $n \geq Ck \log k \log(k/\varepsilon)$ the error bound from Theorem 37 is at most $k((1-\eta)^{cn/(k \log k)} + \gamma)^{1/2}(1 + 2^n\delta) + 2^n\delta \leq O(k((\varepsilon/k)^{100} + 2^{-2n})^{1/2}) + 2^{-n} \leq \varepsilon/2k$.

For the second term, let T' be T conditioned on $|T_i| \leq n/4$ for every $1 \leq i \leq k$. For every fixed $y \in D$ and $t \in T'$, consider the function $f_{y,t}\colon (\{0,1\}^{n/4})^k \to \mathbb{C}_1$ by $f_{y,t}(x) := f(y + t \wedge \mathrm{PAD}_t(x))$. Note that we can group every c functions into one and think of $f_{y,t}$ as a product test of k/c functions on $cn/4$ bits, which can be fooled by $G_{cn/4,k/c}$. Thus, $|\mathrm{E}[f(D + T' \wedge U_m)] - \mathrm{E}[f(G_{n,k}(U_{\ell'}))]|$ equals

$$\begin{aligned}
&|\mathrm{E}[f(D + T' \wedge \mathrm{PAD}_{T'}(U_{m/4}))] - \mathrm{E}[f(D + T' \wedge \mathrm{PAD}_{T'}(G_{cn/4,k/c}(U_\ell)))]| \\
&\leq \mathop{\mathrm{E}}_{y \sim D, t \sim T'}[|\mathrm{E}[f_{y,t}(U_{m/4})] - \mathrm{E}[f_{y,t}(G_{cn/4,k/c}(U_\ell))]|] \\
&\leq \varepsilon.
\end{aligned}$$

Now let E denote the event $|T_i| > n/4$ for some $1 \leq i \leq k$. We bound above $\Pr[E]$. We will use the following tail bound for almost d-wise independence.

Lemma 45 (Lemma B.1 in [RSV13]). *Let $T = (T_1, \ldots, T_n)$ be a γ-almost d-wise independent distribution on n bits where $\mathrm{E}[T_j] = \eta$ for every $1 \le j \le n$. Then for any $\varepsilon \in (0,1)$,*

$$\Pr[|\sum_{i \le n} X_i - \eta n| \ge \varepsilon n] \le \left(\frac{ed}{2\varepsilon^2 n}\right)^{d/2} + \gamma/\varepsilon^d.$$

We apply Lemma 45 with $\eta = \varepsilon = 1/8$, $\gamma = 2^{-2n}$, and $d = \Omega(n)$. This guarantees that for each $1 \le i \le k$ the probability of $|T_i| \le n/4$ is at most $2^{-\Omega(n)}$. By a union bound over $T_1, \ldots, T_k$, we have $\Pr[E] \le k2^{-\Omega(n)} \le \varepsilon/2k$.

Putting everything together we have error $\varepsilon/k + \varepsilon$. □

Finally, we combine these results to prove Theorem 19.

Proof of Theorem 19. Let C be the universal constant in Lemma 43. Suppose $n \ge Ck \log k \log(k/\varepsilon)$. We will first apply Lemma 43 with $c = 2$ for $t := O(\log k)$ times until we are left with a product test of $O(1)$ functions on $O(n/k)$ bits, and then we output the uniform $O(n/k)$-bit string. Note that the condition $n \ge Ck \log k \log(k/\varepsilon)$ holds throughout because n and k are both divided by 2 at each step.

Note that in each application of Lemma 43, we reduce n by at least a half. Hence, the total seed length is at most $\sum_{i=0}^{t} O(n/2^i) + O(n/k) = O(n)$. The error is at most $\sum_{i=0}^{t} 2^i\varepsilon/k \le \varepsilon$.

If $n \le Ck \log k \log(k/\varepsilon)$, pick an integer $c = O(\sqrt{k \log k \log(k/\varepsilon)/n})$ so that $c^2 n \ge O(Ck \log k \log(k/\varepsilon))$. By grouping every c functions into one, f is also a product test of k/c functions on cn bits. Hence, by the previous result we have a generator with seed length $\ell = O(cn) = O(\sqrt{nk \log k \log(k/\varepsilon)})$. □

A potential improvement. We now show that an improvement in the error bound of Theorem 37 would yield much better pseudorandom generators.

