Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 93 (2024)

Low-Degree Polynomials Are Good Extractors

Omar Alrabiah* Jesse Goodman† Jonathan Mosheiff‡ João Ribeiro§

Abstract

We prove that random low-degree polynomials (over $\mathbb{F}_2$) are unbiased, in an extremely general sense. That is, we show that random low-degree polynomials are good *randomness extractors* for a wide class of distributions. Prior to our work, such results were only known for the small families of (1) uniform sources, (2) affine sources, and (3) local sources. We significantly generalize these results, and prove the following.

1. **Low-degree polynomials extract from small families.** We show that a random low-degree polynomial is a good low-error extractor for *any* small family of sources. In particular, we improve the positive result of Alrabiah, Chattopadhyay, Goodman, Li, and Ribeiro (ICALP 2022) for local sources, and give new results for polynomial and variety sources via a single unified approach.
2. **Low-degree polynomials extract from sumset sources.** We show that a random low-degree polynomial is a good extractor for sumset sources, which are the most general *large* family of sources (capturing independent sources, interleaved sources, small-space sources, and more). Formally, for any even d, we show that a random degree d polynomial is an ε-error extractor for n-bit sumset sources with min-entropy $k = O(d(n/\varepsilon^2)^{2/d})$. This is nearly tight in the polynomial error regime.

Our results on sumset extractors imply new complexity separations for linear ROBPs, and the tools that go into its proof may be of independent interest. The two main tools we use are a new structural result on sumset-punctured Reed-Muller codes, paired with a novel type of reduction between extractors. Using the new structural result, we obtain new limits on the power of sumset extractors, strengthening and generalizing the impossibility results of Chattopadhyay, Goodman, and Gurumukhani (ITCS 2024).

*UC Berkeley. oalrabiah@berkeley.edu. Supported by a Saudi Arabian Cultural Mission (SACM) Scholarship, NSF Award CCF-2210823, and a Simons Investigator Award (Venkatesan Guruswami).

†UT Austin. jpmgoodman@utexas.edu. Supported by a Simons Investigator Award (#409864, David Zuckerman).

‡Ben-Gurion University. mosheiff@bgu.ac.il. Supported by an Alon Fellowship and by DOE grant # DE-SC0024124 while visiting the Simons Institute for the Theory of Computing

§Instituto de Telecomunicações and Departamento de Matemática, Instituto Superior Técnico, Universidade de Lisboa. jribeiro@tecnico.ulisboa.pt. Work mainly done while at NOVA LINCS and NOVA School of Science and Technology, and while visiting the Simons Institute for the Theory of Computing supported by DOE grant # DE-SC0024124. Also supported by NOVA LINCS (ref. UIDB/04516/2020) and by FCT/MECI through national funds and when applicable co-funded EU funds under UID/50008: Instituto de Telecomunicações.

ISSN 1433-8092

Contents

1 Introduction

In this work, we are interested in the following open-ended question:

How biased is a random degree d polynomial $f : \mathbb{F}_2^n \to \mathbb{F}_2$?

By *random* degree d polynomial, we mean a polynomial of the form

$$f(x) = \sum_{S \subseteq [n] : |S| \leq d} \alpha_S \cdot x^S,$$

where $x^S := \prod_{i \in S} x_i$ and the coefficients α_S are sampled independently and uniformly at random from $\mathbb{F}_2$. And by *bias*, perhaps the most basic definition is to simply check the difference between how many inputs are mapped to 0 and 1. Or more formally, letting $\mathbf{U}_n$ denote the uniform distribution over $\mathbb{F}_2^n$,

$$\mathrm{bias}(f) := \Pr_{x \sim \mathbf{U}_n}[f(x) = 0] - \Pr_{x \sim \mathbf{U}_n}[f(x) = 1].$$

For these definitions of *random* and *bias*, Ben-Eliezer, Hod, and Lovett [BHL12] – building on a pair of earlier papers [KS05, ABK08] – provided a complete answer to the above question. In particular, they showed sharp concentration bounds on $|\mathrm{bias}(f)|$, concluding that a random degree d polynomial is essentially unbiased on a uniform input, with extremely high probability. More precisely, they proved the following.

Theorem 0 (Random low-degree polynomials are unbiased [BHL12, Lemma 1.2])**.** *For every $\delta \in (0, 1)$ there is a constant $c > 0$ such that the following holds. Let $d \in \mathbb{N}$ be an integer satisfying $1 \leq d \leq (1-\delta)k$. Then for a random degree d polynomial $f : \mathbb{F}_2^n \to \mathbb{F}_2$,*

$$\Pr_f \left[|\mathrm{bias}(f)| > 2^{-cn/d} \right] \leq 2^{-c\binom{n}{\leq d}}.$$

They showed these bounds were tight, and a later work extended these results to all prime fields [BOY20].

While Theorem 0 is interesting in its own right (since low-degree polynomials are fundamental objects), its pursuit was largely motivated by applications in coding theory, complexity theory, and pseudorandomness. Indeed, given this new result, Ben-Eliezer, Hod, and Lovett immediately obtained important corollaries in each of these areas. In coding theory, they obtained new tail bounds on the weight distribution of Reed-Muller codes. In complexity theory, they showed that (random) low-degree polynomials cannot be approximated by polynomials of smaller degree. And in pseudorandomness, they obtained new lower bounds on the seed length of pseudorandom generators (PRGs) for low-degree polynomials.

Since we now understand the bias of a random low-degree polynomial on a *uniform* input, it is natural to ask whether a more general result can be proven for *weakly random* inputs - especially given the connection this problem has to pseudorandomness. But in order to make this question formal, we'll need a more general definition of *bias*, which allows the function to receive a weakly random input. Towards this end, given a random variable ("source") $\mathbf{X}$ over $\mathbb{F}_2^n$, we define

$$\mathrm{bias}_{\mathbf{X}}(f) := \Pr_{x \sim \mathbf{X}}[f(x) = 0] - \Pr_{x \sim \mathbf{X}}[f(x) = 1].$$

Given this definition, a simple observation is that $\mathrm{bias}_{\mathbf{U}_n}(f) = \mathrm{bias}(f)$, and thus establishing concentration bounds for $|\mathrm{bias}_{\mathbf{X}}(f)|$ is a strictly more general problem than doing so for $|\mathrm{bias}(f)|$. But how should we use this generality? And for what distributions $\mathbf{X}$ is it actually interesting to understand $|\mathrm{bias}_{\mathbf{X}}(f)|$? Recall that we would like to understand $|\mathrm{bias}_{\mathbf{X}}(f)|$ for *weakly random* $\mathbf{X}$, but it is still not clear what weakly random should mean. To answer all of these questions and more, we enter the world of *randomness extraction*.

Randomness extractors

Randomness extractors are fundamental objects in pseudorandomness and complexity theory. They are motivated by the fact that nature can only provide us with weak sources of randomness, yet most applications in computer science require perfectly uniform bits. Formally, they are defined as follows.

Definition 1 (Randomness extractor)**.** *Let $\mathcal{X}$ be a family of sources $\mathbf{X} \sim \{0,1\}^n$. A deterministic function $\mathsf{Ext} : \{0,1\}^n \to \{0,1\}^m$ is an* extractor *for $\mathcal{X}$ with error ε if for any $\mathbf{X} \in \mathcal{X}$,*

$$\Delta(\mathsf{Ext}(\mathbf{X}), \mathbf{U}_m) \leq \varepsilon,$$

where $\Delta(\cdot,\cdot)$ denotes statistical distance. For short, we also call Ext *an ε-extractor for $\mathcal{X}$.*

Ever since extractors were first introduced, they have found countless unexpected applications in complexity, cryptography, pseudorandomness, and theoretical computer science. For a survey, see [Vad12, Goo23].

