THEORY OF COMPUTING, Volume 13 (16), 2017, pp. 1–23
www.theoryofcomputing.org

Some Limitations of the Sum of Small-Bias Distributions

Chin Ho Lee* Emanuele Viola†

Received August 3, 2015; Revised September 22, 2016; Published December 14, 2017

Abstract: We present two approaches to constructing ε-biased distributions D on n bits and functions $f\colon \{0,1\}^n \to \{0,1\}$ such that the XOR of two independent copies $(D+D)$ does not fool f. Using them, we give constructions for any of the following choices:

1. $\varepsilon = 2^{-\Omega(n)}$ and f is in $\mathsf{P/poly}$;
2. $\varepsilon = 2^{-\Omega(n/\log n)}$ and f is in NC^2;
3. $\varepsilon = n^{-c}$ and f is a one-way space $O(c \log n)$ algorithm, for any c;
4. $\varepsilon = n^{-\Omega(1)}$ 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. We also give conditional results for AC^0 and DNF formulas.

Meka and Zuckerman (RANDOM 2009) prove 4 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 3.

ACM Classification: F.1.3, G.3, F.2.3

AMS Classification: 68Q17

Key words and phrases: complexity theory, pseudorandomness, RL vs. L, error-correcting codes, k-wise independence, small-bias distributions, sum of small bias

---

A preliminary version of this paper appeared in ECCC [20].

*Supported by NSF grant CCF-1319206. Work done in part at Harvard University.

†Supported by NSF grant CCF-1319206. Work done in part while a visiting scholar at Harvard University, with support from Salil Vadhan's Simons Investigator grant.

 DOI: 10.4086/toc.2017.v013a016

1 Introduction and our results

We consider probability distributions over $\{0,1\}^n$, i. e., functions $D : \{0,1\}^n \to [0,1]$ such that

$$\sum_{x \in \{0,1\}^n} D(x) = 1 .$$

We shall refer to such probability distributions simply as "distributions."

Small-bias distributions, introduced by Naor and Naor [24], cf. [1, 2, 5], are distributions that look balanced to parity functions over $\{0,1\}^n$.

Definition 1.1. A distribution D over $\{0,1\}^n$ is ε-biased if for every nonempty subset $I \subseteq [n]$, we have

$$\left| \mathrm{E}_{x\sim D}\left[(-1)^{\sum_{i\in I} x_i}\right] \right| \le \varepsilon .$$

An ε-biased distribution can be generated using a seed of $O(\log(n/\varepsilon))$ bits. Since their introduction, small-bias distributions 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., bitwise XOR) of several independent copies of small-bias distributions. The first paper to explicitly consider it is [8]. 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 1.2 (RL)**.** Reingold and Vadhan (personal communication) asked whether there exists a constant c such that the sum of two independent copies of any n^{-c}-biased distribution fools one-way logarithmic space, a. k. a. one-way polynomial-size branching programs, which would imply $\mathsf{RL} = \mathsf{L}$. It is known that a small-bias distribution fools one-way width-2 branching programs (Saks and Zuckerman, see also [7] where a generalization is obtained). No such result is known for width-3 programs.

Question 1.3 (Polynomials)**.** The papers [8, 21, 39] show that the sum of d small-bias generators fools $\mathbb{F}_2$-polynomials of degree d. However, the proofs only apply when $d \le (1-\Omega(1))\log n$. It is an open question whether the construction works for larger d. If the construction worked for any $d = \log^{O(1)} n$, it would make progress on long-standing open problems in circuit complexity regarding AC^0 with parity gates [28]. This question is implicit in the works [8, 21, 39] and explicit in [38, Chapter 1] (Open question 4), cf. the survey [38, Chapter 1].

In terms of negative results, Meka and Zuckerman [22] show that the sum of 2 distributions with constant bias does not fool mod 3 linear functions. Bogdanov, Dvir, Verbin, and Yehudayoff [7] show that for $\varepsilon = 2^{-O(\sqrt{n}/k)}$, the sum of k copies of ε-biased distributions does not fool circuits of size $\mathrm{poly}(n)$ and depth $O(\log^2 n)$ (NC^2).

This paper gives two different approaches to improving on both these results and obtain other limitations of the sum of small-bias distributions. One is based on the complexity of decoding, and the other one on bounding the mod 3 rank (see Definition 1.8). Either approach is a candidate to answer negatively the "RL question" (Question 1.2).

1.1 Our results

The following theorem states our main counterexamples. We denote by $D+D$ the bitwise XOR of two independent copies of a distribution D.

Theorem 1.4. *For any c, there exists an explicit ε-biased distribution D over $\{0,1\}^n$ and an explicit function f, such that $f(D+D)=0$ and $\Pr_{x\sim\{0,1\}^n}[f(x)=0]\le p$, where ε, f, p are of any one of the following choices:*

i. $\varepsilon = 2^{-\Omega(n)}$, *$f$ is a uniform* poly(n)*-size circuit, and* $p = 2^{-\Omega(n)}$;

ii. $\varepsilon = 2^{-\Omega(n/\log n)}$, *$f$ is a uniform fan-in 2,* poly(n)*-size circuit of depth* $O(\log^2 n)$, *and* $p = 2^{-n/4}$;

iii. $\varepsilon = 1/n^c$, *f is a one-way* $O(c\log n)$*-space algorithm, and* $p = O(1/n^c)$;

iv. $\varepsilon = n^{-\Omega(1)}$, *$f$ is a* mod 3 *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 bitwise 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. $2c$; iv. $O(\log n/\log\log n)$.

Theorem 1.4.i is tight up to the constant in the exponent because every $\varepsilon 2^{-n}$-biased distribution is ε-close to uniform.

Theorem 1.4.ii would also be true with $\varepsilon = 2^{-\Omega(n)}$, if a decoder for certain algebraic-geometric codes ran in NC^2, which we conjecture it does. [7] prove Theorem 1.4.ii with $\varepsilon = 2^{-O(\sqrt{n}/k)}$.

Theorem 1.4.iii can also be obtained in the following way, pointed out to us by Chen and Zuckerman (personal communication). 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 $\varepsilon = n^{-c}$. Actually, both their proof and ours (presented later) apply to $c > 0.01$; but for smaller c Theorem 1.4.iv kicks in.

Theorems 1.4.iii and 1.4.iv come close to addressing the "RL question," without answering it: Theorem 1.4.iv shows that polynomial bias is necessary even for width-3 regular branching programs, while Theorem 1.4.iii shows that the bias is at least polynomial in the width. [22] prove Theorem 1.4.iv with $\varepsilon = \Omega(1)$.

We have not been able to say anything on the "Polynomials question" (Question 1.3).

There exist other models of interest. For read-once DNFs no counterexample with large error is possible because Chari, Rohatgi, and Srinivasan [12], building on [15], show that (just one) $n^{-O(\log(1/\delta))}$-bias distribution fools any read-once DNF on n variables with error δ, cf. Section 5. The [12] result is rediscovered by De, Etesami, Trevisan, and Tulsiani [14], 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 provided each copy has bias poly(δ/n). 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 DNFs. However, we show that the corresponding result for general DNFs would resolve long-standing problems on circuit lower bounds [35]. This can be interpreted as saying that such a result for DNFs is either false or extremely hard to prove. We also get conditional counterexamples for depth-3 and AC^0 circuits.

Theorem 1.5. *Suppose polynomial time* (P) *has fan-in* 2 *circuits of linear size and logarithmic depth. Then Theorem 1.4 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 of size* $\mathrm{poly}(n)$*, 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 bitwise XOR of* $k \le \log n$ *independent copies of* D *if* ε *is replaced by* $\varepsilon^{2/k}$.