Claim 46. *Let D be an n-wise independent distribution on $m := nk$ bits. Let T be an n-wise independent distribution on m bits which sets each bit to 1 with probability η. Let U be the uniform distribution on m bits.*

Suppose that for any product test $f\colon (\{0,1\}^n)^k \to \mathbb{C}_1$ on m bits we have $|\mathrm{E}[f(U)] - \mathrm{E}[f(D + T \wedge U)]| \le k(1-\eta)^{\Omega(n)}$.

Then there is an explicit generator $G\colon \{0,1\}^\ell \to (\{0,1\}^n)^k$ that ε-fools product tests in any order with seed length $\ell = O((n + \log k \log(m/\varepsilon)) \log m)$.

To prove Claim 46, first we replace Lemma 43 with the following lemma.

Lemma 47. *Suppose $n \ge C \log(m/\varepsilon)$ for a universal constant C. Let c be an integer. If there is an explicit generator $G_{cn/4,k/c}$ that fools product tests that read bits in any order on $(cn/4) \cdot (k/c)$ bits with error ε and uses a seed of length ℓ, then there is an explicit generator $G_{n,k}\colon \{0,1\}^{\ell'} \to (\{0,1\}^n)^k$ that fools product tests that read bits in any order on $m := nk$ bits with error $\varepsilon/m + \varepsilon$ and uses a seed of $\ell' = O(n \log m) + \ell$ bits.*

Proof. The generator is very similar to the one in Lemma 43 except that G now samples an n-wise independent distribution D on m bits and an n-wise independent distribution T on m bits that sets each bit to 1 with probability $1/8$ and 0 otherwise. Now sampling D and T takes a seed of length $O(n \log m)$ [CG89, ABI86].

Now we analyze the error of $G_{n,k}$. Let $f\colon (\{0,1\}^n)^k \to \mathbb{C}_1$ be a product test. As in the proof of Lemma 43 we bound above $|\mathrm{E}[f(U_m)] - \mathrm{E}[f(G_{n,k}(U_{\ell'}))]|$ by

$$|\mathrm{E}[f(U_m)] - \mathrm{E}[f(D + T \wedge U_m)]| + |\mathrm{E}[f(D + T \wedge U_m)] - \mathrm{E}[f(G_{n,k}(U_{\ell'}))]|.$$

By our assumption, the first term is at most $k(1-\eta)^{\Omega(n)} \le \varepsilon/2m$. For the second term, let T' be T conditioned on $|T_i| \le n/4$ for every $1 \le i \le k$. For every fixed $y \in D$ and $t \in T'$, consider the function $f_{y,t}\colon (\{0,1\}^{n/4})^k \to \mathbb{C}_1$ by $f_{y,t}(x) := f(y + t \wedge \mathrm{PAD}_t(x))$. Note that we can group every c functions into one and think of $f_{y,t}$ as a product test of k/c functions on $cn/4$ bits, which can be fooled by $G_{cn/4,k/c}$. Thus, $|\mathrm{E}[f(D + T' \wedge U_m)] - \mathrm{E}[f(G_{n,k}(U_{\ell'}))]|$ equals

$$\begin{aligned}
&|\mathrm{E}[f(D + T' \wedge \mathrm{PAD}_{T'}(U_{m/4}))] - \mathrm{E}[f(D + T' \wedge \mathrm{PAD}_{T'}(G_{cn/4,k/c}(U_\ell)))]| \\
&\le \mathop{\mathrm{E}}_{y \sim D, t \sim T'}[|\mathrm{E}[f_{y,t}(U_{m/4})] - \mathrm{E}[f_{y,t}(G_{cn/4,k/c}(U_\ell))]|] \\
&\le \varepsilon.
\end{aligned}$$

Now let E denote the event $|T_i| > n/4$ for some $1 \le i \le k$. We bound above $\Pr[E]$. Since T is n-wise independent, by the Chernoff bound the probability of $|T_i| \le n/4$ is at most $2^{-\Omega(n)}$. By a union bound over $T_1, \ldots, T_k$, we have $\Pr[E] \le k2^{-\Omega(n)} \le \varepsilon/2m$.