In this paper, we will focus on extractors that output one bit.[1] In this case, the requirement in Definition 1 reduces to a requirement that for any $\mathbf{X} \in \mathcal{X}$, it holds that $|\Pr[\mathsf{Ext}(\mathbf{X}) = 1] - 1/2| \leq \varepsilon$. Or in other words,

$$|\operatorname{bias}_{\mathbf{X}}(\mathsf{Ext})| \leq 2\varepsilon.$$

Thus, returning to our original discussion, we see that getting good bounds on $|\operatorname{bias}_{\mathbf{X}}(f)|$ for a random low-degree polynomial f is *equivalent* to showing that a random low-degree polynomial is a good extractor for $\mathbf{X}$. Thus, in order to figure out interesting distributions $\mathbf{X}$ for which to pursue upper bounds on $|\operatorname{bias}_{\mathbf{X}}(f)|$, it is only natural to borrow some motivation from extractor theory. We do so, below.

Key questions

In order to show that a random low-degree polynomial extracts from a distribution $\mathbf{X} \sim \{0,1\}^n$ (equivalently, in order to upper bound $|\operatorname{bias}_{\mathbf{X}}(f)|$), it is easy to see that an absolute minimum requirement is that $\mathbf{X}$ contains some "randomness." To formalize this notion, it is standard to use *min-entropy*, defined as follows.

$$H_\infty(\mathbf{X}) := \min_{x \in \mathsf{supp}(\mathbf{X})} \log\left(\frac{1}{\Pr[\mathbf{X} = x]}\right).$$

If $\mathbf{X}$ has min-entropy $H_\infty(\mathbf{X}) \geq k$, we often refer to it as an (n,k)-source.

Unfortunately, a well-known impossibility result says that even if each source $\mathbf{X} \in \mathcal{X}$ has nearly *full* min-entropy, it is still impossible to extract [CG88].[2] Thus, in order to make extraction possible, we not only need a lower bound on the min-entropy of each $\mathbf{X} \in \mathcal{X}$, but we also need to assume that each $\mathbf{X} \in \mathcal{X}$ has some *structure*. Towards this end, the oldest trick in the book is to assume that the family $\mathcal{X}$ is not too large. In this case, since a *uniformly random function* extracts from one source $\mathbf{X}$ (with a min-entropy guarantee) with extremely high probability [Vad12, Proposition 6.12], a simple union bound allows one to conclude that there exists a single function that extracts from *all* sources $\mathbf{X} \in \mathcal{X}$.

Given the above discussion, it is natural to ask whether an analogous fundamental result can be established for uniformly random *low-degree polynomials*, which immediately raises the question,

How biased is a random degree d polynomial on a single (n,k)-source $\mathbf{X}$?

[1] As we will see, there are often relatively standard tricks that can boost the output length of an extractor, once one bit is obtained.

[2] Crucially, this is because the extractor Ext must be a single function that works *for all* $\mathbf{X} \in \mathcal{X}$.

If we can show that $|\operatorname{bias}_{\mathbf{X}}(f)|$ is low with extremely high probability over f, then we can conclude that random low-degree polynomials extract from any small family $\mathcal{X}$ of sources. Importantly, many well-studied families of sources are, in fact, very small. In particular, this is true of *all families* for which random low-degree polynomial extractors have been studied: uniform sources [BHL12], affine sources [CT15], and local sources [ACG+22].[3] Thus, showing that a random low-degree polynomial is unbiased on any single source $\mathbf{X}$ of min-entropy k could lead to a result that subsumes (and greatly generalizes) all previous work.

Unfortunately, several important families of sources $\mathcal{X}$ are quite large. For these, the above approach cannot be used to show that random low-degree polynomials extract. Here, the most general (well-studied) family is the family $\mathcal{X}$ of *sumset sources* [CL16], a model inspired by fundamental structures in additive combinatorics. Formally, an (n, k)-sumset source is defined to have the form $\mathbf{X} = \mathbf{A} + \mathbf{B}$, where $\mathbf{A}, \mathbf{B} \sim \{0,1\}^n$ are independent sources of min-entropy at least k, and $+$ denotes bitwise XOR. Sumset sources generalize a huge number of other well-studied large families [CL16, CG22, CL22], including: independent sources [CG88], interleaved sources [RY11], and small-space sources [KRVZ11] (and affine sources [BSHR+01], though this family is small). Thus, to complement our first question, we also ask:

How good is a random degree d polynomial as an extractor for the family of (n, k)-sumset sources?

In the remainder of this paper, our goal is to answer both of the questions presented above. In doing so, we hope to provide new insight into the power of a fundamental computational model (low-degree polynomials) for a fundamental computational task (randomness extraction). As it turns out, however, there are a few other reasons why answering these questions might be useful. Before we formally present our main results, we highlight some of these, below.

Pseudorandomness Low-complexity extractors have important applications in the real world and theory. In the real world, low-complexity extractors are more likely to be implemented and exhibit a reasonable running time. In theory, low-complexity extractors serve as fundamental building blocks in the construction of key cryptographic [Lu02, Vad04] and pseudorandom primitives [Li11, LZ24].[4] Their study has also led to important structural results for well-studied families of distributions [CT15, ACG+22]. Because of this, low-complexity extractors have received a lot of attention in the literature [DM02, Lu02, Vad04, Vio05, DT09, Li11, BG13, GVW15, CT15, CL18, ACG+22, HIV22, CW24, LZ24], with the works of Cohen and Tal [CT15] and Alrabiah, Chattopadhyay, Goodman, Li, and Ribeiro [ACG+22] specifically focusing on the power of *random low-degree polynomials* as extractors (for affine sources and local sources, respectively).

Coding theory Low-degree polynomials are fundamental objects in both algebra and coding theory, and studying whether they are good extractors ultimately requires proving new structural results about them (leading to new insights in these two areas). As an example, the work of Ben-Eliezer, Hod, and Lovett

[3] The family of uniform sources is the trivial family $\mathcal{X} = \{\mathbf{U}_n\}$, while the family of affine sources (of min-entropy k) consists of all $\mathbf{X} \sim \mathbb{F}_2^n$ that are uniform over a k-dimensional affine subspace of $\mathbb{F}_2^n$. Local sources, on the other hand, consist of sources of the form $\mathbf{X} = f(\mathbf{U}_\ell)$, where f is some function where each output bit depends on a bounded number of input bits.

[4] In fact, several well-known explicit extractors are *themselves* low-degree polynomials [CG88, BNS92, BSHR+01, Bou05, Dvi09, LZ19]! For example, the inner-product extractor [CG88, BSHR+01] is a degree 2 polynomial of its input, while the generalized inner-product extractor [BNS92, Dvi09, LZ19] is its natural extension to degree d. A version of Bourgain's two-source extractor [Bou05] can also be seen as a low-degree polynomial over $\mathbb{F}_3$. More precisely, the version of Bourgain's extractor $\mathsf{Bou}(x, y) = \langle (x, x^2), (y, y^2) \rangle$, where $\langle \cdot, \cdot \rangle$ denotes the inner product over $\mathbb{F}_3$, with $x, y \in \mathbb{F}_3^n$ for n prime and x^2 and y^2 computed in $\mathbb{F}_{3^n}$ (see [Dvi12, Section 3.5.2]), is a degree ≤ 4 $\mathbb{F}_3$-polynomial, since the representations in $\mathbb{F}_3^n$ of x^2 and y^2 can be computed as n-variate $\mathbb{F}_3$-polynomials of x and y of degree ≤ 2 [Kop10, Lemma 2.3.1].

[BHL12] (on low-degree extractors for uniform sources) immediately gave new bounds on the weight distribution of Reed-Muller codes. On the other hand, the results of Cohen and Tal [CT15] (on low-degree extractors for affine sources) showed that every low-degree polynomial must have a big subspace in its solution set. And the paper of Alrabiah, Chattopadhyay, Goodman, Li, and Ribeiro [ACG+22] (on low-degree extractors for local sources) proved a new type of "Chevalley-Warning theorem," which established that every (small) system of low-degree polynomial equations must have a solution with low Hamming weight.