Recall that it is still open whether NP has linear-size circuits of logarithmic depth.

Theorem 1.6. *Suppose for every* $\delta > 0$ *there exists a constant* d *such that* NC^2 *has* AC^0 *circuits of size* 2^{n^δ} *and depth* d*. Then Theorem 1.4 also applies to the following choice of parameters:* $\varepsilon = n^{-\log^c n}$*,* f *is an* AC^0 *circuit of size* $n^{O(c)}$ *and depth* $O(c)$ *, and* $p = n^{-\log^{\Omega(1)} n/4}$.

Moreover, our result extends to more copies of D *as follows. The input* $D+D$ *to* f *can be replaced by the bitwise XOR of* $k \le \log^{c+1} n/6(c+1)\log\log n$ *independent copies of* D *if* ε *is replaced by* $\varepsilon^{2/k}$.

Recall that the assumption in Theorem 1.6 holds for NC^1 instead of NC^2, in fact it holds even for log-space. Moreover, the parameters in the conclusion of Theorem 1.6 are 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 [4, 29, 10, 34].

All the above results except Theorem 1.4.iv are based on a new, simple connection between small-bias generators and error-correcting codes, discussed in Section 1.2.

Definition 1.7. A distribution over X^n is *k-wise independent* if its marginal distribution on any k positions is uniformly distributed.

Definition 1.8. Let $S \subseteq \mathbb{F}_p$ be a set of vectors. Define the mod p rank of S, denoted by $\mathrm{rank}_p(S)$ to be the rank of S over $\mathbb{F}_p$. We define the mod p rank of a distribution D to be the mod p rank of its support.

Definition 1.9. The *correlation* of two functions $f, g\colon \{0,1\}^n \to \{0,1\}$ is

$$\mathrm{Cor}(f,g) := \mathrm{E}_{x\sim\{0,1\}^n}[(-1)^{f(x)+g(x)}] .$$

Theorem 1.4.iv instead follows [22] and bounds the mod 3 rank of small-bias distributions. It turns out an upper bound on the mod 3 rank of some k-wise independent distributions over bits would allow us to reduce the bias in Theorem 1.4.iv, assuming long-standing conjectures on correlation bounds for low-degree polynomials (which may be taken as standard).

Claim 1.10. *Suppose*

1. *the parity of* k *copies of* mod 3 *on disjoint inputs of length* m *has correlation* $2^{-\Omega(k)}$ *with any* $\mathbb{F}_2$*-polynomial of degree* $\varepsilon\sqrt{m}$ *for some constant* $\varepsilon > 0$*, and*
2. *for every* c*, there exists an* n^{-c}*-almost* $c\log n$*-wise independent distribution whose support on* $\{0,1\}^n \subseteq \mathbb{F}_3^n = \{0,1,2\}^n$ *has* mod 3 *rank at most* $n^{0.49}$.

Then the "RL *question" has a negative answer, i. e., for every* c*, there exists an* n^{-c}*-biased distribution* D *such that* $D+D$ *does not fool a one-way* $O(\log n)$*-space algorithm. More specifically,* $D+D$ *does not fool a* mod 3 *linear function.*

Contrapositively, an affirmative answer to the "RL question," even for permutation, width-3 branching programs, implies lower bounds on the mod 3 rank of k-wise independent distributions, or that the aforementioned correlation bounds are false.

What we know about the second assumption in Claim 1.10 is in Section 3, where we initiate a systematic study of the mod 3 rank of (almost) k-wise independent distributions, and obtain the following lower and upper bounds. First, we give an $\Omega(k \log n)$ lower bound on the mod 3 rank 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 3.1). This also gives an exponential separation between mod 3 rank and seed length for such distributions.

We then prove the following upper bounds, see Claim 3.9.

Theorem 1.11. *For infinitely many n, there exist k-wise independent distributions over $\{0,1\}^n$ with* mod 3 *rank d for $k = 2$ and $d \le n^{0.73}$.*

We note that an upper bound of $n-1$ on the mod 3 rank of a k-wise independent distribution implies that the distribution is constant on a mod 3 linear test. We ask what is the largest $k^* = k^*(n)$ such that there exists a k-wise independent distribution with mod 3 rank $\le n-1$. We conjectured the bound $k^*(n) = \Omega(n)$. Partial progress towards this conjecture appeared in a preliminary version of this paper [20]. This conjecture was later verified [9].

1.2 Our techniques

All our counterexamples in Theorem 1.4 and 1.5, except Theorem 1.4.iv, 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 finite field of characteristic 2. (Using characteristic 2 allows us to work with small-bias over bits, as opposed to large alphabets, which makes things slightly simpler.) 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 $\mathbb{F}$. Now, define D_e to be the distribution on $n \log|\mathbb{F}|$ bits obtained from adding N_e to C, and we have the following fact.

Fact 1.12. *D_e is $(1 - d^\perp/n)^e$-biased.*

Proof. If a test is on less than $d^\perp$ field elements, D_e has zero bias because it is $(d^\perp - 1)$-wise independent. Otherwise, the bias is nonzero only if none of the symbols touched by the test are hit by random noise, which happens with probability $(1 - d^\perp/n)^e$. □

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,

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

Definition 1.13. An algorithm is a *threshold-e discriminator* for the code C if it decides whether a string is within Hamming distance e of the code.

Now suppose an algorithm is a threshold-$2e$ discriminator for C. Then it can be used to distinguish $D_e + D_e$ from uniform. More generally, if an algorithm is a threshold-ke discriminator for C, then it can distinguish the XOR of k independent copies of D_e from uniform. Contrapositively, if $D_e + D_e$ fools f, then f is not a threshold-$2e$ discriminator for C. Thus, to obtain counterexamples we only have to exhibit an appropriate threshold discriminator. We achieve this by drawing from results in coding theory. This is explained below after two remarks.

Remark 1.14. Our threshold discriminator is only required to tell apart noisy codewords and uniform random strings. This is a weaker condition than decoding. In fact, similar threshold discriminators have been considered in the context of tolerant property testing [17, 19, 30], 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.

Remark 1.15. We note that our connection between ε-bias distributions and linear codes is different from the well-known connection in [24], 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.

As hinted before Fact 1.12, the small-bias property is established through a case analysis based on the weight of the test. This paradigm goes back at least to the original work by Naor and Naor [24]. It was used again more recently in [23, 3]. Our reasoning is especially close to [23, 3] because in both papers small tests are handled by local independence but large tests by sum of independent biased bits.

For general circuits (Theorem 1.4.i), we consider the asymptotically good binary linear code with constant dual relative distance, based on algebraic geometry and exhibited by Guruswami in [32]. We conjecture that the corresponding threshold discriminator can be implemented in NC^2. However, we are unable to verify this. Instead, for NC^2 circuits (Theorem 1.4.ii), we use Reed–Solomon codes and the Peterson–Gorenstein–Zierler syndrome-decoding algorithm [27, 16] which we note is in NC^2. Under the assumption that NC^2 is contained in AC^0 circuits of size 2^{n^δ}, 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 1.6). This result could also be obtained by scaling down a result in [7].