Putting everything together we have error $\varepsilon/m + \varepsilon$. □

Proof of Claim 46. Suppose $n \ge C \log(m/\varepsilon)$. We apply Lemma 47 recursively, in two different ways. One way reduces n and the other reduces k. First, we apply the lemma with $c = 1$ for $t_1 := O(\log n)$ times to bring n down to $n' = O(\log(m/\varepsilon))$. This takes a seed of $\ell_1 := \sum_{i=0}^{t} O(n \log m/4^i) = O(n \log m)$ bits. Now we have a product test of k functions on n' bits. We will instead think of it as a product test of $k/2$ functions on $2n'$ bits, and apply Lemma 47 with $c = 2$, which will reduce it to a product test of $k/4$ functions on n' bits. Now we repeat $t_2 := O(\log k)$ steps to reduce k to $k' = O(1)$. This takes a seed of $\ell_2 := t_2 \cdot O(n' \log m) = O(\log k \log(m/\varepsilon) \log m)$ bits. Now we are left with a product test of k' functions on n' bits, and we can output the uniform string. Therefore the total seed length is $\ell = \ell_1 + \ell_2 + O(\log(m/\varepsilon)) = O((n + \log k \log(m/\varepsilon)) \log m)$. Because in each application of Lemma 47 the input length of the product test decreases by at least half, the error bound is at most $\sum_{i=0}^{t_1+t_2} 2^i \varepsilon/m \le 2^{O(\log m)}\varepsilon/m \le \varepsilon$.

If $n \le C \log(m/\varepsilon)$, we can group the functions and have a product test of k' functions on $C \log(m/\varepsilon)$ bits where $k' \le k$, and reason as before. □

6 Pseudorandomness, III

In this section we prove Theorem 22, giving generators for generalized halfspaces and combinatorial shapes. After that, we discuss the relationship between the results in this paper

and an original motivation [LV].

Lemma 48 ([GKM15])**.** *Suppose $G\colon \{0,1\}^\ell \to (\{0,1\}^n)^k$ is an explicit generator that ε-fools any product test on nk bits that reads bits in any order, then*

1. *G fools any generalized halfspace $h\colon (\{0,1\}^n)^k \to \{0,1\}$ on nk bits that reads bits in any order with error $O(k2^n(n+\log k)\varepsilon)$.*
2. *G fools any combinatorial shape $g\colon (\{0,1\}^n)^k \to \{0,1\}$ on nk bits that reads bits in any order with error $O(k^2 2^n(n+\log k)\varepsilon)$.*

Proof of Lemma 48. (1) Let $U = (U_1,\ldots,U_k)$ be the uniform distribution over $(\{0,1\}^n)^k$ and $X = (X_1,\ldots,X_k) = \pi G(U_\ell) \subseteq (\{0,1\}^n)^k$, where U_ℓ is uniform over $\{0,1\}^\ell$ and π is some permutation on nk bits. Let $Z_1 := \sum_{i\le k} g_i(U_i)$, and $Z_2 := \sum_{i\le k} g_i(X_i)$. Since G fools product tests with error ε, we have for every $\alpha \in [0,1]$,

$$|\mathrm{E}[e^{2\pi i\alpha Z_1}] - \mathrm{E}[e^{2\pi i\alpha Z_2}]| = |\mathrm{E}[\prod_{i\le k} e^{2\pi i\alpha g_i(U_i)}] - \mathrm{E}[\prod_{i\le k} e^{2\pi i\alpha g_i(X_i)}]| \le \varepsilon.$$

By [GKM15, Lemma 9.3], we may assume each $g_i(j)$ and θ are integers of absoluate value $B := (2^n k)^{O(2^n k)}$, and so $-kB \le Z_1, Z_2 \le kB$. It follows form [GKM15, Lemma 9.2] that

$$|\mathrm{E}[h(\pi G(U_\ell))] - \mathrm{E}[h(U)]| \le \max_{-kB\le t\le kB} |\Pr(Z_1 \le t) - \Pr(Z_2 \le t)| \le O(\log(kB))\varepsilon.$$

(2) Since $\sum_{i\le k} g_i(x_i) \in \{0,\ldots,k\}$, it suffices to fool the generalized halfspaces $h(x) := \sum_{i\le k} g_i(x_i) - \theta$ for $\theta \in \{0,\ldots,k\}$, the rest follows from (1) and a union bound. □

Proof of Theorem 22. Combine Lemma 48 and Theorem 19. □

Motivation: The sum of small-bias distributions. One motivation for this work comes from the paper [LV]. That paper shows that the study of bounded independence plus noise is useful in understanding the limitations of the sum of two or more independent small-bias distributions. We refer the reader to [LV] for background, but we mention briefly that while the latter distributions have been shown to fool low-degree polynomials in [BV10, Lov09, Vio09] they are also candidate to giving new circuit lower bounds or RL=L. The paper [LV] lays two approaches to exhibit *distinguishers* for the sum of small-bias distributions, and one approach is related to this work, as discussed next.