Complexity theory Finally, low-complexity extractors can help us establish fine-grained complexity separations (as advocated for in, e.g., [HHTT23]). In more detail, extractors are known to exhibit strong lower bounds against a variety of well-studied complexity classes, including low-depth circuits [HR15, GKW21], general circuits [DK11, GK16, FGHK16, GKW21, LY22], various flavors of branching programs [HR15, CGZ22, GPT22, CL23], and more [CS16, GKW21, GG25, CG25]. Showing that there exist extractors in a low-level complexity class $\mathcal{C}$ would allow one to separate $\mathcal{C}$ from the classes above.

1.1 Our results

In this paper, we show that random low-degree polynomials extract from any small family of sources, and from the (large) family of sumset sources. This answers both questions presented in the introduction, and we present these two results (respectively) in Sections 1.1.1 and 1.1.2.

1.1.1 Low-degree polynomials extract from small families

In order to prove that random low-degree polynomials can extract from any small family, we first show that a random low-degree polynomial extracts from a single source. In particular, we prove the following, which can be viewed as both (1) a "low-degree" version of the classical fact that a random *function* extracts from a single source [Vad12, Proposition 6.12], and (2) a generalization of the result that a random low-degree polynomial has low bias [BHL12, Lemma 1.2] (Theorem 0).

Theorem 1 (Low-degree polynomials extract from a single source)**.** *For every $\delta \in (0,1)$ there is a constant $c > 0$ such that the following holds. Let $\mathbf{X} \sim \mathbb{F}_2^n$ be a source with min-entropy at least k, and let $d \in \mathbb{N}$ be an integer satisfying $1 \le d \le (1-\delta)k$. Then for a random degree d polynomial $f : \mathbb{F}_2^n \to \mathbb{F}_2$,*

$$\Pr_f\left[|\operatorname{bias}_{\mathbf{X}}(f)| > 2^{-ck/d}\right] \le 2^{-c\binom{k}{\le d}}.$$

We highlight some key aspects of this result. First, it has a simple proof, which follows by combining [BHL12, Lemma 1.2] (Theorem 0) with the leftover hash lemma [HILL99]. Second, it is easy to verify that it is tight.[5] Third, we emphasize that the above result works for *any* distribution of min-entropy at least k, not just those that are "flat" (uniform over a subset $S \subseteq \mathbb{F}_2^n$ of size 2^k). This is crucial in some applications.[6]

We also note that by a standard application of the XOR lemma [DLWZ14, Lemma 3.8], it is straightforward to extend Theorem 1 to show that a *sequence* of independent, uniformly random degree d polynomials $f_1, f_2, \ldots, f_m : \mathbb{F}_2^n \to \mathbb{F}_2$ can be concatenated to create a multi-bit extractor for $\mathbf{X}$.[7] In fact, this can further

[5] This follows by considering the (n,k)-source $\mathbf{X}$ which is uniform on the first k bits and constantly 0 on the remaining bits, combined with the tightness of Theorem 0.

[6] Even though arbitrary (n,k)-sources are known to be convex combinations of flat (n,k)-sources [Vad12, Lemma 6.10], this convex combination may end up bringing the source $\mathbf{X}$ out of the "small family" $\mathcal{X}$.

[7] In order to apply the XOR lemma, the only observation needed is that the XOR of (any number of) independent, uniformly random degree d polynomials applied to $\mathbf{X}$ is, itself, a uniformly random degree d polynomial applied to $\mathbf{X}$.

be extended to show that the sequence $f_1, f_2, \ldots, f_m$ not only extracts m uniform bits from $\mathbf{X}$, but has low correlation with any (short) fixed function g applied to $\mathbf{X}$.[8] Finally, if we set $k = n$, $m = 1$, and g to have output length 1, this result can be interpreted as nontrivial bounds on the list-size of Reed-Muller codes at the extreme (relative) radius of $1/2 - 2^{-\Omega(n/d)}$. This appears to be the first result of this form (c.f. [KLP12, ASW15, ASY20]), and naturally extends to punctured Reed-Muller codes (by picking $k < n$).

Returning to our original problem, it is straightforward to combine Theorem 1 to show our unifying result: that random low-degree polynomials extract from any small family of sources, with exponentially small error. We record this corollary below, and instantiate the general result with three interesting small families of sources: local sources, polynomial sources, and variety sources. The family of local sources are easily shown to be small, while the families of polynomial sources and variety sources were recently shown to be small via "input reduction lemmas" [CGG24, CT15].

Corollary 1 (Low-degree polynomials extract from small families)**.** *For every $\delta > 0$ there exists a constant $c > 0$ such that for all $n \geq k \geq d \in \mathbb{N}$ with $d \leq (1-\delta)k$, the following holds. For any family $\mathcal{X}$ of (n,k)-sources with size $|\mathcal{X}| < 2^{c\binom{k}{\leq d}}$, a random degree d polynomial $f : \{0,1\}^n \to \{0,1\}$ is a $2^{-ck/d}$-extractor for $\mathcal{X}$, except with probability at most $2^{-c\binom{k}{\leq d}}$.*

In particular, we obtain the following for a sufficiently large constant $C > 0$ depending only on δ.

- ***Local sources**: There exists a degree $\leq d$ polynomial $f : \{0,1\}^n \to \{0,1\}$ that is a $2^{-ck/d}$-extractor for r-local sources with min-entropy*
$$k \geq Cd(2^r n + rn \log n)^{1/d}.$$
- ***Polynomial sources**: There exists a degree $\leq d$ polynomial $f : \{0,1\}^n \to \{0,1\}$ that is a $2^{-ck/d}$-extractor for degree r polynomial sources with min-entropy*
$$k \geq C\left(\frac{C^r d^d}{r^r} \cdot n\right)^{1/(d-r)}.$$
- ***Variety sources**: There exists a degree $\leq d$ polynomial $f : \{0,1\}^n \to \{0,1\}$ that is a $2^{-ck/d}$-extractor for degree r variety sources with min-entropy*
$$k \geq Cdn^{(r+1)/d}.$$

We make a few brief remarks about this result. First, we note that our result on local sources significantly improves the parameters of the previous best result [ACG+22, Theorem 1.1], which required min-entropy $k \geq C2^r r^2 d(2^r n \log n)^{1/d}$ and had error $\varepsilon = 2^{-ck/(d^3 2^r r^2)}$. Second, we highlight that our result on polynomial sources may be surprising, as it shows that polynomials can be used to extract from polynomial sources. Perhaps this can be used to make progress on (the challenging goal of) constructing explicit extractors for polynomial sources [DGW09, CGG24], as it suggests that it is possible to improve the quality of a polynomial source *while keeping it as a polynomial source*. Third, we mention one interesting application of our result on variety sources. By exploiting known hardness results for variety extractors [LY22, GK16, GKW21], it follows that (general) circuits of size $3.1n - o(n)$ cannot compute (all) degree 3 polynomials, circuits of size $3.11n - o(n)$ cannot compute degree 4 polynomials, and circuits of size $3.9n - o(n)$ cannot compute degree 18 polynomials. Finally, we note that in Section 4.3, we show that Corollary 1 (in fact, Theorem 0) can be used to prove low-degree polynomials are good *linear seeded extractors*.[9]

[8] This can be done by first conditioning on the output of $g(\mathbf{X})$, which will only cause $\mathbf{X}$ to lose a little bit of min-entropy.

[9] Recall that a linear seeded extractor only needs to be linear on the source (fixing the seed), but may be an arbitrarily high-degree polynomial in general.

