

Some limitations of the sum of small-bias distributions

Chin Ho Lee

Emanuele Viola *

January 5, 2015

Abstract

We exhibit ε -biased distributions D on n bits and functions $f: \{0,1\}^n \rightarrow \{0,1\}$ such that the xor of two independent copies ($D + D$) does not fool f , for any of the following choices:

1. $\varepsilon = 2^{-\Omega(n)}$ and f is in P/poly;
2. $\varepsilon = 2^{-\Omega(n/\log n)}$ and f is in NC²;
3. $\varepsilon = n^{-\log^{\Omega(1)} n}$ and f is in AC⁰;
4. $\varepsilon = n^{-c}$ and f is a one-way space $O(c \log n)$ algorithm, for any c ;
5. $\varepsilon = n^{-0.029}$ and f is a mod 3 linear function.

All the results give one-sided distinguishers, and extend to the xor of more copies for suitable ε .

Meka and Zuckerman (RANDOM 2009) prove 5 with $\varepsilon = O(1)$. Bogdanov, Dvir, Verbin, and Yehudayoff (Theory Of Computing 2013) prove 2 with $\varepsilon = 2^{-O(\sqrt{n})}$. Chen and Zuckerman (personal communication) give an alternative proof of 4.

1-4 are obtained via a new and simple connection between small-bias distributions and error-correcting codes. We also give a conditional result for DNF formulas, and show that 5-wise independence does not hit mod 3 linear functions.

*College of Computer and Information Science, Northeastern University. Supported by NSF grant CCF-1319206. Email: {chlee,viola}@ccs.neu.edu. Work partially done while visiting Harvard University.

1 Introduction and our results

Small-bias distributions, introduced by Naor and Naor [NN93], cf. [ABN⁺92, AGHP92, BT13], are distributions that look random to parity functions over $\{0, 1\}^n$. They can be generated using a seed of $O(\log n/\varepsilon)$ bits, if each parity outputs 1 with a probability that is within ε of 1/2. Since their introduction, they have become a fundamental object in theoretical computer science and have found their uses in many areas including derandomization and algorithm design.

In the last decade or so researchers have considered the sum (i.e., bit-wise xor) of several independent copies of small-bias distributions. The first paper to explicitly consider it is [BV10a]. This distribution appears to be significantly more powerful than a single small-bias copy, while retaining a modest seed length. In particular, two main questions have been asked:

Question: RL. Reingold and Vadhan (personal communication) asked whether the sum of two copies of $1/\text{poly}(n)$ -bias distributions fools one-way logarithmic space, aka one-way polynomial-size branching programs, which would imply $\text{RL} = \text{L}$. It is known that a small-bias distribution fools one-way width-2 branching programs (Saks and Zuckerman, see also [BDVY13] where a generalization is obtained). We are not aware of any result showing that more copies help for width-3 programs.

Question: polynomials. The papers [BV10a, Lov09, Vio09b] show that the sum of d small-bias generators fools $\text{GF}(2)$ polynomials of degree d . (We note that by replacing Or with Parity on a random subset of the inputs these results also apply to d -DNF.) However, the proofs only apply when $d \leq (1 - \Omega(1)) \log n$. Still, the construction is candidate to working even for larger d . If true, that would make progress on long-standing open problems in circuit complexity regarding AC^0 with parity gates [Raz87], cf. the survey [Vio09a, Chapter 1].

In this space we highlight the following basic question: what is the smallest $\varepsilon_2 = \varepsilon_2(\varepsilon)$ such that the xor of any two ε -biased distributions over $\{0, 1\}^n$ ε_2 -fools the inner product polynomial $x_1x_2 + x_3x_4 + \dots + x_{n-1}x_n \pmod{2}$? We only know $1.99\varepsilon < \varepsilon_2 \leq O(\sqrt{\varepsilon})$. The details of the first inequality are omitted. The second can be found in [BV10a, Vio09b].

In terms of negative results, Meka and Zuckerman [MZ09] show that the sum of 2 distributions with constant bias does not fool mod3 linear functions. Bogdanov, Dvir, Verbin, and Yehudayoff [BDVY13] show that for $\varepsilon = 2^{-O(\sqrt{n}/k)}$, the sum of k ε -biased distributions does not fool circuits of size $\text{poly}(n)$ and depth $O(\log^2 n)$ (NC^2).

In this work we improve on both these results and obtain other limitations of the sum of small-bias distributions.

1.1 Our results

The following theorem states our main counterexamples. We denote by $D + D$ the bit-wise xor of two independent copies of a distribution D .

Theorem 1. *There exists an explicit ε -biased distribution D over $\{0, 1\}^n$ and an explicit function f , such that $f(D + D) = 0$ and $\Pr_{x \sim \{0, 1\}^n}[f(x) = 0] \leq p$, where ε, f, p are of any one of the following choices:*

- i. $\varepsilon = 2^{-\Omega(n)}$, f is a $\text{poly}(n)$ -size circuit, and $p = 2^{-\Omega(n)}$;
- ii. $\varepsilon = 2^{-\Omega(n/\log n)}$, f is a fan-in 2, $\text{poly}(n)$ -size circuit of depth $O(\log^2 n)$, and $p = 2^{-n/3}$;
- iii. for any c , $\varepsilon = n^{-\log^c n}$, f is an $n^{O(c)}$ -size circuit of unbounded fan-in and depth $O(c)$ (AC^0), and $p = n^{-\log^{\Omega(1)} n/3}$;
- iv. for any c , $\varepsilon = 1/n^c$, f is a one-way $O(c \log n)$ -space algorithm, and $p = O(1/n^c)$;
- v. $\varepsilon = n^{0.029}$, f is a mod3 linear function, and $p = 1/2$.

Moreover, all our results extend to more copies of D as follows. The input $D + D$ to f can be replaced by the bit-wise xor of k independent copies of D if ε is replaced by $\varepsilon^{2/k}$, where k is at most the following quantities corresponding to the above items: i. $n/60$; ii. $n/6 \log n$; iii. $\log^{c+1} n/6(c+1) \log \log n$; iv. $2c$; v. $O(\log n / \log \log n)$.

Theorem 1.i is tight up to the constant in the exponent because every $\varepsilon 2^{-n}$ -biased distribution is ε -close to uniform.

Theorem 1.ii would also be true with $\varepsilon = 2^{-\Omega(n)}$, if a decoder for certain algebraic-geometric codes runs in NC^2 , which we conjecture it does. [BDVY13] prove Theorem 1.ii with $\varepsilon = 2^{-O(\sqrt{n}/k)}$.

Theorem 1.iii is tight in the sense that $n^{-(\log n)^{O(d)}}$ bias fools AC^0 circuits of size n^d and depth d , as shown in the sequence of works [Baz09, Raz09, Bra09, Tal14].

Theorem 1.iv can also be obtained in the following way, pointed out to us by Chen and Zuckerman (personal communication). Observe that since one can distinguish a set of size s from uniform with a width $s+1$ branching program, and there exist ε -bias distributions with support size $O(n/\varepsilon^2)$, the sum of two such distributions can be distinguished from uniform in space $O(c \log n)$ when $\varepsilon = n^{-c}$. Both their proof and ours (presented later) apply to $c > 0.01$; but for smaller c Theorem 1.v kicks in.

Theorems 1.iv and 1.v come close to addressing the “RL question,” without answering it: 1.v shows that polynomial bias is necessary even for width-3 regular branching programs, while 1.iv shows that the bias is at least polynomial in the width. [MZ09] prove Theorem 1.v with $\varepsilon = \Omega(1)$.

We have not been able to say anything on the “Polynomials question.”

There exist other models of interest. For read-once DNF no counterexample with large error is possible because Chari, Rohatgi, and Srinivasan [CRS00], building on [EGL⁺98], show that (just one) $n^{-O(\log^{1/\delta})}$ -bias distribution fools any read-once DNF on n variables with error δ , cf. Appendix A. The [CRS00] result is rediscovered by De, Etesami, Trevisan, and Tulsiani [DETT10], who also show that it is essentially tight by constructing a distribution which is $n^{\Omega(\log^{1/\delta})/\log \log^{1/\delta}}$ -biased yet does not δ -fool a read-once DNF. In particular, fooling with polynomial error requires super-polynomial bias.

It would be interesting to know whether the xor of two copies overcomes this limitation, i.e., if it δ -fools any read-once DNF on n variables if each copy has bias $\text{poly}(n/\delta)$. If true, this would give a generator with seed length $O(\log n/\delta)$, which is open.

We are unable to resolve this for read-once DNF. However we show that the corresponding result for general DNF implies long-standing circuit lower bounds [Val77]. This can be interpreted as saying that such a result for DNF is either false or extremely hard to prove. We also get a conditional counterexample for depth-3 circuits.

Theorem 2. *Suppose polynomial-time (P) has fan-in 2 circuits of linear size and logarithmic depth. Theorem 1 also applies to the following choices of parameters:*

- i. $\varepsilon = n^{-\omega(1)}$, f is a depth-3 circuit of size $n^{o(1)}$ and unbounded fan-in, and $p = n^{-\omega(1)}$.
- ii. $\varepsilon = n^{-\omega(1)}$, f is a DNF formula, and $p = 1 - 1/n^{o(1)}$.