Let D be a linear b-wise independent distribution over $\{0,1\}^m$, and for a parameter η let $E(\eta)$ be a vector of m independent bits which are set to uniform with probability η and 0 otherwise. [LV] makes two observations. First, the distribution $X = D + E(\eta)$, where $+$ denotes bit-wise xor, is $\varepsilon = (1-\eta)^{b+1}$-biased. Second, by the linearity of D we have

$$X + X = D + E(\eta) + D + E(\eta) = D + E(\eta')$$

where $\eta' = 1-(1-\eta)^2 = (2-\eta)\eta < 2\eta$. Hence, the distribution $X+X$ is of the same form as X except for a slight increase in the noise parameter. This structure is useful in exhibiting

tests which are not fooled by the xor of two small-bias distributions, see [LV]. However, it was left open in [LV] whether it can be useful to answer a question posed more than 10 years ago by Reingold and Vadhan, and which we can state in the following form: is it true that for every c there exists a d such that the xor of two m^{-d}-biased distributions on m bits fools one-way algorithms using space $c \log m$? (An affirmative answer implies RL=L.)

The approach in [LV] cannot answer this question in the negative if it turns out that whenever the bias of X is m^{-d} then $X + X$ does fools $c \log m$-space algorithms.

This paper shows that this does turn out to be the case whenever $b = \Omega(m)$, which is also the setting where our bounds are tight. Indeed, in this setting we need $\eta = \Omega(d(\log m)/m)$ to have bias m^{-d}. But then Theorem 5 gives an error bound of $2^{-\Omega(d \log m)}$. This can be made less than m^{-ck} for a constant k by choosing d large enough. And this bound is sufficient to fool one-way space $c \log m$, as remarked in §1.2 before Corollary 20.

This being the failure of an approach to show a limitation, it can be interpreted with optimism.

However, already when $b = m/\log m$ our bounds are not strong enough to show that the [LV] approach fails. The bias condition gives $\eta = c(\log^2 m)/m$, and in this case our bound becomes only $2^{-\Omega(c)}$, which is not sufficient to fool space. This provides further motivation for understanding whether the bounds in Theorem 5 are tight even for $b = o(m)$, and to extend the theorem to other tests.

Bonus results. We note that for fixed order we have the following simple construction that fools product tests.

Claim 49. *Let U_1 be the uniform distribution on n bits and $D_2, \ldots, D_k$ be $k-1$ independent $\varepsilon/(k-1)$-biased distributions on n bits, then the distribution $D := (U_1, U_1 + D_2, \ldots, U_1 + D_k)$ ε-fools any product test $f\colon (\{0,1\}^n)^k \to \mathbb{C}_1$.*

An ε-biased distribution can be sampled using $O(\log(n/\varepsilon))$ bits [NN93, AGHP92]. Hence, D can be sampled using $n + O(k \log(nk/\varepsilon))$ bits, which is optimal when $k = O(1)$.

Proof of Claim 49. We will use the hybrid argument. Let $f := \prod_{i \le k} f_i$ be any product test and $\mu_i = \mathrm{E}[f_i]$. For $1 \le i \le k$, define the hybrid distribution $H_i := (U_1, \ldots, U_i, U_1 + D_{i+1}, \ldots, U_1 + D_k)$, where each U_i is independently uniformly distributed over $\{0,1\}^n$. Note that $H_1 = D$ and H_k is the uniform distribution U. The goal is to show that for $2 \le j \le k$, we have $|\mathrm{E}[H_{i-1}] - \mathrm{E}[H_i]| \le \varepsilon/(k-1)$. Then it follows that

$$|\mathrm{E}[f(D)] - \mathrm{E}[f(U)]| = |\mathrm{E}[f(H_1)] - \mathrm{E}[f(H_k)]| \le \sum_{2 \le j \le k} |\mathrm{E}[f(H_{i-1})] - \mathrm{E}[f(H_i)]| \le \varepsilon.$$