Concurrent work In a concurrent and independent work, Golovnev, Guo, Hatami, Nagargoje, and Yan [GGH+24] prove similar results to those presented above (in Section 1.1.1). In particular, they show that random low-degree polynomials are good extractors for any small family of sources, and instantiate this to obtain results similar to those presented in Corollary 1. Moreover, our proofs both rely on a similar key ingredient on the dimension of punctured Reed-Muller codes [KS05, Theorem 1.5]. The differences are as follows: our result achieves better error (exponentially small vs. polynomially small), and we also establish results for sumset sources (discussed in Section 1.1.2 below). On the other hand, the work [GGH+24] proves a significant generalization of [KS05, Theorem 1.5], in order to get interesting results in algebraic geometry.

1.1.2 Low-degree polynomials extract from sumset sources

In the second main part of our paper, we show that random low-degree polynomials are also good extractors for *sumset sources*, which are the most general well-studied *large* family of sources.

Theorem 2 (Low-degree polynomials extract from sumset sources)**.** *There exists a constant $C > 0$ such that for any $n \geq k \geq d \in \mathbb{N}$ and $\varepsilon > 0$ such that $k \geq Cd(n/\varepsilon^2)^{1/\lfloor d/2 \rfloor}$, a random degree d polynomial $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is an ε-extractor for (n,k)-sumset sources, with probability at least $1 - 2^{-\varepsilon^2 \binom{k/C}{2\lfloor d/2 \rfloor}} \geq 1 - 2^{-n^2/\varepsilon^2}$.*

We highlight a few key specializations of the above theorem – focusing on even d, for simplicity. First, we note that in the *disperser* regime,[10] Theorem 2 shows that there exist degree $\leq d$ polynomials that disperse from sumset sources with min-entropy $k = O(dn^{2/d})$. This is nearly tight, since Cohen and Tal [CT15] show that degree $\leq d$ polynomials cannot extract from *affine sources* (and thus sumset sources) with min-entropy below $k = \Omega(dn^{1/(d-1)})$. (In Section 5.3, we show the same impossibility result holds for *independent sources*, which are the other special case of sumset sources.) Second, we note that in the *polynomial error regime* $\varepsilon = n^{-\gamma}$, Theorem 2 shows that there exist degree $\leq d$ polynomials that are ε-extractors for sumset sources with min-entropy $k = O(dn^{2(1+2\gamma)/d})$. Third, in the *arbitrary error regime* $\varepsilon > 0$, Theorem 2 shows that a random degree $d = O(\log(n/\varepsilon))$ polynomial f is an ε-extractor for sumset sources with min-entropy $k = O(\log(n/\varepsilon))$. This strengthens the existential result of Mrazović [Mra16], who obtained such a min-entropy requirement for a uniformly random function f.

Finally, we highlight that our sumset extractors have interesting consequences for *linear read-once branching programs*, and our proof requires two new tools, which may be of independent interest.

Linear ROBPs In more detail, linear read-once branching programs (ROBPs) are a new type of computational model [GPT22], which simultaneously generalize both standard ROBPs and parity decision trees. At each point in the branching program, instead of querying a single input *variable*, the ROBP is allowed to query an arbitrary *linear function* of the input (so long as it is linearly independent of all previous queries). We observe that by leveraging standard results on finite fields [LN97, Lemma 6.21] (see also [BV10, Lemma 17]), linear ROBPs of constant width $w = O(1)$ can compute any polynomial of degree 2.[11] On the other hand, Theorem 2 (combined with [CL23, Theorem 1]) implies that linear ROBPs require *exponential* width $w = 2^{n-o(n)}$ to compute polynomials of degree 4. This is a huge, perhaps surprising, jump in complexity.[12]

[10] A *disperser* is an extractor $f : \mathbb{F}_2^n \to \mathbb{F}_2$ with nontrivial error $\varepsilon < 1/2$.

[11] In more detail, [BV10, Lemma 17] asserts that for any quadratic $q : \mathbb{F}_2^n \to \mathbb{F}_2$, there is some $B \in \mathbb{F}_2^{m \times n}$ with full row rank and some affine $L : \mathbb{F}_2^n \to \mathbb{F}_2$ such that $q(x) = \langle (Bx)_{\leq m/2}, (Bx)_{>m/2} \rangle + L(x)$. Since each row in B is linearly independent, the inner product can be computed using ≤ 2 bits of storage. Then, using ≤ 1 additional bit of storage (in case L is linearly dependent on the rows in B), one can simultaneously compute $L(x)$. Thus, $q(x)$ can be computed by a constant-width linear ROBP.

[12] Indeed, one might expect the width w to somehow grow proportionately with the degree d of the polynomial that must computed. However, this shows that the width jumps from constant to *exponential*, simply by moving from degree 2 to degree 4.

Sumset-punctured Reed-Muller codes Finally, we highlight that the proof of Theorem 2 requires two new key ingredients, which may be of independent interest. The first key ingredient is a new result about the structure of Reed-Muller codes that are punctured on sumsets. As a bonus application, we use this to build new "evasive sets," which we then use to improve the extractor impossibility results of Chattopadhyay, Goodman, and Gurumukhani [CGG24], and to get a low-error version of Theorem 2 for the special case of degree 4 polynomials and independent sources. Interestingly, the latter result relies on the recent breakthrough resolution of Marton's *PFR conjecture* from additive combinatorics [GGMT23].

A novel reduction between extractors The second key ingredient in the proof of Theorem 2 is a novel type of reduction between extractors. While most reductions between extractors rely on showing that a source can be equipped with structure by breaking it down into a convex combination of well-behaved distributions via a *deterministic process*, we show that doing so via a careful *(correlated) randomized process* can sometimes make this task much easier. In Appendix A, we illustrate a simpler variant of this idea in order to give an alternative proof that a uniformly random function is an extractor for sumset sources with min-entropy $k = O(\log(n/\varepsilon))$ – a result first established by Mrazović [Mra16].

Organization

The remainder of the paper is organized as follows. In Section 2, we provide an overview of our main techniques. Then, in Section 3, we record some basic preliminaries on notation, probability, and extractors. The first main technical part of this paper is Section 4, where we show that random low-degree polynomials extract from small families, and prove all results listed in Section 1.1.1. The second main technical part of the paper consists of Section 5, where we prove our main result from Section 1.1.2. In Section 6, we show that one of the key ingredients from the prior section can be used to establish the existence of "evasive sets," which are then used to obtain a few additional results. (In Appendix A, we show that the other key ingredient can be used to obtain an alternative proof of the result by Mrazović [Mra16].) Finally, we wrap up with some open problems in Section 7.

2 Overview of our techniques

2.1 Low-degree polynomials extract from small families

Recall that our goal in Theorem 1 is to obtain concentration bounds for $|\operatorname{bias}_{\mathbf{X}}(f)| = |\mathbb{E}_{x\sim\mathbf{X}}[(-1)^{f(x)}]|$ with $\mathbf{X}$ an arbitrary (n,k)-source. Our simple argument combines the original result of [BHL12] for $\mathbf{X}$ uniformly distributed over $\mathbb{F}_2^n$ (Theorem 0) and an application of the leftover hash lemma.

We first introduce some useful concepts. For a vector $x \in \mathbb{F}_2^n$, we denote by $\operatorname{eval}_d(x)$ the tuple of evaluations of all monomials of degree at most d on x, i.e.,

$$\operatorname{eval}_d(x) = \left(\prod_{i\in I} x_i\right)_{I\subseteq[n],|I|\le d} \in \mathbb{F}_2^{\binom{n}{\le d}}.$$

Given a set $S \subseteq \mathbb{F}_2^n$, we write $\operatorname{eval}_d(S) = \{\operatorname{eval}_d(x) : x \in S\}$.