Our counterexample for one-way log-space computation (Theorem 1.4.iii) also uses Reed–Solomon codes. The threshold discriminator is simply syndrome decoding: To decode 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 $q = n$, message length $k = d^\perp - 1 = n - O(c)$, 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 2.2 however we show that to obtain n^{-c} bias, the space $O(e \log q)$ 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. [30], 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 1.5) follows from scaling down to barely superlogarithmic input length, and a depth reduction [35] (cf. [38, Chapter 3]) of the counterexample for general circuits (Theorem 1.4.i). We note that the $2^{-\Omega(n)}$-bias in Theorem 1.4.i is essential for this result, in the sense that $2^{-n/\log n}$-bias would be insufficient to obtain Theorem 1.5. We also remark that since $O(\log^2 n)$-wise independence suffices to fool DNF formulas [4], one must consider linear codes with dual distance less than $\log^2 n$ in our construction, and so D_e has bias at least $(1-\log^2 n/n)^e = 2^{-O(\log^2 n)}$. On the other hand, [14] shows that $2^{-O(\log^2 n \log\log n)}$-bias fools DNF formulas.

The connection between codes and small-bias distributions motivate us to study further the complexity of decoding. [37, Chapter 6] and [31], cf. [37, Chapter 6], show that list-decoding requires computing the majority function. In Claim 4.2 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 [10, 34], 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 1.6 from Theorem 1.4.ii.

A different approach. We now explain the high-level ideas in proving Theorem 1.4.iv. Meka and Zuckerman [22] construct the following constant-bias distribution D over $n := \binom{d}{5}$ bits with mod 3 rank less than $\sqrt{n}$. Each output bit is the square of the mod 3 sum of 5 out of the d uniform random bits, which can be written as a degree-5 polynomial over $\mathbb{F}_2$. Since any parity of the output bits is also a degree-5 polynomial over $\{0,1\}^d$, D has constant bias. To show that a mod 3 linear function is always 0 on the support of $D+D$, they observe that for sufficiently large n, D has mod 3 rank at most $d^2 < \sqrt{n}$, and $D+D$ has mod 3 rank at most $(d^2)^2 + d^2 = d^4 + d^2 < n$. (See Fact 2.4.)

We extend their construction using ideas from the Nisan generator [25]: We pick a pseudo-design consisting of n sets where each set has size n^β (we will choose β to be a small constant), and the intersection of any two sets has size $O(\log n)$. Such pseudo-design exists provided the universe has size $n^{2\beta}$. The output distribution is again the square of the mod 3 sum on each set.

For any test of size at least $C\log n$ bits, let J be any $C\log n$ 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 mod 3 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, it corresponds to a set which has intersection at most $\log n$ with each of the sets that correspond to the bits in J. Thus, b is now a mod 3 function on at most $|J|\log n = \log^2 n$ input bits and thus can be written as a degree-$\log^2 n$ polynomial over $\mathbb{F}_2$. Hence, the parity of the bits outside J is also an $\mathbb{F}_2$-polynomial of the same degree, and we call this polynomial p.

Now 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 mod 3 function on n^β bits, by Smolensky's theorem [33], it has constant correlation with p. In Lemma 2.9 we prove a variant of Impagliazzo's XOR lemma [18] 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 $C\log n$. For smaller tests, the above distribution could have constant bias, and hence we XOR it with an $1/n^{\Omega(1)}$-almost $C\log n$-wise independent distribution, which gives us ε bias for tests of size less than $C\log n$ and has sufficiently small rank. We then show that the XOR of the two distributions has rank less than $\sqrt{n}$ and conclude as in the previous paragraph.

We refer the reader to [40] for background on XOR lemmas.

Organization. In Section 2 we describe our counterexamples and prove Theorem 1.4 and 1.5, and Claim 1.10. In Section 3 we prove our lower bounds and upper bounds on the mod 3 rank of k-wise independence. As a bonus, in Section 4 we include some results on the complexity of decoding. For example, we show that for codes with large minimum distance, AC^0 circuits and read-once branching programs cannot decode when the number of errors is close to half of the minimum distance of a code (Claim 4.2). We obtained these results while attempting to devise low-complexity algorithms that *can* decode (which, by our connection, would have consequences for the sum of small-bias generators).

2 Our counterexamples

We are now ready to prove Theorem 1.4 and 1.5, and Claim 1.10. We consider linear codes with different parameters, the bias of D follows from Fact 1.12. 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 [32] exhibits the following family of constant-rate binary linear codes whose primal and dual relative minimum distance are both constant.

Theorem 2.1 (Theorem 4 in [32])**.** *For infinitely many n, there exists a binary linear code C with block length n and dimension $n/2$, which can be constructed, encoded, and decoded from $n/60$ errors in time* $\mathrm{poly}(n)$*. Moreover, the dual of C has minimum distance at least $n/30$.*

Proof of Theorem 1.4.i. Applying Fact 1.12 with $e = n/120$ to the code in Theorem 2.1, we obtain a distribution D that is $2^{-n/3600}$-biased. Our threshold-$2e$ discriminator f for the code C decodes and encodes the input, and accepts if and only if the input and the re-encoded string differ by at most $2e$ positions. Since both the encoding and decoding algorithms run in polynomial time, so does f.

Note that f accepts at most

$$2^{n/2} \cdot \sum_{i=0}^{2e} \binom{n}{i} \leq 2^{n/2} \cdot 2^{nH(1/60)} \leq 2^{0.75n}$$

possible strings, where $H(\cdot)$ is the binary entropy function (cf. [13, Example 11.1.3] for the first inequality). Hence, f distinguishes $D+D$ from the uniform distribution with probability at least $1-2^{-0.25n}$. $\square$

2.2 NC^2 circuits

Proof of Theorem 1.4.ii. Let q be a power of 2. Consider the Reed–Solomon code C over $\mathbb{F}_q$ with block length $q-1$, dimension $q/2$ and minimum distance $q/2$. C has dual minimum distance $q/2+1$ and can decode from $q/4$ errors. Applying Fact 1.12 to C with $e = q/12$, we obtain a distribution D over $n := (q-1)\log q$ bits that is $2^{-\Omega(n/\log n)}$-biased.

Let α be a primitive element of $\mathbb{F}_q$. Let H be a parity-check matrix for C. We first recall the Peterson–Gorenstein–Zierler syndrome-decoding algorithm [27, 16].

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

$$\Lambda_v(x) := \prod_{i \in E}(1-\alpha^i x) = 1 + \sum_{i=1}^{v} \lambda_i x^i$$

be the error locator polynomial. The syndromes and the coefficients of Λ_v are linearly related by

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

for $j > v$. This forms a linear system with unknowns λ_i. The algorithm decodes by attempting to solve the corresponding linear systems with v errors, where v ranges from $2e$ 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, 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 NC^2 [6], f can be computed by an NC^2 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

$$q^{q/2} \cdot \sum_{i=0}^{2e} \binom{q-1}{i} (q-1)^i \le q^{q/2} \cdot 2^{q h_q(1/6)} \le 2^{2n/3+o(n)}$$

possible strings, where

$$h_q(x) := x \log_q(q-1) - x \log_q x - (1-x)\log_q(1-x)$$

is the q-ary entropy function. Therefore, f distinguishes $D+D$ from the uniform distribution with probability at least $1-2^{-n/4}$. □

2.3 One-way log-space computation

Proof of Theorem 1.4.iii. Let q be a power of 2. Consider the $[q-1, q-6c, 6c]_{2^{\log q}}$ Reed–Solomon code C over $\mathbb{F}_{2^{\log q}}$, which has dual minimum distance $q-6c+1$ and can decode from $3c$ errors. Applying Fact 1.12 to C with $e = c$, we obtain a distribution D over $n := (q-1)\log q$ bits that is $O(c \log n/n)^c$-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 space $(2e+O(1))\log q$. Finally, using the Peterson–Gorenstein–Zierler syndrome-decoding algorithm, f accepts if and only if y differs from a codeword of C by at most $2e$ positions.