We now show that $|\mathrm{E}[H_{i-1}] - \mathrm{E}[H_i]| \le \varepsilon/(k-1)$. Note that once we have fixed the values of D_j for $j \ge i$, the corresponding f_j's has the same input as f_1. Thus we can write their products as one function. That is, for every $z_j \in D_j$ where $j > i$, we can define $g_{z_{i+1}, \ldots, z_k}\colon \{0,1\}^n \to \mathbb{C}_1$

by $g_{z_{i+1},\ldots,z_k}(x) := f_1(x)\prod_{j>i} f_j(x+z_j)$. Then

$$\begin{aligned}
|\mathrm{E}[f(H_{i-1})] - \mathrm{E}[f(H_i)]| &= |\mathrm{E}[f_1(U_1)\cdot\prod_{j=2}^{i-1} f_j(U_j)\cdot(f_i(U_1+D_i)-f_i(U_i))\cdot\prod_{j>i} f_j(U_1+D_j)]| \\
&\le (\prod_{j=2}^{i-1}\mu_j)\cdot|\mathrm{E}[(f_1(U_1)\cdot\prod_{j>i} f_j(U_1+D_j))\cdot(f_i(U_1+D_i)-f_i(U_i))]| \\
&\le |\mathop{\mathrm{E}}_{z_j\sim D_j,\forall j>i}[\mathop{\mathrm{E}}_{U_1}[g_{z_{i+1},\ldots,z_k}(U_1)f_i(U_1+D_i)] - \mathrm{E}[g_{z_{i+1},\ldots,z_k}]\cdot\mu_i]| \\
&\le \mathop{\mathrm{E}}_{z_j\sim D_j,\forall j>i}[|\mathop{\mathrm{E}}_{U_1}[g_{z_{i+1},\ldots,z_k}(U_1)f_i(U_1+D_i)] - \mathrm{E}[g_{z_{i+1},\ldots,z_k}\cdot\mu_i]|].
\end{aligned}$$

It follows from Claim 50 below that the inner expectation is at most $\varepsilon/(k-1)$, and the rest follows by averaging over the choices of the z_j's. □

Claim 50. *Let U be the uniform distribution over n bits. Let D be an ε-biased distribution over n bits. Let $f, g\colon \{0,1\}^n \to \mathbb{C}_1$ be two functions. We have $|\mathrm{E}[f(U)g(U+D)] - \mathrm{E}[f]\,\mathrm{E}[g]| \le \varepsilon$.*

Proof. We write f and g in their Fourier expansion. We have

$$\begin{aligned}
|\mathrm{E}[f(U)g(U+D)] - \mathrm{E}[f]\,\mathrm{E}[g]| &= |\mathrm{E}[(\sum_\alpha \hat{f}_\alpha\chi_\alpha(U))(\sum_\beta \hat{g}_\beta\chi_\beta(U+D))] - \hat{f}_\emptyset\hat{g}_\emptyset| \\
&= |\sum_{\alpha,\beta}\hat{f}_\alpha\hat{g}_\beta\,\mathrm{E}[\chi_{\alpha+\beta}(U)]\,\mathrm{E}[\chi_\beta(D)] - \hat{f}_\emptyset\hat{g}_\emptyset| \\
&= |\sum_{\alpha\neq\emptyset}\hat{f}_\alpha\hat{g}_\alpha\,\mathrm{E}[\chi_\alpha(D)]| \\
&\le \varepsilon\sum_{\alpha\neq\emptyset}|\hat{f}_\alpha||\hat{g}_\alpha| \\
&\le \varepsilon,
\end{aligned}$$

where the last equality is because $\mathrm{E}[\chi_\alpha(U)] = 0$ if $\alpha \neq 0$ and equals 1 otherwise, the first inequality is because D is ε-biased, and the last inequality is by Cauchy–Schwarz and the fact that f and g are bounded by 1. □

7 Conclusion

We have shown that distributions with bounded independence (or small-bias) perturbed with noise fool products. We ask for tight bounds on the error ε as a function of the amount of independence and the error parameter η, in any computational model. For products, an immediate question is to understand whether we can remove the factor of $1/k$ in the exponent in our main theorems. This would improve significantly our applications.

Our study also leads us to the following question. For simplicity we focus on the binary case $q = 2$.