In order to obtain the desired concentration, it suffices to appropriately upper bound the high-order moments $\mathbb{E}[\operatorname{bias}_{\mathbf{X}}(f)^t]$ for a large t, which is also the approach followed in [BHL12]. And, also analogously

to [BHL12], it is not hard to show that

$$\begin{aligned}\mathbb{E}_f[\mathrm{bias}_{\mathbf{X}}(f)^t] &= \Pr_{x^{(1)},\dots,x^{(t)}\sim\mathbf{X}}[\mathrm{eval}_d(x^{(1)})+\cdots+\mathrm{eval}_d(x^{(t)})=0] \\ &= \Pr_{x^{(1)},\dots,x^{(t)}\sim\mathbf{X}}[\forall p:\mathbb{F}_2^n\to\mathbb{F}_2, \deg p\le d : p(x^{(1)})+\cdots+p(x^{(t)})=0]. \qquad (1)\end{aligned}$$

Intuitively, we would like to reduce the task of bounding $\mathbb{E}_f[\mathrm{bias}_{\mathbf{X}}(f)^t]$ to the task of bounding $\mathbb{E}_f[\mathrm{bias}(f)^t]$, which can be handled via the concentration bounds from [BHL12]. We establish such a reduction via the leftover hash lemma, which guarantees the existence of a *linear* map $L:\mathbb{F}_2^n\to\mathbb{F}_2^m$ with $m\approx k$ such that $L(\mathbf{X})$ is close (in statistical distance) to the uniform distribution on $\mathbb{F}_2^m$. We claim that

$$\mathbb{E}_f[\mathrm{bias}_{\mathbf{X}}(f)^t]\le\mathbb{E}_g[\mathrm{bias}_{L(\mathbf{X})}(g)^t] \qquad (2)$$

where $g:\mathbb{F}_2^m\to\mathbb{F}_2$ is a random degree d polynomial. This holds since, by Equation (1),

$$\begin{aligned}\mathbb{E}_g[\mathrm{bias}_{L(\mathbf{X})}(g)^t] &= \Pr_{y^{(1)},\dots,y^{(t)}\sim L(\mathbf{X})}[\forall q:\mathbb{F}_2^m\to\mathbb{F}_2,\deg q\le d : q(y^{(1)})+\cdots+q(y^{(t)})=0] \\ &= \Pr_{x^{(1)},\dots,x^{(t)}\sim\mathbf{X}}[\forall q:\mathbb{F}_2^m\to\mathbb{F}_2,\deg q\le d : q(L(x^{(1)}))+\cdots+q(L(x^{(t)}))=0] \\ &\ge (1),\end{aligned}$$

where the inequality uses the fact that $q\circ L:\mathbb{F}_2^n\to\mathbb{F}_2$ has degree at most d (with $q\circ L$ denoting composition), as L is linear.

We are almost done. Informally, since $L(\mathbf{X})\approx\mathbf{U}_m$, it is easy to show that $\mathbb{E}_g[\mathrm{bias}_{L(\mathbf{X})}(g)^t]\approx\mathbb{E}_g[\mathrm{bias}(g)^t]$. Moreover, we can upper bound $\mathbb{E}_g[\mathrm{bias}(g)^t]$ appropriately via the known concentration bound from [BHL12]. Combining this with Equation (2) yields the desired upper bound on $\mathbb{E}_f[\mathrm{bias}_{\mathbf{X}}(f)^t]$, which we translate into a concentration bound on $|\mathrm{bias}_{\mathbf{X}}(f)|$ via Markov's inequality.

2.2 Low-degree polynomials extract from sumset sources

Next, we discuss the approach behind Theorem 2. For simplicity, we focus here on the case of even degree d, and note that it is trivial to extend to odd d (at a slight loss in parameters).

2.2.1 Low-degree polynomials disperse from sumset sources

As a warm-up, we first consider the simpler task of *dispersing*. In this case, we wish to show that a random degree d polynomial $f:\mathbb{F}_2^n\to\mathbb{F}_2$ will satisfy $f(\mathsf{supp}(\mathbf{W}))=\{0,1\}$ simultaneously for all (n,k)-sumset sources $\mathbf{W}$ with $k=O(dn^{2/d})$. Then, we discuss the necessary modifications to obtain sumset extraction with arbitrary error $\varepsilon>0$.

First, as usual, we only need to focus on sumset sources $\mathbf{W}=\mathbf{X}+\mathbf{Y}$ where $\mathbf{X}$ and $\mathbf{Y}$ are independent *flat* (n,k)-sources. Denote the supports of $\mathbf{X}$ and $\mathbf{Y}$ by X and Y, respectively, which have size 2^k. Then, the probability that f is identically 0 on $X+Y$ is

$$\Pr_f[f(X+Y)\equiv 0]\le 2^{-\mathrm{rank}(\mathrm{eval}_d(X+Y))}. \qquad (3)$$

This is because we may write $f(x)=\langle v,\mathrm{eval}_d(x)\rangle$ for a uniformly random vector $v\in\mathbb{F}_2^{\binom{n}{\le d}}$, and so (i) $f(x)$ is uniformly distributed over $\mathbb{F}_2$ for any fixed nonzero x, and (ii) $f(x)$ and $f(y)$ are independent (and uniformly distributed) if $\mathrm{eval}_d(x)$ and $\mathrm{eval}_d(y)$ are linearly independent.

Given Equation (3), it is clear that we must understand $\mathrm{rank}(\mathrm{eval}_d(X+Y))$. If this quantity is suitably large, then the probability that f is constant on any such sumset $X+Y$ is small, and we could hope to survive a union bound over all choices of X and Y. However, this strategy cannot directly work, because $\mathrm{rank}(\mathrm{eval}_d(X+Y))$ will be at most $\binom{n}{\leq d}$ while there are $\binom{2^n}{2^k}^2 \geq 2^{2(n-k)2^k} \gg 2^{\binom{n}{\leq d}}$ choices for X and Y.

A possible way to overcome the barrier to the application of the union bound above is to show that there exist appropriately small subsets $X' \subseteq X$ and $Y' \subseteq Y$ such that $\mathrm{rank}(\mathrm{eval}_d(X'+Y'))$ is still large. If this were the case, we could then just apply the union bound over all possible choices of the now much smaller sets X' and Y'. We can make this approach work by proving the following:

Claim 1 (Informal)**.** *Let $A, B \subseteq \mathbb{F}_2^n$ be sets of size 2^k. Then, there exist subsets $A' \subseteq A$ and $B' \subseteq B$ each of size roughly $\sqrt{\binom{k}{\leq d}}$ such that $\mathrm{rank}(\mathrm{eval}_d(A'+B'))$ is roughly $\binom{k}{\leq d}$.*

We sketch how Claim 1 can be applied to obtain the desired result. Setting $A = X$ and $B = Y$, we obtain $X' \subseteq X$ and $Y' \subseteq Y$ of size about $\sqrt{\binom{k}{\leq d}}$ such that $\mathrm{eval}_d(X'+Y')$ has rank about $\binom{k}{\leq d}$. By Equation (3), this means that the probability that f is identically 0 on $X'+Y'$, and hence on $X+Y$, is at most about $2^{-\binom{k}{\leq d}}$. But now, crucially, we only have to carry out a union bound over the at most about $\binom{2^n}{k^{d/2}}^2$ choices for X' and Y', which is approximately $2^{nk^{d/2}} \ll 2^{\binom{k}{\leq d}}$ when $k \geq Cdn^{2/d}$ for some large enough constant $C > 0$. Note that this strategy would not have worked if X' and Y' were instead of size close to $\binom{k}{\leq d}$.