Moreover, all our results extend to more copies of D as follows. The input $D + D$ to f can be replaced by the bit-wise xor of $k \leq \log n$ independent copies of D if ε is replaced by $\varepsilon^{2/k}$.

All the above results except Theorem 1.v are based on a new, simple connection between small-bias generators and error-correcting codes, discussed in §1.2. We also go the other way around and obtain some results on the complexity of decoding. For example, we show some limitations for low-degree polynomials on detecting codewords with errors (Claim 30), and that for codes with large minimum distance, AC⁰ and read-once branching programs cannot decode when the number of errors is close to half of the minimum distance of a code (Claim 29).

Theorem 1.v instead follows [MZ09] and bounds the mod 3 dimension of small-bias distributions, which is the dimension of the subspace spanned by the support of the distributions over GF(3). It turns out that upper bounds on the mod3 dimension of k -wise independent distributions would allow us to reduce the bias in Theorem 1.v, assuming long-standing conjectures on correlation bounds for low-degree polynomials (which may be taken as standard):

Claim 3. *Suppose*

- 1. *the parity of k copies of mod3 parity on disjoint inputs of length m has correlation $2^{-\Omega(k)}$ with any GF(2) polynomial of degree \sqrt{m} , and*
- 2. *for every c , there exists a $c \log n$ -wise independent distribution whose support on $\{0, 1\}^n \subseteq GF(3)^n = \{0, 1, 2\}^n$ has mod3 dimension $n^{0.49}$.*

Then the “RL question” is false, i.e., there exists an $n^{-\omega(1)}$ -biased distribution D such that $D + D$ does not fool a one-way $O(\log n)$ -space algorithm.

Counterpositively, an affirmative answer to the “RL question,” even for permutation, width-3 branching programs, implies lower bounds on the mod3 dimension of k -wise independent distributions, or that the aforementioned correlation bounds are false.

Therefore, we initiate a systematic study of the mod3 dimension of (almost) k -wise independent distributions, and obtain the following lower and upper bounds. First, we give an $\Omega(k \log n)$ lower bound for almost k -wise independent distributions, specifically, distributions such that any k coordinates are 1/10 close to being uniform over $\{0, 1\}^k$ (Claim 14).

This also gives an exponential separation between mod 3 dimension and seed length for such distributions.

We then prove the following upper bounds, see Claim 23 and Claim 27.

Theorem 4. *For infinitely many n , there exists k -wise independent distributions over $\{0, 1\}^n$ with mod3 dimension d for any of the following choices of k and d :*

- i. $k = 2, d \leq n^{0.72}$;
- ii. $3 \leq k \leq 5, d \leq n - 1$.

We note that an upper bound of $n - 1$ on the mod3 dimension of a k -wise independent distribution is equivalent to saying that there exists a k -wise independent distribution that is supported on the set of binary strings whose Hamming weight is divisible by three. In particular, such a distribution does not fool mod3 in a strong, one-sided sense. We ask what is the largest $k^* = k^*(n)$ such that there exists a k -wise independent distribution with mod 3 dimension $\leq n - 1$. We present a general framework using duality and symmetrization towards obtaining such bounds. By combining our framework with a computer program we have shown that for every $n \leq 900$, $k^* \geq n/3$. Hence, we conjecture the bound $k^*(n) = \Omega(n)$. Analytically, we show $k^* \geq 5$ for infinitely many n , as reported in Theorem 4.ii.

1.2 Our techniques

All our counterexamples in Theorem 1 and 2, except Theorem 1.v, come from a new connection between small-bias distributions and linear codes, which we now explain. Let $C \subseteq \mathbb{F}^n$ be a linear error correcting code over a field of characteristic 2. Abusing notation, we also use C to denote the uniform distribution over the code C . It is well-known that if C^\perp has minimum distance d^\perp , then C is $(d^\perp - 1)$ -wise independent.

Define N_e to be the “noise” distribution over \mathbb{F}^n obtained by repeating the following process e times: Pick a uniformly random position from $[n]$, and set it to a uniform symbol in \mathbb{F} . Then every linear test has zero bias if one of its positions is set to random. Thus, N_e has bias at most $(1 - d^\perp/n)^e$ over tests of size at least d^\perp .

Now, define D_e to be the small-bias distribution obtained from adding N_e to C , and we have the following fact.

Fact 5. D_e is $(1 - d^\perp/n)^e$ -biased.

Our main observation is that *the xor of two noisy codewords is also a noisy codeword*, with the number of errors injected to the codeword doubled. That is,

$$D_e + D_e = C + N_e + C + N_e = C + N_{2e} = D_{2e}.$$

Now suppose an algorithm can distinguish whether a string is within $2e$ errors of a codeword or if it is uniform. Then it can be used to distinguish $D_e + D_e$ from uniform. More generally, if a circuit can distinguish a codeword with ke errors from a uniform string, then it can distinguish the xor of k independent copies of D_e from uniform. Contrapositively, if $D + D$ fools f , then f cannot detect codewords in C within $2e$ errors. Thus, to obtain counterexamples we only have to exhibit appropriate distinguishers. We achieve this by drawing from results in coding theory. This is explained below after a remark.

Remark 1. Our distinguisher is only required to tell apart noisy codewords and uniform random strings. This is a weaker condition than decoding. In fact, similar distinguishers have been considered in the context of tolerant property testing [GR05, KS09, RU10], where tolerant testers are designed to decide if the input is close to being a codeword or far from every codeword, by looking at as few positions of the input as possible.

We also note that our connection between ε -bias distributions and linear codes is different from the well-known connection in [NN93], which shows that for a binary linear code with relative minimum and maximum distance $\geq 1/2 - \varepsilon$ and $\leq 1/2 + \varepsilon$ respectively, the columns of its $k \times n$ generator matrix form the support of an ε -biased distribution over $\{0, 1\}^k$. However, the connection to codes is lost once we consider the sum of the same distributions. In contrast, the sum of our distributions bears the code structure of a single copy.

For general circuits (Theorem 1.i), we consider the asymptotically good binary linear code with constant dual relative distance, based on algebraic geometry and exhibited by Guruswami in [Shp09]. We conjecture that the corresponding distinguisher can be implemented in NC^2 . However we are unable to verify this. Instead, for NC^2 circuits (Theorem 1.ii), we use Reed-Solomon codes and the Peterson-Gorenstein-Zierler syndrome-decoding algorithm [Pet60, GZ61] which we note is in NC^2 . By scaling the NC^2 result down to polylog n bits followed by a depth reduction we obtain our results for AC^0 circuits (Theorem 1.iii). This result could also be obtained by scaling down a result in [BDVY13], although it was not stated there.

Our counterexample for one-way log-space computation (Theorem 1.iv) also uses Reed-Solomon codes. The decoder is simply syndrome decoding: from e errors it can be realized by computing the syndrome in a one-way fashion using space $O(e \log q)$, where q is the size of the underlying field of the code. For a given constant c , setting $q = n$, $k = d^\perp - 1 = n - O(c)$, and $e = O(c)$ we obtain a one-way space $O(c \log n)$ distinguisher for the sum of two distributions with bias n^{-c} .

Naturally, one might try to eliminate the dependence on c in the $O(c \log n)$ space bound with a different choice of e and q , which would answer the “RL question” in the negative. In Claim 8 however we show that to obtain n^{-c} bias, the space $e \log q$ for syndrome decoding must be of $\Omega(c \log n)$, regardless of the code and the alphabet. Thus our result is the best possible that can be obtained using syndrome decoding. We raise the question of whether syndrome decoding is optimal for one-way decoding in this setting of parameters, and specifically if it is possible to devise a one-way decoding algorithm using space $o(e \log q)$. There do exist alternative one-way decoding algorithms, cf. [RU10], but apparently not for our setting of parameters of $e = O(1)$ and $k = n - O(1)$.

Our conditional result for depth-3 circuits and DNF formulas (Theorem 2) follows from scaling down to barely superlogarithmic input length, and a depth reduction [Val77] (cf. [Vio09a, Chapter 3]) of the counterexample for general circuits (Theorem 1.i). We note that the $2^{-\Omega(n)}$ -bias in Theorem 1.i is essential for this result, in the sense that $2^{-n/\log n}$ -bias would be insufficient to obtain Theorem 2. We also remark that since $O(\log^2 n)$ -wise independence suffices to fool DNF formulas [Baz09], one must consider linear codes with dual distance less than $\log^2 n$ in our construction, and so D has bias at most $2^{-O(\log^2 n)}$.

The connection between codes and small-bias distributions motivate us to study further the complexity of decoding. [Vio06, Chapter 6] and [SV10], cf. [Vio06, Chapter 6], show that

list-decoding requires computing the majority function. In Claim 29 we extend their ideas and prove that the same requirement holds even for decoding up to half of the minimum distance. This gives some new results for AC^0 and for branching programs. Finally, since $\log^{O(1)} n$ -wise independence fools AC^0 [Bra09, Tal14], we obtain that AC^0 cannot distinguish a codeword from a code with $\log^{\Omega(1)} n$ dual distance from uniform random strings. This also gives some explanation of why scaling is necessary to obtain Theorem 1.iii from Theorem 1.ii.