Since f accepts at most

$$q^{q-6c} \cdot \sum_{i=0}^{2e} \binom{q-1}{i} (q-1)^i \leq q^{q-6c} \cdot 2q^{4c} \leq O(q^{q-2c})$$

strings, f distinguishes $D+D$ from uniform with probability $1-O(\log n/n)^{2c}$. $\square$

Computing the input for syndrome decoding requires space $(2e+O(1))\log q$. We now show that in order to obtain n^{-c} bias via our construction, we always have $2e\log q = \Omega(c\log n)$. Thus, one cannot answer the "RL question" in the negative via syndrome decoding.

Claim 2.2. *For every $q \geq n+1$, let C be an $[n,k,d]$ code over $\mathbb{F}_q$ which decodes from e errors, and $d^\perp$ be its dual minimum distance. If C satisfies $(1-d^\perp/n)^e < q^{-c}$ for sufficiently large c, then we have $e\log q = \Omega(c\log n)$.*

Proof. If $d^\perp > (1-1/q)n$, then by the Plotkin bound on the dual code, $n-k = O(1)$. By the Singleton bound, $e \leq d \leq n-k$ and so we have $e = O(1)$. Hence, $(1-d^\perp/n)^e = (1/q)^e \geq q^{-c}$ for sufficiently large 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 $(1-d^\perp/n)^e < q^{-c}$ implies $e\log q > c\log q > c\log n$. $\square$

2.4 Depth 3 circuits, DNF formulas and AC^0 circuits

Proof of Theorem 1.5. We will use Valiant's depth reduction [35, 36] (cf. [38, Theorem 25]).

Theorem 2.3 ([35, 36])**.** *Let $C\colon \{0,1\}^n \to \{0,1\}$ be a circuit of size cn, depth $c\log n$ and fan-in 2. The function computed by C can also be computed by an unbounded fan-in circuit of size $2^{c'n/\log\log n}$ and depth 3 with inputs $x_1, x_2, \ldots, x_n, \overline{x_1}, \overline{x_2}, \ldots, \overline{x_n}$, where c' depends only on c.*

By the assumption that P has fan-in 2 circuits of linear size and logarithmic depth and the fact that f in Theorem 1.4.i is in P, we can apply Theorem 2.3 to f and obtain an unbounded fan-in depth-3 circuit f' of size $2^{O(n/\log\log n)}$ that computes the same function. Then we scale down n to $n' = \log n \log\log\log\log n$ bits (we set the rest of the $n-n'$ bits uniformly at 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 1.5.i.

To prove Theorem 1.5.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. $\square$

Proof of Theorem 1.6. Let D and f be the distribution and distinguisher in Theorem 1.4.ii, respectively. 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 uniformly at random.) $D_{n'}$ has bias $2^{-\Omega(n'/\log n')} = n^{-\Omega(\log^c n)}$. By our assumption, $f_{n'}$ is in AC^0 and distinguishes $D_{n'}+D_{n'}$ from uniform with probability $1-n^{-\log^c n/4}$. $\square$

2.5 Mod 3 linear functions

Recall the definition of mod p rank in Definition 1.8.

Fact 2.4 (Lemma 7.1 and 7.2 in [22])**.** *Let S,T be two sets of vectors in $\mathbb{F}_3^n$. 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) $\mathrm{rank}_3(S^2) \leq \mathrm{rank}_3(S)^2$;

(2) when S and T are subsets in $\{0,1\}^n \subseteq \mathbb{F}_3^n$,

$$\mathrm{rank}_3(S +_2 T) \leq \mathrm{rank}_3(S) + \mathrm{rank}_3(T) + \mathrm{rank}_3(S)\,\mathrm{rank}_3(T)\,.$$

If $S = T$, then $\mathrm{rank}_3(S +_2 S) \leq \mathrm{rank}_3(S) + \mathrm{rank}_3(S)^2$.

Proof. Let $d_S := \mathrm{rank}_3(S)$ and $d_T := \mathrm{rank}_3(T)$. Let $\{\beta_1, \ldots, \beta_{d_S}\}$ be a basis of S and $\{\gamma_1, \ldots, \gamma_{d_T}\}$ be a basis of T. Let

$$x = \sum_{i=1}^{d_S} c_i \beta_i \in S \quad \text{and} \quad y = \sum_{j=1}^{d_T} r_j \gamma_j \in T$$

be any two vectors. We have

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

Thus $\{\beta_i \times_3 \beta_j\}_{i,j \in [d_S]}$ spans S^2, proving (1). For (2), observe that for any $a,b \in \{0,1\} \subseteq \mathbb{F}_3$, we have $a +_2 b = a +_3 b +_3 a \times_3 b$. Hence we have

$$x +_3 y +_3 x \times_3 y = \sum_{i=1}^{d_S} c_i \beta_i + \sum_{j=1}^{d_T} r_j \gamma_j + \sum_{i \in [d_S], j \in [d_T]} c_i r_j (\beta_i \times_3 \gamma_j)\,,$$

and thus

$$\{\beta_i\}_{i \in [d_S]} \cup \{\gamma_j\}_{j \in [d_T]} \cup \{\beta_i \times_3 \gamma_j\}_{i \in [d_S], j \in [d_T]}$$

spans $S +_2 T$. □

The following lemma is well-known (cf. [25]). We include a proof here for completeness.

Lemma 2.5. *There exists a pseudo-design $(S_1, \ldots, S_n)$ 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 $d = n^{2\beta}, t = n^{\beta}$, and $\hat{t} = \log n$ for any $\beta < 0.5$.

We will use the following Chernoff bound in the proof.

Claim 2.6 (Chernoff bound)**.** *Let $X_1, \ldots, X_n \in \{0,1\}$ be n independent and identically distributed variables with $\mathrm{E}[X_i] = \mu$ for each i. We have*

$$\Pr\left[\left|\sum_{i=1}^{n} X_i - \mu n\right| \geq t\right] \leq e^{-t^2/4n}\,.$$

Proof of Lemma 2.5. 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 $[d]$ to be in S with probability $p = 0.1\log n/n^{\beta}$. We have $\mathrm{E}[|S|] = pd \geq 2n^{\beta}$. By the Chernoff bound,

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

We also have $\mathrm{E}[|S \cap S_j|] = pt = 0.1\log n$. Again by the 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 1.4.iv. Let $\alpha < 1/36$ and $\beta = 4\alpha$. Also let $d,t,\hat{t}$ be the parameters and $S_1,\dots,S_n$ be the pseudo-design specified in Lemma 2.5. Define the function $L\colon \{0,1\}^d \to \{0,1\}^n$ whose i-th output bit y_i equals

$$\mathrm{mod}_3^2(x_{S_i}) := \Big(\sum_{j\in S_i} x_j\Big)^2 \bmod 3\,.$$

Let T_1 be the image set of L. Without the square, this set has mod 3 rank d and so by Fact 2.4, $\mathrm{rank}_3(T_1) = O(d^2) = O(n^{16\alpha})$. Let T_2 be an ε-almost k-wise independent set, where $\varepsilon = 1/n^{\alpha}$ and $k = 2\log n$. Known constructions [2, Theorem 2] (see also [24]) produce such a set of size $O((k\log n)/\varepsilon)^2$ and therefore $\mathrm{rank}_3(T_2)$ is at most $O(n^{2\alpha}\log^4 n)$.