Before discussing the simple proof of Claim 1, it is instructive to consider existing results of a similar flavor. Keevash and Sudakov [KS05] (and later Ben-Eliezer, Hod, and Lovett [BHL12, Lemma 1.4]) proved that for any set $S \subseteq \mathbb{F}_2^n$ of size 2^k there exists a subset $S' \subseteq S$ of size at least $\binom{k}{\leq d}$ such that $\mathrm{rank}(\mathrm{eval}_d(S')) = |S'|$. This lemma is one of the most important steps in the proof of the main result of [BHL12], and its proof hinges on the construction of an intricate "rank-preserving surjection" from S to $\mathbb{F}_2^k$. More precisely, this is a map $g : \mathbb{F}_2^n \to \mathbb{F}_2^k$ that is surjective on S and satisfies

$$\mathrm{rank}(\mathrm{eval}_d(S)) \geq \mathrm{rank}(\mathrm{eval}_d(g(S))) = \mathrm{rank}(\mathrm{eval}_d(\mathbb{F}_2^k)) = \binom{k}{\leq d}.$$

Applying this same rank-preserving surjection in the setting of Claim 1 is not guaranteed to work, because now the subset that witnesses the $\binom{k}{\leq d}$ rank lower bound must be a sumset $A+B$ with $|A|, |B| \approx \sqrt{\binom{k}{\leq d}}$. We overcome this by observing that if we are fine with worse constants, then we can replace the complicated rank-preserving surjection from [KS05, BHL12] with a *linear* map, guaranteed by the leftover hash lemma. In fact, by the leftover hash lemma, there exists a linear map $L : \mathbb{F}_2^n \to \mathbb{F}_2^{k'}$, for $k' = \Omega(k)$, which is surjective on both X and Y. We choose $X' \subseteq X$ and $Y' \subseteq Y$ such that $L(X') = L(Y') = \mathcal{B}_{d/2}^{k'}(0)$, the radius-$d/2$ Hamming ball in $\mathbb{F}_2^{k'}$ centered at 0. Then, the linearity of L plus basic properties of evaluation vectors allows us to conclude that

$$\begin{aligned}\mathrm{rank}(\mathrm{eval}_d(X'+Y')) &\geq \mathrm{rank}(\mathrm{eval}_d(L(X'+Y'))) \\ &= \mathrm{rank}(\mathrm{eval}_d(\mathcal{B}_{d/2}^{k'}(0) + \mathcal{B}_{d/2}^{k'}(0))) = \mathrm{rank}(\mathrm{eval}_d(\mathcal{B}_d^{k'}(0))),\end{aligned}$$

and it is well known that $\mathrm{rank}(\mathrm{eval}_d(\mathcal{B}_d^{k'}(0))) = \binom{k'}{\leq d}$.

2.2.2 From dispersers to extractors via random convex combinations

The argument discussed above shows that a random degree d polynomial is a k-sumset disperser for min-entropy $k = O(dn^{2/d})$ with high probability. It remains to see how we can extend this to get sumset extraction for similar min-entropy k with arbitrary error $\varepsilon > 0$.

Our first observation is that if $\mathbf{W} = \mathbf{X} + \mathbf{Y}$ is a sumset source with flat $\mathbf{X}$ and $\mathbf{Y}$ whose supports X and Y satisfy $\mathrm{rank}(\mathrm{eval}_d(X + Y)) = |X| \cdot |Y|$, then $\Pr[f(\mathbf{W}) = 0] \approx 1/2$ holds with high probability over the choice of a random degree d polynomial f. In other words, $f(\mathbf{W})$ is close (in statistical distance) to uniformly distributed over $\{0, 1\}$ with high probability over the choice of f.[13] This happens because, under the conditions above, $\mathbf{W}$ is a flat source and all the vectors in $\mathrm{eval}_d(X+Y)$ are linearly independent, which means that the bits $(f(w) = \langle v, \mathrm{eval}_d(w)\rangle)_{w \in X+Y}$ are independent and uniformly distributed when v is a uniformly random vector over $\mathbb{F}_2^{\binom{n}{\leq d}}$.

Now, set $k = Cd(n/\varepsilon^2)^{2/d}$ for a large enough constant $C > 0$, and call a sumset source $\mathbf{W} = \mathbf{X} + \mathbf{Y}$ *special* if $\mathbf{X}$ and $\mathbf{Y}$ are flat, $\mathrm{rank}(\mathrm{eval}_d(X + Y)) = |X| \cdot |Y|$, and $|X|, |Y| \approx \sqrt{\binom{k}{\leq d}}$. Combining the previous paragraph with a union bound over the choices of X and Y (analogous to the one in Section 2.2.1) shows that a random degree d polynomial will be, with high probability, an ε-extractor for the class of special sumset sources.

Of course, most sumset sources are far from special. We overcome this by showing that every (n, k)-sumset source $\mathbf{W} = \mathbf{X} + \mathbf{Y}$ with flat $\mathbf{X}$ and $\mathbf{Y}$ is $2^{-\Omega(k)}$-close to a convex combination of special sumset sources. Combining this with the observations above lets us conclude that a random degree d polynomial will be a $(k, \varepsilon' = 2^{-\Omega(k)} + \varepsilon \approx \varepsilon)$-sumset extractor with high probability.

It remains to argue why every (n, k)-sumset source $\mathbf{W} = \mathbf{X} + \mathbf{Y}$ with flat $\mathbf{X}$ and $\mathbf{Y}$ is $2^{-\Omega(k)}$-close to some convex combination of special sumset sources. To better highlight the main underlying ideas, we consider here only the particular case where $k = n$ and $\mathbf{X}$ and $\mathbf{Y}$ are uniformly distributed over $\mathbb{F}_2^k$, and present a less optimized version of our final argument. It is not hard to reduce the general case to a scenario very similar to this particular case through an application of the leftover hash lemma.

We consider an alternative way of (approximately) sampling from $\mathbf{W} = \mathbf{X} + \mathbf{Y}$ – the convex combination will be implicit in this sampling procedure. The idea is to first sample uniformly random subsets $X' \subseteq X = \mathbb{F}_2^k$ and $Y' \subseteq Y = \mathbb{F}_2^k$ each of size $\binom{k/3}{d/2}$ (which is *very roughly* $\sqrt{\binom{k}{\leq d}}$), and then sample $\mathbf{X}'$ and $\mathbf{Y}'$ uniformly at random from X' and Y', respectively. If X' and Y' are sampled independently, then it is not hard to show that $\mathbf{W}' = \mathbf{X}' + \mathbf{Y}'$ is distributed exactly like $\mathbf{W}$. However, we would like to claim that the resulting sumset source $\mathbf{X}' + \mathbf{Y}'$ will be special with high probability over the choice of subsets X' and Y'. To this end, we do not sample the subsets X' and Y' independently from each other, but rather couple the randomness used in their sampling carefully. This coupling will ensure that $\mathbf{W}' = \mathbf{X}' + \mathbf{Y}'$ is always a special sumset source for any fixing of X' and Y', while we still have $\mathbf{W}' \approx_{2^{-\Omega(k)}} \mathbf{W}$.

To sample the (*correlated*) random subsets X' and Y', we proceed as follows. Let $B_1 = \{u_1, \ldots, u_t\}$ be the set of weight-$d/2$ vectors supported on $\{1, \ldots, k/3\}$, and let $B_2 = \{v_1, \ldots, v_t\}$ be the set of weight-$d/2$ vectors supported on $\{2k/3 + 1, \ldots, k\}$. Since any two vectors u_i and v_j have disjoint supports and are non-zero, each sum $u_i + v_j$ is a distinct non-zero vector in the radius-d Hamming ball, and so

$$\mathrm{rank}\, \mathrm{eval}_d(B_1 + B_2) = |B_1| \cdot |B_2| = \binom{k/3}{d/2}^2.$$

We couple the sampling of X' and Y' by choosing a uniformly random *invertible* matrix $\mathbf{L} \in \mathbb{F}_2^{k \times k}$ and

[13] The exact shape of the "high probability" comes from a standard Chernoff bound.

setting $X' = \mathbf{L}B_1 = \{\mathbf{L}u_1, \ldots, \mathbf{L}u_t\}$ and $Y' = \mathbf{L}B_2 = \{\mathbf{L}v_1, \ldots, \mathbf{L}v_t\}$. Now, because $\mathbf{L}$ is invertible, we know that

$$\operatorname{rank}\operatorname{eval}_d(X' + Y') = \operatorname{rank}\operatorname{eval}_d(B_1 + B_2) = |B_1| \cdot |B_2| = |X'| \cdot |Y'|.$$

Therefore, if $\mathbf{X}'$ and $\mathbf{Y}'$ are sampled independently and uniformly at random from X' and Y', respectively, then $\mathbf{W}' = \mathbf{X}' + \mathbf{Y}'$ is a special sumset source, as desired.