Theorem 1.v. Meka and Zuckerman [MZ09] construct the following constant-bias distribution D over $n := \binom{d}{5}$ bits with mod3 dimension less than \sqrt{n} : Each output bit is the square of mod3 of 5 out of the d uniform random bits, which can be written as a degree-5 GF(2)-polynomial. Since any parity of the output bits is also a degree-5 polynomial over $\{0, 1\}^d$, D has constant bias. To show that $D + D$ does not hit a mod3 function, they observe that D has mod3 dimension at most $d^2 < \sqrt{n}$, and from Fact 10 that $D + D$ has mod3 dimension at most $(d^2)^2 = d^4 < n$.

We extend their construction using ideas from the Nisan-Wigderson generator [NW94]: We pick a design consisting of n sets where each set has size n^β and the intersection of any two sets has size $\log n$. Such design exists provided the universe has size $n^{2\beta}$. The output distribution is again the square of mod3 on each set.

For any test of size at least $\log n$ bits, let J be any $\log n$ bits of the test. We fix the intersections of their corresponding sets in the universe to make them independent. After we do this, every bit in J is still a mod3 function on $n^\beta - |J| \log n \geq 0.9n^\beta$ bits.

We further fix every bit outside the $|J|$ sets in the universe. This will not affect the bits in J . Now consider any bit b in the test that is not in J , this corresponds to a set which has intersection at most $\log n$ with each of the sets corresponding to the bits in J . Thus, b is now a mod3 function on at most $|J| \log n = \log^2 n$ input bits and thus can be written as a degree $\log^2 n$ GF(2) polynomial. Hence, the parity of the bits outside J is also a GF(2) polynomial of the same degree, and we call this polynomial p .

Observe that the bias of the test equals to the correlation between the parity of the bits in J and p . Since each bit in J is a mod3 function on n^β bits, it has constant correlation with p . In Lemma 13 we prove a variant of Impagliazzo's XOR lemma [Imp95] to show that the xor of $\log n$ independent such bits makes the correlation drop from constant to $\varepsilon = n^{-\beta/4}$. This variant of XOR lemma may be folklore, but we are not aware of any reference.

This handles tests of size at least $\log n$. For smaller tests we xor the above distribution with an $1/n^{\Omega(1)}$ -almost $\log n$ -wise independent distribution, which gives us ε bias for tests less than $\log n$ and has sufficiently small dimension. We then show that the xor of the two distributions has dimension less than \sqrt{n} and conclude as in the previous paragraph.

Organization. In §2 we describe our counterexamples and prove Theorem 1 and 2, and Claim 3. In §3 we prove our lower bounds and upper bounds on the mod3 dimension of k -wise independence. The results on the complexity of decoding are in §4.

2 Our counterexamples

We are now ready to prove Theorem 1 and 2, and Claim 3. We consider linear codes with different parameters, the bias of D follows from Fact 5. Then we present our distinguishers. In the end, we explain how our results hold for k copies instead of 2.

2.1 General circuits

Venkatesan Guruswami [Shp09] exhibits the following family of constant-rate binary linear codes whose primal and dual relative minimum distance are both constant.

Theorem 6 (Theorem 4 in [Shp09]). *For infinitely many n , there exists a binary linear $[n, n/2]$ code C which can be constructed, encoded, and decoded from $n/60$ errors in time $\text{poly}(n)$. Moreover, the dual of C has minimum distance at least $n/30$.*

Proof of Theorem 1.i. Applying Fact 5 with $e = n/120$ to the code in Theorem 6, we obtain a distribution D that is $2^{-n/1800}$ -biased. To detect C within $2e$ errors, the detector f decodes and encodes the input, and accepts if and only if the input and the re-encoded string differ by at most $2e$ blocks. Since both the encoding and decoding algorithms run in polynomial time, so does f .

Note that f accepts at most $2^{n/2} \cdot \sum_{i=0}^{2e} \binom{n}{i} = 2^{n/2} \cdot 2^{2nH(1/60)+o(n)} \leq 2^{0.75n}$ possible strings, where $H(\cdot)$ is the binary entropy function. Hence, f distinguishes $D + D$ from the uniform distribution with probability at least $1 - 2^{-0.25n}$. \square

2.2 NC² circuits

Proof of Theorem 1.ii. Consider the $[q, q/2, q/2 + 1]$ Reed-Solomon code C over $\mathbb{F}_{2^{\log q}}$. C has dual minimum distance $q/2 + 1$ and can decode from $q/4$ errors. Applying Fact 5 to C with $e = q/12$, we obtain a distribution D over $n := q \log q$ bits that is $2^{-\Omega(n/\log n)}$ -biased.

Let α be a primitive element of $\mathbb{F}_{2^{\log q}}$. Let H be a parity check matrix of C . We first recall the Peterson-Gorenstein-Zierler syndrome-decoding algorithm [Pet60, GZ61].

Given a corrupted codeword y , let $(s_1, \dots, s_{q/2})^T := Hy$ be the syndrome of y . Suppose y has $v < q/2$ errors. Let E denote the set of its corrupted positions. Let $\Lambda_v(x) := \prod_{i \in E} (x - \alpha^i) = 1 + \sum_{i=1}^v \lambda_i x^i$ be the error locator polynomial. The syndromes and the coefficients of Λ are linearly related by

$$\lambda_v s_{j-v} + \lambda_{v-1} s_{j-t+1} + \dots + \lambda_1 s_{j-1} = 0,$$

for $j > v$. These form a linear system with unknowns λ_i 's. The algorithm decodes by attempting to solve the corresponding linear systems with v errors, where v ranges from e to 1.

Note that the system has a unique solution if and only if y and some codeword differ by exactly v positions, for some v between 1 and $2e$. Thus, our detector f computes the determinants of the $2e < q/4$ systems and accepts if and only if one of them is nonzero. Since computing determinant is in NC² [Ber84], f can be computed by an NC² circuit. The system always has a solution when inputs are under $D + D$ and so f always accepts. On

the other hand, f accepts at most $q^{q/2} \cdot \sum_{i=0}^{2e} \binom{q}{i} (q-1)^i \leq q^{q/2} \cdot 2^{qH(1/6)+o(1)} q^{q/6} \leq 2^{2n/3+o(n)}$ possible strings, where $H(\cdot)$ is the binary entropy function. Therefore, f distinguishes $D + D$ from the uniform distribution with probability at least $1 - 2^{-n/3}$. \square

2.3 Constant-depth circuits

By scaling down the previous counterexample to polylog n bits and applying standard depth reduction, we have the following result for AC^0 circuits.

Proof of Theorem 1.iii. Let D and f be the distribution and distinguisher in Theorem 1.ii respectively. Since f is a fan-in 2 circuit of $\text{poly}(n)$ -size and depth $O(\log^2 n)$, it can be evaluated by traversing the circuit once while maintaining the values of the evaluated subtree at each level, which takes time $\text{poly}(n)$ and space $O(\log^2 n)$.

Let $D_{n'}$ and $f_{n'}$ be the scaled distribution and distinguisher of D and f on $n' = \log^{c+1} n$ bits respectively (We set the rest of the $n - n'$ bits at uniformly random). $D_{n'}$ has bias $2^{-\Omega(\frac{n'}{\log n'})} = n^{-\Omega(\log^c n)}$. From the above observation, $f_{n'}$ runs in time $\log^{cb} n$ for some universal constant b and space $S := O((\log \log n)^2)$, and distinguishes $D_{n'} + D_{n'}$ from uniform with probability $1 - n^{-\log^c n/3}$. It follows from the following claim that $f_{n'}$ can be implemented by a circuit of polynomial-size and depth $O(c)$. \square

Claim 7. *Every problem on n bits that is decidable in time $\log^c n$ and space $S = o(\log n)$ can be decided by a circuit of polynomial-size and depth $2c + O(1)$.*

Proof. Let A be the decider. Consider all the possible configurations of A . Define $\phi^i: \{0, 1\}^{2S} \rightarrow \{0, 1\}$ so that $\phi^i(C_s, C_t)$ takes two configurations C_s and C_t , and accepts if and only if on input x , A 's configuration goes from C_s to C_t in $(\frac{\log n}{S} - 1)^i$ steps. ϕ^i can be written recursively as

$$\phi^i(C_s, C_t) := \bigvee_{C_s = C_1, C_2, \dots, C_{\frac{\log n}{S}} = C_t} \bigwedge_{j=1}^{\frac{\log n}{S} - 1} \phi^{i-1}(C_j, C_{j+1}).$$

Note that $\phi^0(C_s, C_t)$ has a constant size circuit and we have $|\phi^i| = \left(\frac{2^S}{\frac{\log n}{S}}\right)(\frac{\log n}{S} - 1)|\phi^{i-1}|$. Therefore, A can be implemented by a circuit of polynomial-size and depth $2c + O(1)$. \square

2.4 One-way log-space computation

Proof of Theorem 1.iv. Consider the $[q, q - 6c, 6c + 1]_{2^{\log q}}$ Reed-Solomon code C over $\mathbb{F}_{2^{\log q}}$, which has dual minimum distance $6c$ and can decode from $3c$ errors. Applying Fact 5 to C with $e = c$, we obtain a distribution D over $n := q \log q$ bits that is $(\log n/n)^c$ -biased.