Consider the set $T := T_1 +_2 T_2$. By Fact 2.4, T has rank at most $O(n^{18\alpha}\log^4 n)$. By the same fact, $T +_2 T$ has rank at most $O(n^{36\alpha}\log^8 n) < n$ because $\alpha < 1/36$. Therefore, there is a non-zero mod 3 linear function ℓ such that $\ell(y) \equiv 0 \pmod 3$ for any $y \in T$, while $\Pr[\ell(y) = 0] \leq 1/2$ for a uniform y in $\{0,1\}^n$. It remains to show that T is $O(1/n^{0.99\alpha})$-biased. For any test on $I \subseteq [n]$, we consider the cases (1) when $|I| \leq k$, and (2) when $|I| > k$ separately.

Write $y = y_1 + y_2$, 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 2.7. *For any $|I| > k$, we have $|\mathrm{E}_{y_1\in T_1}[\chi_I(y_1)]| \leq O(1/n^{0.99\alpha})$, where $\chi_I(z) := (-1)^{\sum_{i\in I} z_i}$.*

Proof. Pick a subset $J \subseteq I$ of size k. Define $f,p\colon \{0,1\}^n \to \{0,1\}$ to be

$$f(x) := \sum_{i\in J} \mathrm{mod}_3^2(x_{S_i}) \quad \text{and} \quad p(x) := \sum_{i\in I\setminus J} \mathrm{mod}_3^2(x_{S_i})\,,$$

respectively. Observe that

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

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, $B_1 := \{\ell \in [d] : \ell \in S_i \cap S_j$ for some distinct $i,j \in J\}$. Fixing the value of every $x_\ell \in B_1$, each $\mathrm{mod}_3^2(x_{S_j})$ 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 S_j for $j \in J$. The bits in B_2 do not affect the outputs in J. Fixing their values, each $\text{mod}_3^2(x_{S_j})$ in p is a function of at most $\hat{t} \cdot k = O(\log^2 n)$ bits and so can be written as a polynomial of degree $O(\log^2 n)$ over $\mathbb{F}_2$. Since p is a parity of values $\text{mod}_3^2(x_{S_j})$, it can also be written as a polynomial of degree $O(\log^2 n)$ over $\mathbb{F}_2$.

We will use the following theorem by Smolensky [33] (cf. [38, Chapter 1]). The proof in [38] has the condition that n is divisible by 3. This condition can be removed. For example, when $n = 3\ell + 1$, we can set a random bit of the uniform distribution to zero. This distribution is close to uniform, but now we can apply [38] as stated.

Theorem 2.8 ([33]). *There exists an absolute constant $\varepsilon > 0$ such that for every n that is divisible by 3 and for every polynomial $p\colon \mathbb{F}_2^n \to \mathbb{F}_2$ of degree at most $\varepsilon\sqrt{n}$, we have*

$$\mathrm{E}_{x \sim \{0,1\}^n}\left[(-1)^{\text{mod}_3^2(x)+p(x)}\right] \le 0.9\,.$$

To build intuition, note that after fixing the input bits in B_1 and B_2, for each of the $\text{mod}_3^2(x_{S_j})$ in f, by Theorem 2.8 we have

$$\left|\mathrm{E}_{x_i : i \in [d]}\left[(-1)^{(\text{mod}_3^2(x_{S_j})+p(x))}\right]\right| \le 1 - \Omega(1)\,.$$

In the following lemma we prove a variant of Impagliazzo's XOR Lemma [18] to show that

$$\left|\mathrm{E}_{x_i : i \in [d]}\left[(-1)^{(f(x)+p(x))}\right]\right| \le O(1/m^{0.249}) = O(1/n^{0.99\alpha})\,.$$

Averaging over the values of the x_k in B_1 and B_2 finishes the proof. □

Lemma 2.9. *Let $k = 2\log m$, define $f\colon \{0,1\}^{m\times k} \to \{0,1\}$ by*

$$f(x^{(1)}, \dots, x^{(k)}) := \text{mod}_3^2(x^{(1)}) +_2 \cdots +_2 \text{mod}_3^2(x^{(k)})\,.$$

Let $p\colon \{0,1\}^{m\times k} \to \{0,1\}$ be any polynomial of degree $O(\log^2 m)$. We have

$$\mathrm{Cor}(f, p) := \mathrm{E}_{x\sim\{0,1\}^{m\times k}}\left[(-1)^{f(x)+p(x)}\right] \le O(1/m^{0.249})\,.$$

Proof. We will use the fact that Theorem 2.8 holds for degree $n^{\Omega(1)}$ polynomials to get correlation $1/n^{\Omega(1)}$ for polynomials of degree $\mathrm{polylog}(n)$.

As in the proof in [18] we first show the existence of a measure $M\colon \{0,1\}^m \to [0,1]$ of size at most $|M| := \sum_x M(x) = 2^m/4$ such that with respect to its induced distribution $D(x) := M(x)/|M|$, the function mod_3^2 is $1/2m^{0.249}$-hard for any polynomial p of degree $O(\log^2 m)$, i. e.,

$$\Pr_{x\sim D}[\text{mod}_3^2(x) = p(x)] \le 1/2 + 1/4m^{0.249}\,.$$

Suppose not. Lemma 1 in [18] implies that one can obtain a function q by taking the majority of $O(m^{0.498})$ polynomials of degree $O(\log^2 m)$ such that

$$\Pr_{x\sim\{0,1\}^m}[\text{mod}_3^2(x) = q(x)] > 0.99\,.$$

Note that q can be represented as a degree $O(m^{0.498}\log^2 m)$ polynomial. By Theorem 2.8,

$$\Pr_{x\sim\{0,1\}^m}[\mathrm{mod}_3^2(x)=p(x)]\le 0.95$$

for any degree $\varepsilon m^{1/2}$ polynomial p, a contradiction.

Now we show that there is a set $S\subseteq\{0,1\}^m$ of size $2^m/8$ 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.,

$$\Pr_{x\sim S}[\mathrm{mod}_2^3(x)=p(x)]\le 1/2+1/2m^{0.249}.$$

Let p be any degree-$O(\log^2 m)$ polynomial. For any measure $M\colon \{0,1\}^m\to[0,1]$, define $\mathrm{Adv}_p(M)$ by

$$\mathrm{Adv}_p(M):=\sum_x M(x)(-1)^{\mathrm{mod}_3^2(x)+p(x)}.$$

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 $\mathrm{E}_S[\mathrm{Adv}_p(M_S)]=\mathrm{Adv}_p(M)\le |M|/(2m^{0.249})$. Note that $\mathrm{Adv}_p(M_S)$ is the sum of 2^m independent random variables, where each variable is over $[-1,0]$ or $[0,1]$. By Hoeffding's inequality,

$$\Pr_S[\mathrm{Adv}_p(M_S)>|M|/m^{0.249}]\le 2^{-2|M|^2/(2^m\cdot 4m^{0.498})}=2^{-2^m/32m^{0.498}}.$$

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

It follows that there exists a set of inputs $S\subseteq\{0,1\}^m$ of size at most $2^m/8$ such that mod_3^2 is $1/m^{0.249}$-hard-core on S for any polynomial of degree $O(\log^2 m)$. Now we apply the following lemma, which is stated in [18] for circuits, but the same proof applies to polynomials.

Lemma 2.10 (Lemma 4 in [18]). *If g is ε-hard-core for some set of $\delta 2^n$ inputs for polynomials of degree d, then the function $f(x^{(1)},\ldots,x^{(k)}):=g(x^{(1)})+_2\cdots+_2 g(x^{(k)})$ is $\varepsilon+(1-\delta)^k$-hard-core for polynomials of the same degree.*

Applying this lemma with our choice of k, we have for any polynomial p of degree $O(\log^2 m)$,

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

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

Proof of Claim 1.10. We replace the pseudo-design in the proof of Theorem 1.4.iv 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 2.5 one can show that such pseudo-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 function on $m=O(\log^4 n)$ bits and a polynomial of degree $O(\log^2 n)$, respectively. This gives bias $O(1/n^c)$. Note that the image set T_1 now has mod 3 rank $d^2=O(\log^{16} n)$.