However, because the choices of X' and Y' are now correlated, we still need to argue that this overall sampling process produces something statistically close to $\mathbf{W} = \mathbf{X} + \mathbf{Y}$, with $\mathbf{X}$ and $\mathbf{Y}$ independent and uniformly distributed over $\mathbb{F}_2^k$. If $\mathbf{L} \in \mathbb{F}_2^{k \times k}$ was a uniformly random matrix, then this would be immediate. Indeed, let I and J be the random indices associated to the choices of $\mathbf{X}'$ and $\mathbf{Y}'$ from X' and Y'. Then, $\mathbf{X}' = \mathbf{L}u_I$ and $\mathbf{Y}' = \mathbf{L}v_J$ would be independent and uniformly distributed over $\mathbb{F}_2^k$, since u_I and v_J are linearly independent for all choices of I and J. To argue that this is still approximately true when $\mathbf{L}$ is required to be invertible, we use the fact that the supports of u_I and v_J lie in a subset of $2k/3$ coordinates. Therefore, it suffices to focus on $2k/3$ columns of $\mathbf{L}$. Since a collection of $2k/3$ uniformly random vectors over $\mathbb{F}_2^k$ will be linearly independent except with probability $2^{-\Omega(k)}$, we conclude that any collection of $2k/3$ columns of $\mathbf{L}$ will be $2^{-\Omega(k)}$-close in statistical distance to a collection of $2k/3$ uniformly random vectors. This gives that $(\mathbf{X}', \mathbf{Y}') \approx_{2^{-\Omega(k)}} (\mathbf{X}, \mathbf{Y})$, where $\mathbf{X}$ and $\mathbf{Y}$ are independent, and so $\mathbf{W}' = \mathbf{X}' + \mathbf{Y}'$ is $2^{-\Omega(k)}$-close to the true sumset $\mathbf{W} = \mathbf{X} + \mathbf{Y}$.

3 Preliminaries

Notation We denote random variables by boldfaced uppercase letters such as $\mathbf{X}$ and $\mathbf{Y}$ and denote sets by uppercase letters such as A and B or, at times, by calligraphic uppercase letters. In this work we focus on random variables supported on finite sets, and write $\mathsf{supp}(\mathbf{X})$ for the support of the random variable $\mathbf{X}$. We denote the uniform distribution over $\mathbb{F}_2^m$ by $\mathbf{U}_m$, and we write $\log$ for the base-2 logarithm. We use $\mathsf{wt}(x)$ to denote the Hamming weight of a vector $x \in \mathbb{F}_2^n$, and we let $\mathcal{B}_r^n(v)$ denote the Hamming ball in $\mathbb{F}_2^n$ that is centered at v and has radius r. Finally, we define $\binom{n}{\leq r} = \sum_{i=0}^{r} \binom{n}{i}$.

Binomial coefficients We will need the following fact about binomial coefficients.

Fact 1 ([BHL12, Proposition 8])**.** *For any $\beta, \delta \in (0,1)$, there exists a constant $\gamma > 0$ such that the following holds for all $n \in \mathbb{N}$. If $1 \leq d \leq \delta n$, $t \geq d$, and $t \geq (1 - \gamma/d)n$, then it holds that $\binom{t}{\leq d} \geq \beta\binom{n}{\leq d}$.*

3.1 Probability

In this section we collect some basic notions from probability theory that will be useful throughout.

Definition 2 (Statistical distance)**.** *The statistical distance between discrete random variables $\mathbf{X}$ and $\mathbf{Y}$ supported on S, denoted $\Delta(\mathbf{X}, \mathbf{Y})$, is given by*

$$\Delta(\mathbf{X}, \mathbf{Y}) = \max_{T \subseteq S} |\Pr[\mathbf{X} \in T] - \Pr[\mathbf{Y} \in T]| = \frac{1}{2} \sum_{x \in S} |\Pr[\mathbf{X} = x] - \Pr[\mathbf{Y} = x]| .$$

*We say that $\mathbf{X}$ and $\mathbf{Y}$ are ε-*close*, and write $\mathbf{X} \approx_\varepsilon \mathbf{Y}$, if $\Delta(\mathbf{X}, \mathbf{Y}) \leq \varepsilon$.*

We will heavily exploit the following standard result about statistical distance.

Fact 2 (Data-processing inequality)**.** *For any random variables $\mathbf{X}, \mathbf{Y} \sim V$ and function $f : V \to W$,*

$$\Delta(\mathbf{X}, \mathbf{Y}) \geq \Delta(f(\mathbf{X}), f(\mathbf{Y})).$$

Definition 3 (Min-entropy)**.** *The* min-entropy *of a random variable $\mathbf{X}$ is defined as*

$$H_\infty(\mathbf{X}) := \min_{x \in \mathsf{supp}(\mathbf{X})} \log\left(\frac{1}{\Pr[\mathbf{X} = x]}\right).$$

Definition 4 ((n,k)-source)**.** *We say that $\mathbf{X} \sim \mathbb{F}_2^n$ is an (n,k)-*source *if $H_\infty(\mathbf{X}) \geq k$.*

Definition 5 ((n,k)-sumset source)**.** *We say that $\mathbf{W} \sim \mathbb{F}_2^n$ is an (n,k)-*sumset source *if there exist independent (n,k)-sources $\mathbf{X}, \mathbf{Y} \sim \mathbb{F}_2^n$ such that $\mathbf{W} = \mathbf{X} + \mathbf{Y}$.*

We also use the following version of the Chernoff bound, which can be found in [Vad12, Theorem 2.21].

Lemma 1 (Chernoff bound)**.** *Let $\mathbf{X}_1, \ldots, \mathbf{X}_t$ be independent random variables supported on $[0,1]$, and let $\mathbf{X} = \frac{1}{t}\sum_{i=1}^t \mathbf{X}_i$. Then,*

$$\Pr[|\mathbf{X} - \mathbb{E}[\mathbf{X}]| > \varepsilon] \leq 2 \cdot e^{-\varepsilon^2 t/4}.$$

Finally, we will need the following result from [CG25]. We include its short proof, for completeness.

Definition 6. *We say that $g : Y \to X$ is a* pseudoinverse *of $f : X \to Y$ if $f(g(f(x))) = f(x)$ for all $x \in X$.*

Lemma 2 (Dependency reversal [CG25])**.** *For any random variable $\mathbf{X} \sim X$ and deterministic function $f : X \to Y$, there exists an independent random variable $\mathbf{A} \sim A$ and deterministic function $g : Y \times A \to X$ such that*

$$g(f(\mathbf{X}), \mathbf{A}) \equiv \mathbf{X},$$

and such that $g(\cdot, a)$ is a pseudoinverse of f, for all $a \in A$.

Proof. We define an independent random variable $\mathbf{A} \sim A := X^Y$ as a sequence of independent random variables $\mathbf{A} = (\mathbf{A}_y)_{y \in Y}$ where each $\mathbf{A}_y \sim X$ is defined as

$$\Pr[\mathbf{A}_y = x] := \Pr[\mathbf{X} = x \mid f(\mathbf{X}) = y] \text{ for all } x \in X.$$

If we then define the deterministic function $g : Y \times A \to V$ as

$$g(y, a) := a_y,$$

it directly follows that $g(f(\mathbf{X}), \mathbf{A}) \equiv \mathbf{X}$ via the law of total probability, as desired. □

3.2 Extractors

This section collects additional basic definitions of randomness extractors and useful auxiliary results.