Let H be a parity check matrix of C . On input $y \in \mathbb{F}_{2^{\log q}}^q$, our distinguisher f computes s_{2e+1}, \dots, s_{4e} from the syndrome $s := Hy$. Clearly this can be implemented in one-pass and $(4e + O(1)) \log q$ space. Finally f accepts if and only if there exists a u with Hamming weight at most $2e$ such that $Hu = Hy$.

Since f accepts at most $q^{q-3c} \cdot O(q^{2c}) = O(q^{q-c})$ strings, f distinguishes $D + D$ from uniform with probability $1 - O(\log n/n)^c$. \square

Computing the input for syndrome decoding requires space $2e \log q$. We now show that in order to obtain n^{-c} bias via our construction, $2e \log q = \Omega(c \log n)$.

Claim 8. *Let C be a $[n, k, d]$ code over \mathbb{F}_q which decodes from e errors, and d^\perp be its dual minimum distance. If C satisfies $(1 - d^\perp/n)^e < n^{-c}$ for sufficiently large c , then we have $e \log q = \Omega(c \log n)$.*

Proof. If $d^\perp > (1 - 1/q)n$, then by Plotkin bound on the dual code, $n - k = O(1)$. By Singleton bound, $e \leq d \leq n - k$ and so we have $e = O(1)$. Hence, $(1 - d^\perp/n)^e > 1/n^c$, and therefore the condition is not satisfied.

On the other hand, suppose $d^\perp \leq (1 - 1/q)n$. Then $(1 - d^\perp/n)^e \geq (1/q)^e$. The condition $(1/q)^e \leq n^{-c}$ implies $e \log q \geq 2c \log n$. \square

2.5 Depth 3 circuits and DNF formulas

Proof of Theorem 2. First we apply Valiant's depth reduction [Val77, Val83], cf. [Vio09a, Theorem 25], to f in Theorem 1.i, which gives us an unbounded fan-in depth-3 circuit f' of size $2^{n/\log\log n}$. Then we scale down n to $n' = \log n \log\log\log\log n$ bits (We set the rest of the $n - n'$ bits at uniformly random) to get an $n^{-\omega(1)}$ -biased distribution $D_{n'}$ and a circuit $f'_{n'}$ of size $n^{o(1)}$ and depth 3 that distinguishes $D_{n'} + D_{n'}$ from uniform with probability at least $1 - n^{-\omega(1)}$. This proves Theorem 1.i.

To prove Theorem 1.ii, note that $f'_{n'}$ accepts with probability 1 under $D_{n'} + D_{n'}$ and without loss of generality we can assume $f'_{n'}$ is an AND-OR-AND circuit. Hence, it contains a DNF f'' such that (1) f'' accepts under $D_{n'} + D_{n'}$ with probability 1, and (2) f'' rejects with probability at least $1/2n^{o(1)}$ under the uniform distribution. \square

2.6 Mod 3 linear functions

First we define our key concept.

Definition 9. *Identify \mathbb{F}_3 with $\{0, 1, 2\}$. Let $S \subseteq \mathbb{F}_3^n$ be a set of vectors. Define the mod 3 dimension of S , denoted by $\dim_3(S)$, to be the dimension of its spanning subspace over \mathbb{F}_3 . We also define the mod 3 dimension of a distribution D to be the mod 3 dimension of its support.*

Fact 10 (Lemma 7.1 and 7.2 in [MZ09]). *Let S be a set of vectors in $\{0, 1\}^n \subseteq \mathbb{F}_3^n$. Define S^2 to be the set $\{x \times_3 x : x \in S\}$, where $x \times_3 y$ denote the pointwise product of two vectors x and y (over \mathbb{F}_3). Then (1) $\dim_3(S^2) \leq \dim_3(S)^2$ and (2) $\dim_3(S +_2 S) \leq \dim_3(S) + \dim_3(S)^2$.*

Proof. Let $d = \dim_3(S)$ and $\{\beta_1, \dots, \beta_d\}$ be a basis of S . Let $x = \sum_{i=1}^d c_i \beta_i$ and $y = \sum_{j=1}^d d_j \beta_j$ be any two vectors in S . We have

$$x \times_3 x = \sum_{i,j \in [d]} c_i c_j (\beta_i \times_3 \beta_j).$$

Thus $\{\beta_i \times_3 \beta_j\}_{i,j \in [d]}$ forms a basis of S^2 , proving (1). For (2), observe that for any $a, b \in \mathbb{F}_3$, $a +_2 b = a +_3 b +_3 a \times_3 b$. Hence we have

$$x +_3 y +_3 x \times_3 y = \sum_{i=1}^d (c_i + d_i) \beta_i + \sum_{i,j \in [d]} c_i d_j (\beta_i \times_3 \beta_j),$$

and thus $\{\beta_i\}_{i \in [d]} \cup \{\beta_i \times_3 \beta_j\}_{i,j \in [d]}$ forms a basis of $S +_2 S$. \square

The following lemma is well-known (cf. [Nis91]). We include a proof here for completeness.

Lemma 11. *There exists a design (S_1, \dots, S_n) over the universe $[d]$ such that*

1. $|S_i| = t$ for every $i \in [d]$, and
2. $|S_i \cap S_j| \leq \hat{t}$ for every $i \neq j \in [d]$,

where $d = n^{2\beta}$, $t = n^\beta$, $\hat{t} = \log n$ for any $\beta < 0.5$.

Proof. It suffices to show that given S_1, \dots, S_{i-1} , there exists a set S such that $|S| \geq t$ and $|S \cap S_j| \leq \hat{t}$ for $j < i$. Consider picking each element in $[d]$ with probability $p = 0.1 \log n / n^\beta$. We have $E[|S|] = pd \geq 2n^\beta$. By Chernoff bound,

$$\Pr[|S| < t = n^\beta] \leq 2^{-n^\beta/4} < 1/2.$$

We also have $E[|S \cap S_j|] = pt = 0.1 \log n$. Again by Chernoff bound,

$$\Pr[|S \cap S_j| > \hat{t} = \log n] \leq 2^{-4 \log n} < 1/2n.$$

It follows by a union bound that with nonzero probability there is an S which satisfies the two conditions above. \square

Proof of Theorem 1.v. Let $\alpha = 1/34$ and $\beta = 4\alpha$. Also let d, t, \hat{t} be the parameters and S_1, \dots, S_n be the design specified in Lemma 11. Define the function $L: \{0, 1\}^d \rightarrow \{0, 1\}^n$ whose i -th output bit y_i equals

$$\text{mod}_3^2(x_{S_i}) := \left(\sum_{j \in S_i} x_j \right)^2 \bmod 3.$$

Let T_1 be the image set of L . Without the square, this set has mod 3 dimension d and so by Fact 10, $\dim_3(T_1) = O(d^2) = O(n^{16\alpha})$. Let T_2 be an ε -almost k -wise independent set, where $\varepsilon = 1/n^\alpha$ and $k = 2 \log n$. Known constructions [NN93, AGHP92] produce such a set of size $O((k \log n) / \varepsilon)$ and therefore $\dim_3(T_2)$ is at most $O(n^\alpha \log^2 n)$.

Consider the set $T := T_1 +_2 T_2$. By Fact 10, $T +_2 T$ has dimension at most $O(n^{34\alpha} \log^4 n) < n$ because $\alpha < 1/34$. Therefore, there is a non-zero mod 3 linear function ℓ such that $\ell(y) = 0 \pmod{3}$ for any $y \in T$, while $\Pr[\ell(y)] \leq 1/2$ for a uniform y . It remains to show that T is $O(1/n^\alpha)$ -biased. For any test on $I \subseteq [n]$, we consider the cases (1) when $|I| \leq k$, and (2) when $|I| > k$ separately.

Write $y = y_1 \oplus y_2$, where $y_1 \in T_1$ and $y_2 \in T_2$. Case (1) follows from the fact that T_2 is $1/n^\alpha$ -almost k -wise independent. Case (2) follows from the following claim. \square

Claim 12. For any $|I| > k$, $|\mathbb{E}_{y_1}[\chi_I(y_1)]| \leq O(1/n^\alpha)$, where $\chi_I(y) := (-1)^{\sum_{i \in I} y_i}$.

Proof. Pick a subset $J \subseteq I$ of size k . Define $f, p: \{0, 1\}^n \rightarrow \{0, 1\}$ by $f(x) := \sum_{i \in J} \text{mod}_3^2(x_{S_i})$ and $p(x) := \sum_{i \in I \setminus J} \text{mod}_3^2(x_{S_i})$ respectively. Observe that

$$|\mathbb{E}_{x_i: i \in [d]}[\chi_I(y_1)]| = |\mathbb{E}_{x_i: i \in [d]}[(-1)^{(f(x)+p(x))}]|,$$

which is the correlation between f and p .