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

By Fact 2.4, $T:=T_1+_2 T_2$ has mod 3 rank $O(n^{0.49}\log^{16} n)=o(n^{0.5})$. Hence, $T+_2 T$ has rank less than n and the claim follows. □

2.6 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 1.4, 1.5 and 1.6.

Proof of "Moreover" part in Theorem 1.4, 1.5 and 1.6. To prove Theorem 1.4.i, 1.4.ii and 1.4.iii, 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, $2e/k \geq 1$. The rest follows by noting the sum of k copies of D' is identical to $D+D$.

By scaling down the above small-bias distributions D' for Theorem 1.4.i and 1.4.ii to n' bits as in the proofs of Theorem 1.5 and 1.6, 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 1.6 and 1.5. Moreover, k scales from $k(n)$ to $k(n')$.

For Theorem 1.4.iv, 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 an $n^{-2\alpha/k}$-almost $2\log n$-wise independent set in the proof, and call them T_1' and T_2', respectively. We now have $\mathrm{rank}_3(T_1') = O(n^{32\alpha/k})$ and $\mathrm{rank}_3(T_2') = O(n^{4\alpha/k}\log^4 n)$. Thus, the set $T' := T_1' +_2 T_2'$ has rank at most $O(n^{36\alpha/k}\log^4 n)$ and therefore the sum of k copies has rank at most $\mathrm{rank}_3(T')^k = O(n^{36\alpha}\log^{4k} n) < n$, 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 $2\log n$, and T_1' has bias $O(n^{-2\alpha/k})$ for tests of size greater than $2\log n$. □

3 Mod 3 rank of k-wise independence

In this section, we begin a systematic investigation on the mod 3 rank of k-wise independent distributions.

Recall Definition 1.8 of mod p rank. We also define the mod p rank of a matrix over the integers to be its rank over $\mathbb{F}_p$. We also write rank_p for mod p rank.

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 mod 3 rank 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 mod 3, 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 rank 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 3.1. *Let D be any subset $\{0,1\}^n$. If $\mathrm{rank}_3(D) = t$, then D is not $1/10$-almost $ct/\log(n/t)$-wise independent, for a universal constant c.*

This gives an exponential separation between seed length and rank for almost k-wise independence. Indeed, for $k = O(1)$, the seed length is $\Theta(\log\log n)$, whereas the rank 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. Moreover, by the Singleton bound, $d^\perp - 1 \le t$. By the Hamming bound, that is,

$$3^{n-t} \sum_{i=0}^{\lfloor \frac{d^\perp - 1}{2} \rfloor} \binom{n}{i} 2^i \le 3^n\,,$$

we have

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

where we use the fact that $d^\perp - 1 \le t$ in the last inequality. Hence, $d^\perp \le O(t/\log(n/t))$. Now, let y be a codeword in $C^\perp$ with Hamming weight $d^\perp$. Let $I := \{i \mid y_i \ne 0\}$. 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 in $\{0,1\}^I$ we have $\langle y, x \rangle_3 = 0$ with probability at most $1/2$. Therefore, D is constant bounded away from uniform on the $d^\perp$ bits indexed by I, and thus cannot be close to $d^\perp$-wise independent. $\square$

3.2 Pairwise independence

We now show that the mod 3 rank of a pairwise independent set can be as small as $n^{0.73}$. Then we give evidence that our approach cannot do any better.

Definition 3.2. We say H is an Hadamard matrix of order n if its entries are ± 1 and it satisfies $HH^T = nI_n$, where I_n is the $n \times n$ identity matrix.

It is well-known that by removing the all-ones row of an Hadamard matrix H, which can always be created by multiplying each column by its first element, the uniform distribution over the columns of the truncated matrix is pairwise independent.

Henceforth we will work with vectors whose entries are from $\{-1,1\} = \{2,1\} \subseteq \mathbb{F}_3$. The following two claims show that certain Hadamard matrices cannot have dimension smaller than $n/2$. They are taken from [41], and here we give a self-contained proof for completeness. First, we would give a lower bound to the mod p rank from the determinant of any square matrix.

Claim 3.3 (Theorem 1 in [41])**.** *Let A be an $n \times n$ matrix over the integers. Assume $p^{e+1} \nmid \det(A)$. Then* $\mathrm{rank}_p(A) \geq n - e$.

Proof. Suppose $\mathrm{nullity}_p(A) = n - r$. Let $(\beta_1, \ldots, \beta_{n-r})$ be a basis of the null space of A over $\mathbb{F}_p$. Extend the basis to $(\beta_1, \ldots, \beta_n)$ so that it forms a basis of $\mathbb{F}_p^n$. Let B be the matrix whose columns are the β_i. Note that $\det(B) \not\equiv 0 \pmod p$ and $\det(AB) = \det(A)\det(B)$. Thus, $p^s \mid \det(A)$ if and only if $p^s \mid \det(AB)$. By construction, $\beta_1, \ldots, \beta_{n-r}$ 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 $p^{n-r} \mid \det(A)$. □

Claim 3.4 (Theorem 2 in [41])**.** *Let H be an $n \times n$ Hadamard matrix. Let p be an odd prime such that $p^2 \nmid n$. Then* $\mathrm{rank}_p(H) \geq n/2$.

Proof. Since H is an 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 on p we have that $p^{n/2+1} \nmid n^{n/2}$. Hence, it follows from Claim 3.3 that $\mathrm{rank}_p(H) \geq n/2$. □

The following claim characterizes Hadamard matrices with mod p rank at most $n/2$.

Claim 3.5. *Let H be an $n \times n$ Hadamard matrix. If $p \mid n$, then* $\mathrm{rank}_p(H) \leq n/2$. *Otherwise,* $\mathrm{rank}_p(H) = n$.

Proof. If $p \mid n$, then by Sylvester's rank inequality, we have

$$\mathrm{rank}_p(H) + \mathrm{rank}_p(H^T) - n \leq \mathrm{rank}_p(HH^T) = \mathrm{rank}_p(nI_n) = 0\,.$$

Hence, $\mathrm{rank}_p(H) \leq n/2$, proving the first part. For the second part, if $p \nmid n$ then

$$\det(H)^2 = \det(H)\det(H^T) \not\equiv 0 \pmod p$$

and so $\det(H) \not\equiv 0 \pmod p$. Hence $\mathrm{rank}_p(H) = n$. □

Now we give a construction of Hadamard matrices whose orders do not satisfy the condition in Claim 3.4. These matrices have much smaller mod p ranks than the lower bound stated in Claim 3.4. Note that the affine bijection $L\colon \{-1,1\}^n \to \{0,1\}^n$ 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 3.6. *Let $S \subseteq \{-1,1\}^n$ be a set containing the all-ones vector. Then* $\mathrm{rank}_3(L(S)) \leq \mathrm{rank}_3(S)$.

Fact 3.7. *If A and B are two Hadamard matrices then $A \otimes B$ is also an Hadamard matrix, where $\otimes$ indicates the Kronecker product.*

The following is well-known.

Fact 3.8. *Let A, B be two matrices over any field. Then we have* $\mathrm{rank}(A \otimes B) = \mathrm{rank}(A) \cdot \mathrm{rank}(B)$.