Question 51. *Let $X = (X_1, X_2, \ldots, X_k)$ be an ε-biased distribution over $(\mathbb{F}_2^n)^k$. Let U be uniform over $\mathbb{F}_2^n$. Let $f_1, f_2, \ldots, f_k$ be functions from $\{0,1\}^n \to \{0,1\}$ with expectations $\mu_1, \mu_2, \ldots, \mu_k$. Is it true that*

$$|\mathop{\mathrm{E}}_{U,X}[\prod_{i \le k} f_i(U + X_i)] - \prod_{i \le k} \mu_i| \le \varepsilon'$$

for an ε' which is independent of n and that, say, goes to 0 for any fixed k and vanishing ε?

Note that the X_i may be correlated. It can be shown that the distribution $D + E$ in Theorem 1.1, when D is linear, has the above format, by writing the generator matrix in systematic form. (In fact, it is the sum of several independent samples of such distributions.)

Acknowledgments. We thank Andrej Bogdanov for useful discussions.

References

[AAK+07] Noga Alon, Alexandr Andoni, Tali Kaufman, Kevin Matulef, Ronitt Rubinfeld, and Ning Xie. Testing k-wise and almost k-wise independence. In *ACM Symp. on the Theory of Computing (STOC)*, pages 496–505, 2007.

[ABI86] Noga Alon, László Babai, and Alon Itai. A fast and simple randomized algorithm for the maximal independent set problem. *Journal of Algorithms*, 7:567–583, 1986.

[AGHP92] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures & Algorithms*, 3(3):289–304, 1992.

[AGM03] Noga Alon, Oded Goldreich, and Yishay Mansour. Almost k-wise independence versus k-wise independence. *Inf. Process. Lett.*, 88(3):107–110, 2003.

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

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

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

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

[Baz09] Louay M. J. Bazzi. Polylogarithmic independence can fool DNF formulas. *SIAM J. Comput.*, 38(6):2220–2272, 2009.

[BJMM12] Anja Becker, Antoine Joux, Alexander May, and Alexander Meurer. Decoding random binary linear codes in $2^{n/20}$: How $1 + 1 = 0$ improves information set decoding. In *Int. Conf. on the Theory and Applications of Cryptographic Techniques (EUROCRYPT)*, pages 520–536, 2012.

[BM05] Louay Bazzi and Sanjoy K. Mitter. Encoding complexity versus minimum distance. *IEEE Transactions on Information Theory*, 51(6):2103–2112, 2005.

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

[BPW12] Andrej Bogdanov, Periklis A. Papakonstantinou, and Andrew Wan. Pseudorandomness for linear length branching programs and stack machines. In *Workshop on Randomization and Computation (RANDOM)*, pages 447–458, 2012.

[Bra10] Mark Braverman. Polylogarithmic independence fools AC^0 circuits. *J. of the ACM*, 57(5), 2010.

[BV10] Andrej Bogdanov and Emanuele Viola. Pseudorandom bits for polynomials. *SIAM J. on Computing*, 39(6):2464–2486, 2010.

[BYRST02] Ziv Bar-Yossef, Omer Reingold, Ronen Shaltiel, and Luca Trevisan. Streaming through combinatorial objects. In *Seventeenth IEEE Conference on Computational Complexity*. IEEE Computer Soc., Los Alamitos, CA, 2002.

[CG89] Benny Chor and Oded Goldreich. On the power of two-point based sampling. *Journal of Complexity*, 5(1):96–106, 1989.

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

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

[CW79] J. Lawrence Carter and Mark N. Wegman. Universal classes of hash functions. *J. of Computer and System Sciences*, 18(2):143–154, 1979.

[De15] Anindya De. Beyond the central limit theorem: Asymptotic expansions and pseudorandomness for combinatorial sums. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 883–902, 2015.

[DGJ+10] Ilias Diakonikolas, Parikshit Gopalan, Ragesh Jaiswal, Rocco A. Servedio, and Emanuele Viola. Bounded independence fools halfspaces. *SIAM J. on Computing*, 39(8):3441–3462, 2010.

[DKN10] Ilias Diakonikolas, Daniel Kane, and Jelani Nelson. Bounded independence fools degree-2 threshold functions. In *51th IEEE Symp. on Foundations of Computer Science (FOCS)*. IEEE, 2010.