Consider the sets $S_j \subseteq [d]$ with $j \in J$. Let B_1 be the set of indices appearing in their pairwise intersections. That is, $B_1 := \{k \in [d] : k \in S_i \cap S_j \text{ for some distinct } i, j \in J\}$. Fixing the value of every $x_k \in B_1$, each $\text{mod}_3^2(S_j)$ in f becomes a function on $m := n^\beta - \hat{t} \cdot k \geq 0.9n^\beta$ bits.

Let B_2 be the set of indices in $[d]$ outside the subsets indexed by J , which do not affect the bits in J . Fixing their values, each $\text{mod}_3^2(S_j)$ in p is a function on at most $\hat{t} \cdot k = O(\log^2 n)$ bits and so can be written as a degree $O(\log^2 n)$ GF(2)-polynomial. Since p is a parity of the $\text{mod}_3^2(S_j)$'s, it can also be written as a degree $O(\log^2 n)$ GF(2) polynomial.

To build intuition, note that after fixing the input bits in B_1 and B_2 , for each of the $\text{mod}_3^2(S_j)$ in f , by [Smo87] we have $|\mathbb{E}_{x_i: i \in [d]}[(-1)^{(\text{mod}_3^2(S_j) + p(x))}]| \leq 1 - \Omega(1)$. In the following lemma we prove a variant of Impagliazzo's XOR Lemma [Imp95] to show that

$$|\mathbb{E}_{x_i: i \in [d]}[(-1)^{(f(x)+p(x))}]| \leq O(1/m^{0.249}) = O(1/n^\alpha).$$

Averaging over the values of the x_k 's in B_1 and B_2 finishes the proof. \square

Lemma 13. Let $k = 2 \log m$, define $f: \{0, 1\}^{m \times k} \rightarrow \{0, 1\}$ by $f(x^{(1)}, \dots, x^{(k)}) := \text{mod}_3^2(x^{(1)}) + \dots + \text{mod}_3^2(x^{(k)})$. Let $p: \{0, 1\}^{m \times k} \rightarrow \{0, 1\}$ be any polynomial of degree $O(\log^2 m)$. We have $\text{Cor}(f, p) := \mathbb{E}_{x \sim \{0, 1\}^{m \times k}}[(-1)^{f(x)+p(x)}] \leq O(1/m^{0.249})$.

Proof. We will use the fact that [Smo87] holds for degree $n^{\Omega(1)}$ polynomials to get correlation $1/n^{\Omega(1)}$ for polynomials of much smaller degree ($\text{polylog}(n)$).

As in the proof in [Imp95] we first show the existence of a measure $M: \{0, 1\}^{m \times k} \rightarrow [0, 1]$ of size $|M| := \sum_x M(x) = 2^{mk}/4$ such that with respect to its induced distribution $D(x) := \frac{M(x)}{|M|}$, mod_3^2 is $1/m^{0.249}$ -hard for any polynomial p of degree $O(\log^2 m)$, i.e.,

$$\Pr_{x \sim D}[\text{mod}_3^2(x) = p(x)] \leq 1/2 + 1/m^{0.249}.$$

Suppose not. Lemma 1 in [Imp95] implies that one can obtain a function q by taking the majority of $32m^{0.499}$ polynomials of degree $O(\log^2 m)$ such that

$$\Pr_{x \sim \{0, 1\}^{m \times k}}[\text{mod}_3^2(x) = q(x)] > 3/4.$$

Note that q can be represented as a degree $O(m^{0.499} \log^2 m)$ polynomial. From [Smo87], $\Pr_{x \sim \{0, 1\}^{m \times k}}[\text{mod}_3^2(x) = p(x)] \leq 3/4 + \Theta(\delta)$ for any degree $\delta m^{1/2}$ polynomial p , a contradiction.

Now we show that there is a set $S \subseteq \{0, 1\}^{m \times k}$ of size $2^{mk}/8$ such that mod_3^2 is $1/m^{0.249}$ -hard-core on S for any polynomial p of degree $O(\log^2 m)$, i.e.,

$$\Pr_{x \sim S}[\text{mod}_3^2(x) = p(x)] \leq 1/2 + 1/m^{0.249}.$$

Let p be any degree- $O(\log^2 m)$ polynomial. For any measure $M: \{0, 1\}^{m \times k} \rightarrow [0, 1]$, define $Adv_p(M)$ by $Adv_p(M) := \sum_x M(x)(-1)^{\text{mod}_3^2(x)+p(x)}$. We construct S probabilistically by picking each x to be in S with probability $M(x)$. Let M_S be the indicator function of S . Then $E_S[Adv_p(M_S)] = Adv_p(M) \leq \frac{|M|}{2m^{0.249}}$. Note that $Adv_p(M_S)$ is the sum of 2^{mk} independent random variables, where each variable is over $[-1, 0]$ or $[0, 1]$. By Hoeffding's inequality,

$$\Pr_S[Adv_p(M_S) > |M|/m^{0.249}] \leq 2^{-|M|^2/2^{m \cdot 2m^{0.49}}} = 2^{-2^m/128m^{0.49}}.$$

Note that there are $2^{mkO(\log^2 m)}$ polynomials of degree $\log^2 m$. Moreover, since $E_S[|S|] = 2^{mk}/4$, again by Hoeffding's inequality, $\Pr_S[|S| < 2^{mk}/8] \leq 1/2$. Hence, by a union bound, the required S exists.

It follows that there exists a set of inputs $S \subseteq \{0, 1\}^{m \times k}$ of size $2^{mk}/8$ such that mod_3^2 is $1/m^{0.249}$ -hard-core on S for any polynomial of degree $O(\log^2 m)$. By Lemma 4 in [Imp95] and our choice of k , for any polynomial p of degree $O(\log^2 m)$,

$$\Pr_x[f(x) = p(x)] \leq 1/2 + 1/m^{0.249} + (7/8)^k = 1/2 + O(1/m^{0.249}).$$

Hence f is $O(1/m^{0.249})$ -hard for any polynomial of degree $O(\log^2 m)$, and the lemma follows. \square

Proof of Claim 3. We replace the design in the proof of Theorem 1.v with one that has set size $t = O(\log^4 n)$ and intersection size $\hat{t} = O(\log n)$. Using the same idea as in the proof of Lemma 11 one can show that such design exists provided the universe is of size $d = O(\log^8 n)$. Now, using the same argument, for tests of size larger than $c \log n$, we apply (1) to f and p , which are the parity of $c \log n$ copies of mod 3 parity on $m = O(\log^4 n)$ bits and an $O(\log^2 n)$ degree polynomial respectively. This gives bias $O(1/n^c)$. Note that the image set T_1 now has mod 3 dimension $d^2 = O(\log^{16} n)$.

For tests of size at most $c \log n$, we replace the almost k -wise independent set with the k -wise independent distribution given by (2), which has zero bias, and we denote the support of the distribution by T_2 .

By Fact 10, $T := T_1 +_2 T_2$ has mod 3 dimension $O(n^{0.49} \log^{16} n) < n$. Hence, $T +_2 T$ has dimension less than n and the claim follows. \square

2.7 Sum of k copies of small-bias distributions

We now show that the results hold for k copies when ε is replaced by $\varepsilon^{2/k}$, proving the “Moreover” part in Theorem 1 and 2.

Proof of “Moreover” part in Theorem 1 and 2. To prove Theorem 1.i, 1.ii and 1.iv, we can replace e by $2e/k$ in their proofs to obtain distributions D' that are $\varepsilon^{2/k}$ -biased. Since we have to throw in at least one error, $k \leq 2e$. The rest follows by noting the sum of k copies of D' is identical to $D + D$.

By scaling down the above small-biased distributions D' for Theorem 1.i and 1.ii to n' bits as in the proofs of Theorem 1.iii and 2 respectively, we obtain $\varepsilon^{2/k}$ -biased distributions $D'_{n'}$ so that the sum of k copies of $D'_{n'}$ is identical to $D_{n'} + D_{n'}$ in Theorem 1.iii and 2. Since we have to throw in at least one error, $k \leq e_{n'}$, where $e_{n'}$ is the scaled down $e = e(n)$.

For Theorem 1.v, let $\alpha := \log(1/\varepsilon)/\log n$ and so $\varepsilon^{2/k} = n^{-2\alpha/k}$. We set $\beta = 8\alpha/k$ instead of 4α in the construction of T_1 and replace T_2 by a $1/n^{-2\alpha/k}$ -almost $2\log n$ -wise independent set in the proof, and call them T'_1 and T'_2 respectively. We now have $\dim_3(T'_1) = O(n^{32\alpha/k})$ and $\dim_3(T'_2) = O(n^{2\alpha/k} \log^2 n)$. Thus, the set $T' := T'_1 +_2 T'_2$ has dimension at most $O(n^{34\alpha/k} \log^2 n)$ and therefore the sum of k copies has dimension at most $\dim_3(T')^k = O(n^{34\alpha} \log^{2k} n) < n$, for $k < O(\log n / \log \log n)$. The bias of T' follows from the facts that T'_2 has bias $n^{-2\alpha/k}$ against tests of size at most k , and T_1 has bias $O(n^{-2\alpha/k})$ for tests of size greater than k . \square

3 Mod 3 dimension of k -wise independence

In this section, we begin a systematic investigation on the mod3 dimension of k -wise independent distributions.