Claim 3.9. *For infinitely many values of n, there exists a pairwise independent distribution over $\{0,1\}^n$ with* mod 3 *rank at most $n^{0.73}$.*

Proof. Paley [26] constructed a $(q-1)\times(q-1)$ Hadamard matrix for every prime power $q \equiv 1 \pmod 4$. Starting with an Hadamard matrix H_{12} over $\{-1,1\} = \{2,1\} \subseteq \mathbb{F}_3$ using Paley's construction, 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 3.4 and 3.5 that $\mathrm{rank}_3(H_{12}) = 6$. Hence, by Fact 3.8, H_n has rank $6^{\log_{12} n} \le n^{0.73}$. As discussed above, we can assume H_{12} contains an all-ones row. Thus, H_n also contains an all-ones row. and the claim follows from Fact 3.6. □

We start from an $m \times m$ Hadamard matrix with mod 3 rank $m/2$, for some m. The smaller m we start from, the better exponent we get. Since Hadamard matrices must be of order 1, 2, or multiple of 4, Claim 3.4 implies that 12 is indeed the smallest possible m.

4 Complexity of decoding

In this section we prove some negative results on the complexity of decoding.

In [37, Chapter 6] and [31] 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 ε-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 at least $2/3$ 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 the 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 by at most $(1/2-\varepsilon)d$, respectively.

The results in this section have been obtained while attempting to bound the complexity of the threshold discriminator explained in Section 1.2: distinguish a random noisy codeword from uniform. The ε-DECODE problem is different, as it asks to distinguish between two noisy codewords.

Lemma 4.1. *If a function $D\colon \{0,1\}^n \to \{0,1\}$ solves ε-DECODE, then a restriction of D solves ε-MAJ on d bits.*

Proof. Let $x, y \in C$ be the codewords at Hamming distance d. 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. □

Shaltiel and Viola [31] show that depth-c $\mathsf{AC}^0[\oplus]$ circuits can solve ε-MAJ only if ε is at least $1/O(\log n)^{(c+2)}$. Brody and Verbin [11] 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 4.1, we have the following claim.

Claim 4.2. *Let* $D\colon \{0,1\}^n \to \{0,1\}$ *be a function.*

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

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

Claim 4.3. *Let* $C \subseteq \{0,1\}^n$ *be an* $[n,k,d]$ *code with dual minimum distance* $d^{\perp}$*. If*

$$2^{-t} > 16\left(1 - \frac{d^{\perp}}{n}\right)^{e/2^{t-1}}$$

for some constant t and $e \leq \lfloor \frac{d-1}{2} \rfloor$*, then no degree-$t$ polynomial over* $\mathbb{F}_2$ *can be a threshold-te discriminator for C.*

Proof. Suppose on the contrary a polynomial P is a threshold-te discriminator for C. By Fact 1.12 and the Schwartz–Zippel Lemma, there exists an $\varepsilon := (1-d^{\perp}/n)^e$-biased distribution D such that P distinguishes the sum of t independent copies of D from uniform with probability at least 2^{-t}. But by [39], the sum of t copies of D fools P with probability $16\varepsilon^{1/2^{t-1}}$, a contradiction. □

5 Fooling read-once DNF formulas

In this section we state and prove that an $m^{-O(\log(1/\delta))}$ bound on the bias suffices to δ-fool any read-once DNF formulas with m terms. This directly follows from Lemma 5.2 in [12].

Claim 5.1. *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 [12], $|\Pr_{x\sim D}[\phi(x)] - \Pr_{x\sim\{0,1\}^n}[\phi(x)]|$ is upper bounded by

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

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. □

Acknowledgments. A previous version of this paper made an unconditional claim about AC^0. Specifically, we claimed Theorem 1.6 without the assumption in the first sentence. We are grateful to Andrej Bogdanov for pointing out this mistake.

We are grateful to Xue Chen and David Zuckerman for telling us an alternative proof of Theorem 1.4.iii mentioned in Section 1.1. We also thank Xue Chen for pointing out that the proof of Theorem 1.4.iv was written with the wrong pseudo-design parameters. We are also grateful to Ravi Boppana for the detailed feedback on a preliminary version of this paper, in particular pointing out several inaccuracies in our definitions and in our empirical results.

We wish to thank the referees for the detailed and useful feedback, and in particular for pointing out several inaccuracies.

References

[1] 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. *IRE Trans. Inform. Theory*, 38(2):509–516, 1992. Preliminary version in ISIT'91. [doi:10.1109/18.119713] 2

[2] NOGA ALON, ODED GOLDREICH, JOHAN HÅSTAD, AND RENÉ PERALTA: Simple construction of almost k-wise independent random variables. *Random Structures & Algorithms*, 3(3):289–304, 1992. Preliminary version in FOCS'90. [doi:10.1002/rsa.3240030308] 2, 12

[3] BENNY APPLEBAUM, ANDREJ BOGDANOV, AND ALON ROSEN: A dichotomy for local small-bias generators. *J. Cryptology*, 29(3):577–596, 2016. Preliminary version in TCC'12. [doi:10.1007/s00145-015-9202-8] 6

[4] LOUAY M. J. BAZZI: Polylogarithmic independence can fool DNF formulas. *SIAM J. Comput.*, 38(6):2220–2272, 2009. Preliminary version in FOCS'07. [doi:10.1137/070691954] 4, 7

[5] AVRAHAM BEN-AROYA AND AMNON TA-SHMA: Constructing small-bias sets from algebraic-geometric codes. *Theory of Computing*, 9(3):253–272, 2013. Preliminary version in FOCS'09. [doi:10.4086/toc.2013.v009a005] 2

[6] 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. [doi:10.1016/0020-0190(84)90018-8] 9

[7] ANDREJ BOGDANOV, ZEEV DVIR, ELAD VERBIN, AND AMIR YEHUDAYOFF: Pseudorandomness for width-2 branching programs. *Theory of Computing*, 9(7):283–293, 2013. [doi:10.4086/toc.2013.v009a007] 2, 3, 6

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

[9] RAVI B. BOPPANA, JOHAN HÅSTAD, CHIN HO LEE, AND EMANUELE VIOLA: Bounded independence vs. moduli. In *Proc. 20th Internat. Workshop on Randomization and Computation (RANDOM'16)*, LIPIcs, pp. 24:1–24:9. Leibniz-Zentrum fuer Informatik, 2016. [doi:10.4230/LIPIcs.APPROX-RANDOM.2016.24] 5

[10] MARK BRAVERMAN: Poly-logarithmic independence fools AC^0 circuits. *J. ACM*, 57(5):28:1–28:10, 2010. Preliminary version in CCC'09. [doi:10.1145/1754399.1754401] 4, 7

[11] JOSHUA BRODY AND ELAD VERBIN: The coin problem and pseudorandomness for branching programs. In *Proc. 51st FOCS*, pp. 30–39. IEEE Comp. Soc. Press, 2010. [doi:10.1109/FOCS.2010.10] 18

[12] SURESH CHARI, PANKAJ ROHATGI, AND ARAVIND SRINIVASAN: Improved algorithms via approximations of probability distributions. *J. Comput. System Sci.*, 61(1):81–107, 2000. Preliminary version in STOC'94. [doi:10.1006/jcss.1999.1695] 3, 19

[13] THOMAS M. COVER AND JOY A. THOMAS: *Elements of Information Theory*. Wiley-Interscience, 2006. 8

[14] ANINDYA DE, OMID ETESAMI, LUCA TREVISAN, AND MADHUR TULSIANI: Improved pseudorandom generators for depth 2 circuits. In *Proc. 14th Internat. Workshop on Randomization and Computation (RANDOM'10)*, volume 6302 of *LNCS*, pp. 504–517. Springer, 2010. [doi:10.1007/978-3-642-15369-3_38] 3, 7