[EGL+92] Guy Even, Oded Goldreich, Michael Luby, Noam Nisan, and Boban Velickovic. Approximations of general independent distributions. In *ACM Symp. on the Theory of Computing (STOC)*, pages 10–16, 1992.

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

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

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

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

[GMRZ13] Parikshit Gopalan, Raghu Meka, Omer Reingold, and David Zuckerman. Pseudorandom generators for combinatorial shapes. *SIAM J. on Computing*, 42(3):1051–1076, 2013.

[GOWZ10] Parikshit Gopalan, Ryan O'Donnell, Yi Wu, and David Zuckerman. Fooling functions of halfspaces under product distributions. In *25th IEEE Conf. on Computational Complexity (CCC)*, pages 223–234. IEEE, 2010.

[Gro06] Andre Gronemeier. A note on the decoding complexity of error-correcting codes. *Information Processing Letters*, 100(3):116–119, 2006.

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

[Hås87] Johan Håstad. *Computational limitations of small-depth circuits*. MIT Press, 1987.

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

[IMP12] Russell Impagliazzo, William Matthews, and Ramamohan Paturi. A satisfiability algorithm for AC^0. In *ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 961–972, 2012.

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

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

[Lov09] Shachar Lovett. Unconditional pseudorandom generators for low degree polynomials. *Theory of Computing*, 5(1):69–82, 2009.

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

[LV] Chin Ho Lee and Emanuele Viola. Some limitations of the sum of small-bias distributions. *Theory of Computing*.

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

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

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

[Raz09] Alexander A. Razborov. A simple proof of Bazzi's theorem. *ACM Transactions on Computation Theory (TOCT)*, 1(1), 2009.

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

[Shp09] Amir Shpilka. Constructions of low-degree and error-correcting epsilon-biased generators. *Computational Complexity*, 18(4):495–525, 2009.

[Sub61] B. A. Subbotovskaya. Realizations of linear functions by formulas using +, *, -. *Soviet Mathematics-Doklady*, 2:110–112, 1961.

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

[Tal14] Avishay Tal. Tight bounds on The Fourier Spectrum of AC^0. *Electronic Colloquium on Computational Complexity*, Technical Report TR14-174, 2014. www.eccc.uni-trier.de/.

[Vio09] Emanuele Viola. The sum of d small-bias generators fools polynomials of degree d. *Computational Complexity*, 18(2):209–217, 2009.

[Vio12] Emanuele Viola. The complexity of distributions. *SIAM J. on Computing*, 41(1):191–218, 2012.

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

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

[Yao85] Andrew Yao. Separating the polynomial-time hierarchy by oracles. In *26th IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 1–10, 1985.

A A lower bound on b and η

For completeness we now prove by standard arguments a lower bound on b and η such that a result as in Theorem 5 may apply. We obviously require $b \geq n$ if $\eta = 0$, whereas for general η one requires a more elaborate argument. Let $k = 1$ and let M be a uniformly chosen $n \times t$ matrix over $\mathbb{F}_q$. The probability that the corresponding code has minimum distance $\leq d$ is at most $q^t V_q(d)/q^n$. Hence a code C' exists with minimum distance $> d$ for $n - t = \lceil \log_q V_q(d) \rceil$. By Fact 11 the uniform distribution D over the dual C of C' is d-uniform. This distribution can be generated by an $n \times (n-t)$ matrix. Hence the support size of this distribution is $q^{n-t} \leq O(V_q(d))$.

Moreover, by Lemma 34 we can sample with $O(\eta \log(q/\eta)n)$ bits a distribution that is $2^{-\Omega(\eta n)}$-close to the noise vector E.

Hence $D+E$ is $2^{-\Omega(\eta n)}$-close to a distribution supported on a set S of size $O(V_q(d))2^{O(\eta \log(q/\eta)n)} \leq 2^{d \log O(enq/d) + O(\eta \log(q/\eta)n)}$. The function f_1 is taken to be the characteristic function of S. By the lemma the function outputs 1 on $D + E$ with probability $1 - 2^{-\Omega(\eta n)}$.

On the other hand, the function outputs 1 on a uniform input with probability $1 - |S|/q^n$. In particular, for any $d = (1-\varepsilon)n$ and sufficiently large q this shows that f_1 has a constant distinguishing advantage for all η less ε', where ε' depends only on ε.