Recall Definition 9 of mod3 dimension. We also define the mod3 dimension of a matrix to be the mod3 dimension of its rows. We also write rank_3 for mod3 dimension. This notion is naturally generalized to dimension over \mathbb{F}_p for arbitrary p , written rank_p .

We will sometimes work with vectors over $\{-1, 1\}$ instead of $\{0, 1\}$. Note that the map $(1 - x)/2$ convert the values 1 and -1 to 0 and 1 respectively, and so the mod3 dimension of a set will differ by at most 1 when we switch vector values from $\{-1, 1\}$ to $\{0, 1\}$, and vice versa.

While we state our results for mod3, all the results in this section can be extended to mod p for any odd prime p naturally.

3.1 Lower bound for almost k -wise independence

In the following claim we give a dimension lower bound on almost k -wise independent distributions. Here “almost” is measured with respect to statistical distance (another possible definition is the max bias of any parity).

Claim 14. *Let D be any set in $\{0, 1\}^n$. If $\dim_3(D) = t$, then D is not $1/10$ -almost $ct/\log n$ -wise independent, for a universal constant c .*

This actually rules out even $k - 1$ independence because $x_i \in \{0, 1\}$. Moreover, this gives an exponential separation between seed length and dimension for almost d^\perp -wise independence. Indeed, for $k = O(1)$, the seed length is $\Theta(\log \log n)$, whereas the dimension must be $\Omega(\log n)$.

Proof. Let C be the span of D over \mathbb{F}_3 and C^\perp be its orthogonal complement. C^\perp has dimension $n - t$. We view C^\perp as a linear code over \mathbb{F}_3 and let d^\perp be its minimum distance. Since C^\perp is linear, d^\perp equals the minimum Hamming weight of its non-zero elements. We

have

$$\begin{aligned}
(d^\perp \log_2 n)/4 &\leq \left\lfloor \frac{d^\perp - 1}{2} \right\rfloor \log_2 n - \left\lfloor \frac{d^\perp - 1}{2} \right\rfloor \log_2 \left\lfloor \frac{d^\perp - 1}{2} \right\rfloor + \left\lfloor \frac{d^\perp - 1}{2} \right\rfloor \\
&= \log_2 \left(\frac{n}{\left\lfloor \frac{d^\perp - 1}{2} \right\rfloor} \right)^{\left\lfloor \frac{d^\perp - 1}{2} \right\rfloor} + \left\lfloor \frac{d^\perp - 1}{2} \right\rfloor \\
&\leq \log_2 \left(\left(\frac{n}{\left\lfloor \frac{d^\perp - 1}{2} \right\rfloor} \right) 2^{\left\lfloor \frac{d^\perp - 1}{2} \right\rfloor} \right) \\
&\leq \log_2 \left(\sum_{i=0}^{\left\lfloor \frac{d^\perp - 1}{2} \right\rfloor} \binom{n}{i} 2^i \right) \\
&\leq t \log_2 3,
\end{aligned}$$

where the last inequality follows from the Hamming bound:

$$3^{n-t} \sum_{i=0}^{\left\lfloor \frac{d^\perp - 1}{2} \right\rfloor} \binom{n}{i} 2^i \leq 3^n.$$

Hence $d^\perp \leq O(t/\log n)$. Let y be a codeword in C^\perp with Hamming weight d^\perp . Let $I := \{i \mid y_i \neq 0\}$. Note that for every $x \in D$, we have $\langle y, x \rangle_3 = 0$ on I . On the other hand, for a uniformly distributed x we have $\langle y, x \rangle_3 = 0$ with constant probability. (Here we are using that $d^\perp \geq 2$ w.l.o.g.) Therefore, D is constant bounded away from uniform on the bits indexed in I . \square

3.2 2-wise independence

We now show that the mod 3 dimension of a 2-wise independent set can be as small as $n^{0.72}$. Then we give evidence that our approach cannot do any better.

Definition 15. We say H is a Hadamard matrix of order n if it satisfies $HH^T = nI_n$, where I_n is the $n \times n$ identity matrix.

It is well-known that the rows of a Hadamard matrix form a 2-wise independent set.

In the following we will work with vectors whose entries are from $\{-1, 1\} = \{2, 1\}$. The following two claims show that certain Hadamard matrices cannot have dimension smaller than $n/2$. They are taken from [Wil12]. First we give a lower bound to the mod p rank from the determinant of any square matrix.

Claim 16 (Theorem 1 in [Wil12]). *Let A be an $n \times n$ matrix. Then $\text{rank}_p(A) \geq n - e$, where e is the largest s such that $p^s \mid \det(A)$, i.e., $p^e \mid \det(A)$ and $p^{e+1} \nmid \det(A)$.*

Proof. Suppose $\text{nullity}_p(A) = n - r$. Let $(\beta_1, \dots, \beta_{n-r})$ be a basis of null space of A over \mathbb{F}_p . Extend the basis to $(\beta_1, \dots, \beta_n)$ so that it forms a basis of \mathbb{F}_p^n . Let B be the matrix whose columns are β_i 's. Note that $\det(B) \neq 0 \pmod{p}$ and $\det(AB) = \det(A)\det(B)$. Thus, $p^s \mid \det(A)$ if and only if $p^s \mid \det(AB)$. By construction, $\beta_1, \dots, \beta_{n-r}$ are in the null space of A over \mathbb{F}_p and so the first $n - r$ columns of AB are zero mod p . Hence $p^{n-r} \mid \det(A)$. \square

Claim 17 (Theorem 2 in [Wil12]). *Let H be an $n \times n$ Hadamard matrix. Let p be an odd prime such that $p \mid n$ and $p^2 \nmid n$. Then $\text{rank}_p(H) \geq n/2$.*

Proof. Since H is a Hadamard matrix, we have $HH^T = nI$ and so $\det(H)\det(H^T) = \det(H)^2 = n^n$. Hence $|\det(H)| = n^{n/2}$. By the condition of p , $p^{n/2} \mid n^{n/2}$ and $p^{n/2+1} \nmid n^{n/2}$. Hence, it follows from Claim 16 that $\text{rank}_p(H) \geq n/2$. \square

The following claims characterize Hadamard matrices with mod p rank at most $n/2$.

Claim 18. *Let A be an $n \times m$ matrix such that $AA^T = mI_n$. If $p \mid m$, then $\text{rank}_p(A) \leq m/2$.*

Proof. If $p \mid m$, $AA^T = 0 \pmod{p}$. Suppose $\text{rank}_p(A) = \text{rank}_p(A^T) = k$. We know that the basis of A^T are contained in the null space of A . Hence $\text{nullity}(A) = m - k \geq k$ and therefore $k \leq m/2$. \square

Claim 19. *Let H be an $n \times n$ Hadamard matrix. If $p \mid n$, then $\text{rank}_p(H) \leq n/2$. Otherwise, $\text{rank}_p(H) = n$.*

Proof. The first part follows from the previous claim. For the second part, if $p \nmid n$ then $\det(H)^2 = \det(H)\det(H^T) \neq 0 \pmod{p}$ and so $\det(H) \neq 0 \pmod{p}$. Hence $\text{rank}_p(H) = n$. \square

Now we give a generic transformation that reduces the dimension of Hadamard matrices whose order violates the condition in Claim 17. Note that the affine bijection $L: \{-1, 1\}^n \rightarrow \{0, 1\}^n$ defined by $L(v) = (\mathbf{1} - v)/2$, where $\mathbf{1}$ is the all-ones vector, maps vectors from $\{-1, 1\}^n$ to $\{0, 1\}^n$. We have the following facts.