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

[16] DANIEL GORENSTEIN AND NEAL ZIERLER: A class of error-correcting codes in p^m symbols. *J. Soc. Indust. Appl. Math.*, 9(2):207–214, 1961. [doi:10.1137/0109020] 6, 9

[17] VENKATESAN GURUSWAMI AND ATRI RUDRA: Tolerant locally testable codes. In *Proc. 9th Internat. Workshop on Randomization and Computation (RANDOM'05)*, volume 3624 of *LNCS*, pp. 306–317. Springer, 2005. [doi:10.1007/11538462_26] 6

[18] RUSSELL IMPAGLIAZZO: Hard-core distributions for somewhat hard problems. In *Proc. 36th FOCS*, pp. 538–545. IEEE Comp. Soc. Press, 1995. [doi:10.1109/SFCS.1995.492584] 7, 13, 14

[19] SWASTIK KOPPARTY AND SHUBHANGI SARAF: Tolerant linearity testing and locally testable codes. In *Proc. 13th Internat. Workshop on Randomization and Computation (RANDOM'09)*, volume 5687 of *LNCS*, pp. 601–614. Springer, 2009. [doi:10.1007/978-3-642-03685-9_45] 6

[20] CHIN HO LEE AND EMANUELE VIOLA: Some limitations of the sum of small-bias distributions. *Electron. Colloq. on Comput. Complexity (ECCC)*. TR15-005, 2016. 1, 5

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

[22] RAGHU MEKA AND DAVID ZUCKERMAN: Small-bias spaces for group products. In *Proc. 13th Internat. Workshop on Randomization and Computation (RANDOM'09)*, volume 5687 of *LNCS*, pp. 658–672. Springer, 2009. [doi:10.1007/978-3-642-03685-9_49] 2, 3, 4, 7, 11

[23] ELCHANAN MOSSEL, AMIR SHPILKA, AND LUCA TREVISAN: On epsilon-biased generators in NC^0. *Random Structures & Algorithms*, 29(1), 2005. Preliminary versions in FOCS'03 and ECCC. [doi:10.1109/SFCS.2003.1238188] 6

[24] JOSEPH NAOR AND MONI NAOR: Small-bias probability spaces: efficient constructions and applications. *SIAM J. Comput.*, 22(4):838–856, 1993. Preliminary version in STOC'90. [doi:10.1137/0222053] 2, 6, 12

[25] NOAM NISAN: Pseudorandom bits for constant depth circuits. *Combinatorica*, 11(1):63–70, 1991. Preliminary version in FOCS'88. [doi:10.1007/BF01375474] 7, 11

[26] RAYMOND E. A. C. PALEY: On orthogonal matrices. *J. Math. Phys.*, 12(1-4):311–320, 1933. [doi:10.1002/sapm1933121311] 18

[27] WILLIAM WESLEY PETERSON: Encoding and error-correction procedures for the Bose-Chaudhuri codes. *IRE Trans. Inform. Theory*, 6(4):459–470, 1960. [doi:10.1109/TIT.1960.1057586] 6, 9

[28] ALEXANDER A. RAZBOROV: Lower bounds on the dimension of schemes of bounded depth in a complete basis containing the logical addition function. *Mat. Zametki Akad. Nauk SSSR*, 41(4):598–607, 1987. English translation in Math. Notes of the Acad. Sci. of the USSR, 41(4):333-338, 1987. [doi:10.1007/BF01137685] 2

[29] ALEXANDER A. RAZBOROV: A simple proof of Bazzi's theorem. *ACM Trans. Comput. Theory*, 1(1):3:1–3:5, 2009. [doi:10.1145/1490270.1490273] 4

[30] ATRI RUDRA AND STEVE UURTAMO: Data stream algorithms for codeword testing. In *Proc. 37th Internat. Colloq. on Automata, Languages and Programming (ICALP'10)*, pp. 629–640. Springer, 2010. [doi:10.1007/978-3-642-14165-2_53, arXiv:1004.4601] 6, 7

[31] RONEN SHALTIEL AND EMANUELE VIOLA: Hardness amplification proofs require majority. *SIAM J. Comput.*, 39(7):3122–3154, 2010. Preliminary version in STOC'08. [doi:10.1137/080735096] 7, 18

[32] AMIR SHPILKA: Constructions of low-degree and error-correcting ε-biased generators. *Comput. Complexity*, 18(4):495–525, 2009. Preliminary version found in CCC'06. [doi:10.1007/s00037-009-0281-5] 6, 8

[33] ROMAN SMOLENSKY: Algebraic methods in the theory of lower bounds for boolean circuit complexity. In *Proc. 19th STOC*, pp. 77–82. ACM Press, 1987. [doi:10.1145/28395.28404] 7, 13

[34] AVISHAY TAL: Tight bounds on the Fourier spectrum of AC^0. In *Proc. 32nd IEEE Conf. on Computational Complexity (CCC'17)*, volume 79 of *LIPIcs*, pp. 15:1–15:31. Leibniz-Zentrum fuer Informatik, 2017. Preliminary version in ECCC'14. [doi:10.4230/LIPIcs.CCC.2017.15] 4, 7

[35] LESLIE G. VALIANT: Graph-theoretic arguments in low-level complexity. In *6thSymp. Math. Found. Comp. Sci. (MFCS'77)*, volume 53 of *LNCS*, pp. 162–176. Springer, 1977. [doi:10.1007/3-540-08353-7_135] 3, 7, 10

[36] LESLIE G. VALIANT: Exponential lower bounds for restricted monotone circuits. In *Proc. 15th STOC*, pp. 110–117. ACM Press, 1983. [doi:10.1145/800061.808739] 10

[37] EMANUELE VIOLA: *The Complexity of Hardness Amplification and Derandomization*. Ph. D. thesis, Harvard University, 2006. 7, 18

[38] EMANUELE VIOLA: On the power of small-depth computation. *Found. Trends Theor. Comput. Sci.*, 5(1):1–72, 2009. [doi:10.1561/0400000033] 2, 7, 10, 13

[39] EMANUELE VIOLA: The sum of d small-bias generators fools polynomials of degree d. *Comput. Complexity*, 18(2):209–217, 2009. Preliminary version in CCC'08. [doi:10.1007/s00037-009-0273-5] 2, 19

[40] EMANUELE VIOLA AND AVI WIGDERSON: Norms, XOR lemmas, and lower bounds for polynomials and protocols. *Theory of Computing*, 4(7):137–168, 2008. Preliminary version in CCC'07. [doi:10.4086/toc.2008.v004a007] 8

[41] RICHARD M. WILSON: Combinatorial analysis lecture notes. 2012. 16, 17

AUTHORS

Chin Ho Lee
Ph. D. student
College of Computer and Information Science
Northeastern University
Boston, MA
chlee@ccs.neu.edu
http://www.ccs.neu.edu/~chlee

Emanuele Viola
Associate professor
College of Computer and Information Science
Northeastern University
Boston, MA
viola@ccs.neu.edu
http://www.ccs.neu.edu/~viola

ABOUT THE AUTHORS

CHIN HO LEE is a Ph. D. student at the College of Computer and Information Science of Northeastern University, under the supervision of Emanuele Viola. He obtained his M. Phil. in Computer Science and Engineering (advised by Andrej Bogdanov) and B. S. in Mathematics from the Chinese University of Hong Kong.

EMANUELE VIOLA moved four times in the academic year 2015–2016. He also has a blog! Check it out at https://emanueleviola.wordpress.com/.