Fact 20. *Let $S \subseteq \{-1, 1\}^n$ be a set consisting of the all-ones vector. Then $\dim_3(L(S)) \leq \dim_3(S)$.*

Fact 21. *If A and B are two Hadamard matrices over $\{-1, 1\}$, then $A \otimes B$ is also a Hadamard matrix.*

Fact 22. *Let A, B be two matrices over any field. We have $\text{rank}_3(A \otimes B) \leq \text{rank}_3(A) \cdot \text{rank}_3(B)$, where $A \otimes B$ is the tensor product of A and B .*

Proof. Let $\alpha_1, \dots, \alpha_u$ and β_1, \dots, β_v be two bases of A and B respectively. We show that $(\alpha_i \otimes \beta_j)_{i,j}$ is a basis of $A \otimes B$. Indeed, given any vector $a := \sum_{i=1}^u c_i \alpha_i$ and $b := \sum_{j=1}^v d_j \beta_j$,

$$a \otimes b = \left(\sum_{i=1}^u c_i \alpha_i \right) \otimes \left(\sum_{j=1}^v d_j \beta_j \right) = \sum_{i=1}^u \sum_{j=1}^v c_i d_j (\alpha_i \otimes \beta_j). \quad \square$$

Claim 23. *There exists an infinite family of 2-wise independent distributions over $\{0, 1\}^n$ with mod 3 dimension at most $n^{0.72}$.*

Proof. Starting with a Hadamard matrix H_{12} over $\{-1, 1\} = \{2, 1\} \subseteq \mathbb{F}_3$ (its existence is guaranteed by Paley construction [Pal33]), for every n that is a power of 12, we construct the Hadamard matrix $H_n := H_{12}^{\otimes r}$, where $r = \log_{12} n$. It follows from Claim 17 and 19 that $\text{rank}_3(H_{12}) = 6$. Hence, by Fact 22, H_n has dimension $6^{\log_{12} n} = n^{0.72}$. By rows and column operations, we can assume H_{12} contains the all-ones vector. Thus, H_n also contains the all-ones tensor, and the claim follows from Fact 20. \square

The smaller m we start from any $m \times m$ Hadamard matrix with dimension $m/2$, the better exponent we get. Since Hadamard matrices must be of order 1, 2, or multiple of 4, Claim 17 implies that 12 is indeed the smallest possible m .

3.3 k -wise independence with dimension $n - 1$

We restrict our attention on the subspace $M := \{x \in \{-1, 1\}^n \mid \sum_{i=1}^n x_i = 0 \bmod 3\}$ and look for the largest k so that there exists a k -wise distribution supported on M . To this end, we give a sufficient condition for its existence and show that $k \geq 5$. We note that by formulating the condition as a linear program, our empirical results from the LP solver show that for $n \leq 900$, $k \geq n/3$.

Claim 24. *The following statements are equivalent:*

- (1) *There exists a k -wise independent distribution over $\{-1, 1\}^n$ supported on M .*
- (2) *For every polynomial $p: \{-1, 1\}^n \rightarrow \mathbb{R}$ of degree k with no constant term, there exists an $x \in M$ such that $p(x) \leq 0$.*

Proof. For every $S \subseteq [n]$, define $\chi_S: \{-1, 1\}^n \rightarrow \{-1, 1\}$ by $\chi_S(x) := \prod_{i \in S} x_i$. We formulate (1) as the following linear system:

$$\begin{aligned} \sum_{x \in M} \mu_x &= 1, \\ \sum_{x \in M} \mu_x \chi_S(x) &= 0 \quad \forall S : 0 < |S| \leq k, \\ \mu_x &\geq 0 \quad \forall x \in M. \end{aligned}$$

By Farkas' lemma, a solution exists if and only if the following linear system has *no* solution:

$$\begin{aligned} \sum_{S: 0 \leq |S| \leq k} \hat{p}_S \chi_S(x) &\geq 0 \quad \forall x \in M, \\ \hat{p}_\emptyset &< 0. \end{aligned}$$

This is equivalent to

$$\sum_{S: 1 \leq |S| \leq k} \hat{p}_S \chi_S(x) > 0 \quad \forall x \in M. \quad \square$$

The next step is symmetrization.

Definition 25. *Given a degree- k polynomial $p(x) : \{-1, 1\}^n \rightarrow \mathbb{R}$, $p(x) := \sum_{S: 1 \leq |S| \leq k} \hat{p}_S \chi_S(x)$, let $\tilde{p}: \{0, 1, \dots, n\} \rightarrow \mathbb{R}$ be its symmetrization, defined by*

$$\tilde{p}(t) := \sum_{x: |x|=t} p(x) = \sum_{S: 1 \leq |S| \leq k} \hat{p}_S \sum_{x: |x|=t} \chi_S(x) = \sum_{i=1}^k c_i m_i(t)$$

for some $c_1, \dots, c_k \in \mathbb{R}$, where

$$m_i(t) := \sum_{j=0}^i \binom{t}{j} \binom{n-t}{i-j} (-1)^j.$$

Define Z to be the set

$$Z := \{t \in \{0, 1, 2, \dots, n\} \mid t \bmod 3 = 0\}.$$

Because symmetrization does not change Hamming weight, if $p(x) > 0$ for every $x \in M$, then $\tilde{p}(t) > 0$ for every $t \in Z$. Thus, we have the following claim:

Claim 26. *If for every $c_1, \dots, c_k \in \mathbb{R}$, there exists a $t \in Z$ such that $\tilde{p}(t) := \sum_{i=1}^k c_i m_i(t) \leq 0$, then there exists a k -wise independent distribution supported on M .*

Claim 26 suggests a method to exhibit k -wise independent distributions that do not fool mod3.

Claim 27. *For sufficiently large n divisible by 6, there exists a 5-wise independent distribution supported on M .*

Proof. Let $q(t) := \hat{p}(t) + \hat{p}(n-t) = 2(c_2 m_2(t) + c_4 m_4(t))$. It suffices to show that for every c_2 and c_4 , $q(t) \leq 0$ for some $t \in Z$. There are three cases:

If $c_4 \leq 0$ and $c_2 > 0$, we have $q(n/2) = -2c_2n + 2c_4n(n-2)/4 \leq 0$.

If $c_4 \leq 0$ and $c_2 \leq 0$, we have $q(0) = 2(c_2 \binom{n}{2} + c_4 \binom{n}{4}) \leq 0$.

If $c_4 > 0$, then $c_4 m_4(t)$ is negative between the two zeros at $n/2 - c^+(n)/2$ and $n/2 - c^-(n)/2$, where for sufficiently large n ,

$$\begin{aligned} c^+(n) &:= \sqrt{(3n-4) + \sqrt{2}\sqrt{3n^2 - 9n + 8}} \geq 2.33\sqrt{n}, \text{ and} \\ c^-(n) &:= \sqrt{(3n-4) - \sqrt{2}\sqrt{3n^2 - 9n + 8}} \leq 0.75\sqrt{n}. \end{aligned}$$

Moreover, between these two points, $m_2(t)$ has a zero at its non-critical point $n/2 - \sqrt{n}/2$, and so there is a $t_0 \in Z$ near $n/2 - \sqrt{n}/2$ such that $c_2 m_2(t_0) \leq 0$. Therefore, $q(t_0) \leq 0$ and the rest follows from the previous claim. \square

4 Complexity of decoding

In [Vio06, Chapter 6] and [SV10] it is shown that list-decoding binary codes from error rate $1/2 - \varepsilon$ requires computing the majority function on $1/\varepsilon$ bits, which implies lower bounds for list decoding over several computational models.

Using a similar approach, we give lower bounds on the decoding complexity for AC^0 circuits and read-once branching programs. We give a reduction from ε -approximating the majority function to decoding $(1/2 - \varepsilon)d$ errors of a code, where d is the minimum distance.

Define $\varepsilon\text{-MAJ}$ to be the promise problem on $\{0, 1\}^n$, where the YES and NO instances are strings of Hamming weight at least $(1/2 + \varepsilon)n$ and at most $(1/2 - \varepsilon)n$, respectively. We

say that a probabilistic circuit solves ε -MAJ if it accepts every YES instance with probability $2/3$ and accepts every NO instance with probability at most $1/3$.

Let $C \subseteq \{0, 1\}^n$ be a code with minimum distance d and let m_x, m_y be two messages whose codewords x and y differ by exactly d positions, respectively. Define ε -DECODE to be the promise problem on $\{0, 1\}^n$, where the YES and NO instances are strings that differ from x and y at most $(1/2 - \varepsilon)d$, respectively.

Lemma 28. *If a function $D: \{0, 1\}^n \rightarrow \{0, 1\}$ solves ε -DECODE, then a restriction of D solves ε -MAJ on d bits.*

Proof. Let $x, y \in C$ be the codewords of m_x and m_y respectively. Without loss of generality, we assume x and y differ in the first d positions. We further assume $x_i = 0$ and $y_i = 1$ for $i \in [d]$. Given an ε -MAJ instance w of length d , let z be the n -bit string where $z_i = w_i$ for $i \in [d]$ and $z_i = x_i (= y_i)$ otherwise. If w has weight at most $(1/2 - \varepsilon)d$, then w and x disagree in at most $(1/2 - \varepsilon)d$ positions and therefore D accepts. Similarly, if w has weight at least $(1/2 + \varepsilon)d$ then D rejects. \square

Shaltiel and Viola [SV10] show that depth- c $\text{AC}^0[\oplus]$ circuits can solve ε -MAJ only if ε is at least $1/O(\log n)^{(c+2)}$. Brody and Verbin [BV10b] show that ε -MAJ can be solved by a read-once width- w branching program whenever ε is at least $1/(\log n)^{\Theta(w)}$. Combining these results with Lemma 28, we have the following claim.

Claim 29. *Let $D: \{0, 1\}^n \rightarrow \{0, 1\}$ be a function.*

1. *If D is computable by an $\text{AC}^0[\oplus]$ circuit of depth c , then it can only solve ε -DECODE with $\varepsilon \geq 1/O(\log n)^{c+2}$.*
2. *If D is computable by a read-once width- w branching program, then it can only solve ε -DECODE with $\varepsilon \geq 1/(\log n)^{\Theta(w)}$.*

We also note the following negative result for decoding by low-degree polynomials.

Claim 30. *Let $C \subseteq \{0, 1\}^n$ be an $[n, k, d]$ code with dual minimum distance d^\perp , respectively. If*

$$2^{-t} - 2^{k-n} \sum_{i=0}^{te} \binom{n}{i} > 16 \left(1 - \frac{d^\perp}{n}\right)^{e/2^{t-1}}$$

for some constant t and $e \leq \lfloor \frac{d-1}{2} \rfloor$, then any degree- t $GF(2)$ -polynomial cannot detect codewords of C within te errors.

Proof. Suppose on the contrary a polynomial P can detect codewords of C within te errors. By Fact 5 and Schwarz-Zippel lemma, there exists an $\varepsilon := (1 - \frac{d^\perp}{n})^e$ -biased distribution D such that P distinguishes the sum of t independent copies of D from uniform with probability at least $2^{-t} - 2^{k-n} \sum_{i=0}^{te} \binom{n}{i}$. But by [Vio09b], the sum of t copies of D fools P with probability $16\varepsilon^{1/2^{t-1}}$, a contradiction. \square

Acknowledgments. We are grateful to Xue Chen and David Zuckerman for telling us an alternative proof of Theorem 1.iv. We also thank Xue Chen for pointing out that, in a preliminary version of this paper, the proof of Theorem 1.v was written with the wrong design parameters.

References

- [ABN⁺92] Noga Alon, Jehoshua Bruck, Joseph Naor, Moni Naor, and Ron M. Roth. Construction of asymptotically good low-rate error-correcting codes through pseudo-random graphs. *IEEE Transactions on Information Theory*, 38(2):509–516, 1992.
- [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.
- [Baz09] Louay M. J. Bazzi. Polylogarithmic independence can fool DNF formulas. *SIAM J. Comput.*, 38(6):2220–2272, 2009.
- [BDVY13] Andrej Bogdanov, Zeev Dvir, Elad Verbin, and Amir Yehudayoff. Pseudorandomness for width-2 branching programs. *Theory Comput.*, 9:283–292, 2013.
- [Ber84] Stuart J. Berkowitz. On computing the determinant in small parallel time using a small number of processors. *Inform. Process. Lett.*, 18(3):147–150, 1984.
- [Bra09] Mark Braverman. Poly-logarithmic independence fools AC^0 circuits. In *24th IEEE Conf. on Computational Complexity (CCC)*. IEEE, 2009.
- [BT13] Avraham Ben-Aroya and Amnon Ta-Shma. Constructing small-bias sets from algebraic-geometric codes. *Theory of Computing*, 9:253–272, 2013.
- [BV10a] Andrej Bogdanov and Emanuele Viola. Pseudorandom bits for polynomials. *SIAM J. on Computing*, 39(6):2464–2486, 2010.
- [BV10b] Joshua Brody and Elad Verbin. The coin problem, and pseudorandomness for branching programs. In *51th IEEE Symp. on Foundations of Computer Science (FOCS)*, 2010.
- [CRS00] Suresh Chari, Pankaj Rohatgi, and Aravind Srinivasan. Improved algorithms via approximations of probability distributions. *J. Comput. System Sci.*, 61(1):81–107, 2000.
- [DETT10] Anindya De, Omid Etesami, Luca Trevisan, and Madhur Tulsiani. Improved pseudorandom generators for depth 2 circuits. In *Approximation, randomization, and combinatorial optimization*, volume 6302 of *Lecture Notes in Comput. Sci.*, pages 504–517. Springer, Berlin, 2010.
- [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.
- [GR05] Venkatesan Guruswami and Atri Rudra. Tolerant locally testable codes. In *Approximation, randomization and combinatorial optimization*, volume 3624 of *Lecture Notes in Comput. Sci.*, pages 306–317. Springer, Berlin, 2005.

- [GZ61] Daniel Gorenstein and Neal Zierler. A class of error-correcting codes in p^m symbols. *J. Soc. Indust. Appl. Math.*, 9:207–214, 1961.
- [Imp95] Russell Impagliazzo. Hard-core distributions for somewhat hard problems. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 538–545, 1995.
- [KS09] Swastik Kopparty and Shubhangi Saraf. Tolerant linearity testing and locally testable codes. In *Approximation, randomization, and combinatorial optimization*, volume 5687 of *Lecture Notes in Comput. Sci.*, pages 601–614. Springer, Berlin, 2009.
- [Lov09] Shachar Lovett. Unconditional pseudorandom generators for low degree polynomials. *Theory of Computing*, 5(1):69–82, 2009.
- [MZ09] Raghu Meka and David Zuckerman. Small-bias spaces for group products. In *Approximation, randomization, and combinatorial optimization*, volume 5687 of *Lecture Notes in Comput. Sci.*, pages 658–672. Springer, Berlin, 2009.
- [Nis91] Noam Nisan. Pseudorandom bits for constant depth circuits. *Combinatorica. An Journal on Combinatorics and the Theory of Computing*, 11(1):63–70, 1991.
- [NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: efficient constructions and applications. *SIAM J. on Computing*, 22(4):838–856, 1993.
- [NW94] Noam Nisan and Avi Wigderson. Hardness vs randomness. *J. of Computer and System Sciences*, 49(2):149–167, 1994.
- [Pal33] Raymond EAC Paley. On orthogonal matrices. *J. Math. Phys.*, pages 311–320, 1933.
- [Pet60] William W. Peterson. Encoding and error-correction procedures for the Bose-Chaudhuri codes. *Trans. IRE*, IT-6:459–470, 1960.
- [Raz87] Alexander Razborov. Lower bounds on the dimension of schemes of bounded depth in a complete basis containing the logical addition function. *Akademiya Nauk SSSR. Matematicheskie Zametki*, 41(4):598–607, 1987. English translation in Mathematical Notes of the Academy of Sci. of the USSR, 41(4):333–338, 1987.
- [Raz09] Alexander A. Razborov. A simple proof of Bazzi’s theorem. *ACM Transactions on Computation Theory (TOCT)*, 1(1), 2009.
- [RU10] Atri Rudra and Steve Uurtamo. Data stream algorithms for codeword testing. In *Automata, Languages and Programming*, pages 629–640. Springer, 2010.
- [Shp09] Amir Shpilka. Constructions of low-degree and error-correcting epsilon-biased generators. *Computational Complexity*, 18(4):495–525, 2009.
- [Smo87] Roman Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *19th ACM Symp. on the Theory of Computing (STOC)*, pages 77–82. ACM, 1987.
- [SV10] Ronen Shaltiel and Emanuele Viola. Hardness amplification proofs require majority. *SIAM J. on Computing*, 39(7):3122–3154, 2010.
- [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/.

- [Val77] Leslie G. Valiant. Graph-theoretic arguments in low-level complexity. In *6th Symposium on Mathematical Foundations of Computer Science*, volume 53 of *Lecture Notes in Computer Science*, pages 162–176. Springer, 1977.
- [Val83] L. G. Valiant. Exponential lower bounds for restricted monotone circuits. In *15th ACM Symp. on the Theory of Computing (STOC)*, pages 110–117. ACM, 1983.
- [Vio06] Emanuele Viola. *The Complexity of Hardness Amplification and Derandomization*. PhD thesis, Harvard University, 2006.
- [Vio09a] Emanuele Viola. On the power of small-depth computation. *Foundations and Trends in Theoretical Computer Science*, 5(1):1–72, 2009.
- [Vio09b] Emanuele Viola. The sum of d small-bias generators fools polynomials of degree d . *Computational Complexity*, 18(2):209–217, 2009.
- [Wil12] Richard Wilson. Combinatorial analysis lecture notes. 2012.

A Fooling read-once DNF formulas

The following claim shows that $m^{-O(\log(1/\delta))}$ -bias suffices to δ -fool any read-once DNF formulas with m terms. This directly follows from Lemma 5.2 in [CRS00].

Claim 31. *Let ϕ be a read-once DNF formula with m terms. For $1 \leq k \leq m$, every ε -biased distribution D fools ϕ with error $O(2^{-\Omega(k)} + \varepsilon m^k)$.*

Proof. Write $\phi(x) := \bigvee_{i=1}^m C_i$. By Lemma 5.2 in [CRS00], $|\Pr_{x \sim D}[\phi(x)] - \Pr_{x \sim \{0,1\}^n}[\phi(x)]|$ is upper bounded by

$$2^{-k} + e \cdot e^{-k/2e} + \sum_{\ell=1}^k \sum_{S \subseteq [m]: |S|=\ell} \left| \Pr_{x \sim D} \left[\bigwedge_{i \in S} C_i \right] - \Pr_{x \sim \{0,1\}^n} \left[\bigwedge_{i \in S} C_i \right] \right|.$$

The rest follows from the fact that D fools each $\bigwedge_{i \in S} C_i$ with error ε because it is an AND of AND terms. \square