Definition 7 (Strong seeded extractor)**.** *A function $\mathsf{Ext} : \mathbb{F}_2^n \times \mathbb{F}_2^s \to \mathbb{F}_2^m$ is said to be a (k,ε)-*strong seeded extractor *if for every (n,k)-source $\mathbf{X}$ it holds that*

$$\mathsf{Ext}(\mathbf{X}, \mathbf{U}_s), \mathbf{U}_s \approx_\varepsilon \mathbf{U}_m, \mathbf{U}_s,$$

where $\mathbf{U}_s$ is independent of $\mathbf{X}$ and $\mathbf{U}_m$. Moreover, we say that Ext is linear *if $\mathsf{Ext}(\cdot, y)$ is a linear function for all $y \in \mathbb{F}_2^s$.*

Lemma 3 (Leftover Hash Lemma [HILL99])**.** *For every $0 < k < n$, $\varepsilon > 0$, and $m \leq k - 2\log(1/\varepsilon)$ there exists an explicit linear (k, ε)-strong seeded extractor $\mathsf{Ext} : \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{F}_2^m$.*

We will use the following simple corollary of Lemma 3.

Lemma 4. *For every $n > k > 100$ there exists a family of linear maps $L : \mathbb{F}_2^n \to \mathbb{F}_2^{k/100}$ such that for every (n, k)-source $\mathbf{X}$ we have that at least a 0.8-fraction of linear maps in the family are surjective on $\mathsf{supp}(\mathbf{X})$.*

Proof. We apply the leftover hash lemma with $\varepsilon = 2^{-k/5}$ and $m = k/100 \leq k - 2\log(1/\varepsilon) = 3k/5$. Fix any flat (n, k)-source $\mathbf{X}$. By an averaging argument, we have $L(\mathbf{X}) \approx_{10\varepsilon} \mathbf{U}_1$ for more than a 0.8-fraction of linear maps L in the family given by the leftover hash lemma. Since $10\varepsilon < 2^{-k/100} = 2^{-m}$ when $k > 100$, we conclude that every such map L is surjective on $\mathsf{supp}(\mathbf{X})$. □

The lemma below is a special case of a much more general result of Dhar and Dvir on leftover hashing with ℓ_∞ guarantees [DD22, Theorem II.4].

Lemma 5 ([DD22, Theorem II.4, special case])**.** *There exists a constant $C > 0$ such that for all $n \geq k \geq 2$, the following holds. Fix any set $S \subseteq \mathbb{F}_2^n$ of size 2^k. Then at least 0.99 fraction all linear maps $L : \mathbb{F}_2^n \to \mathbb{F}_2^m$ with output length $m = \lfloor k - C\log k \rfloor$ are surjective when restricted to S.*

We now define various special types of extractors and dispersers.

Definition 8 (Two-source disperser)**.** *A function $\mathsf{Disp} : \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{F}_2$ is said to be a k-*two-source disperser *if for any two independent (n, k)-sources $\mathbf{X}$ and $\mathbf{Y}$ it holds that $\mathsf{supp}(\mathsf{Disp}(\mathbf{X}, \mathbf{Y})) = \{0, 1\}$.*

Definition 9 (Two-source extractor)**.** *A function $\mathsf{Ext} : \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{F}_2^m$ is said to be a (k, ε)-*two-source extractor *if for any two independent (n, k)-sources $\mathbf{X}$ and $\mathbf{Y}$ it holds that*

$$\mathsf{Ext}(\mathbf{X}, \mathbf{Y}) \approx_\varepsilon \mathbf{U}_m.$$

Definition 10 (Sumset disperser)**.** *A function $\mathsf{Disp} : \mathbb{F}_2^n \to \mathbb{F}_2$ is said to be a k-*sumset disperser *if for any (n, k)-sumset source $\mathbf{W}$ it holds that $\mathsf{supp}(\mathsf{Disp}(\mathbf{W})) = \{0, 1\}$.*

Definition 11 (Sumset extractor)**.** *A function $\mathsf{Ext} : \mathbb{F}_2^n \to \mathbb{F}_2^m$ is said to be a (k, ε)-*sumset extractor *if for any (n, k)-sumset source $\mathbf{W}$ it holds that*

$$\mathsf{Ext}(\mathbf{W}) \approx_\varepsilon \mathbf{U}_m.$$

Note that every k-sumset disperser Disp induces a k-two-source disperser Disp' by setting $\mathsf{Disp}'(x, y) = \mathsf{Disp}(x + y)$. The same holds in the setting of extractors.

4 Low-degree polynomials extract from small families

4.1 Low-degree polynomials extract from a single source

In this section we prove Theorem 1, which we restate here.

Theorem 1 (Low-degree polynomials extract from a single source)**.** *For every $\delta \in (0, 1)$ there is a constant $c > 0$ such that the following holds. Let $\mathbf{X} \sim \mathbb{F}_2^n$ be a source with min-entropy at least k, and let $d \in \mathbb{N}$ be an integer satisfying $1 \leq d \leq (1 - \delta)k$. Then for a random degree d polynomial $f : \mathbb{F}_2^n \to \mathbb{F}_2$,*

$$\Pr_f \left[|\operatorname{bias}_{\mathbf{X}}(f)| > 2^{-ck/d} \right] \leq 2^{-c\binom{k}{\leq d}}.$$

Recall that for a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ and a random variable $\mathbf{X} \sim \mathbb{F}_2^n$ we define the bias of f with respect to $\mathbf{X}$ as

$$\mathrm{bias}_{\mathbf{X}}(f) = \Pr_{x\sim\mathbf{X}}[f(x) = 0] - \Pr_{x\sim\mathbf{X}}[f(x) = 1] = \mathbb{E}_{x\sim\mathbf{X}}\left[(-1)^{f(x)}\right].$$

When $\mathbf{X}$ is uniformly distributed over a set $X \subseteq \mathbb{F}_2^n$, we may write $\mathrm{bias}_{\mathbf{X}}(f) = \mathrm{bias}_X(f)$, and when $X = \mathbb{F}_2^n$ we simply write $\mathrm{bias}(f)$ for $\mathrm{bias}_X(f)$. We will need the following characterization about the moments of the bias, whose statement and proof are analogous to those of [BHL12, Claim 2.1] for the special case of a uniform input distribution.

Lemma 6 (Simple extension of [BHL12, Claim 2.1])**.** *If $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is a random degree d polynomial, it holds that*

$$\mathbb{E}_f\left[\mathrm{bias}_{\mathbf{X}}(f)^t\right] = \Pr_{x^{(1)},\ldots,x^{(t)}\sim\mathbf{X}}[\forall p : \mathbb{F}_2^n \to \mathbb{F}_2, \deg(p) \le d : p(x^{(1)}) + \cdots + p(x^{(t)}) = 0]$$

for any random variable $\mathbf{X} \sim \mathbb{F}_2^n$.

Proof. The argument follows the same lines as that behind [BHL12, Claim 2.1]. Observe that the lemma statement follows immediately if we show that

$$\mathbb{E}_f[\mathrm{bias}_{\mathbf{X}}(f)^t] = \Pr_{x^{(1)},\ldots,x^{(t)}\sim\mathbf{X}}\left[\mathrm{eval}_d(x^{(1)}) + \cdots + \mathrm{eval}_d(x^{(t)}) = 0\right].$$

We begin by writing f as

$$f(x) = \sum_{I\subseteq[n]:|I|\le d} \alpha_I x_I,$$

where the α_I are independent and uniformly distributed over $\mathbb{F}_2$ and $x_I = \prod_{i\in I} x_i$. Then, we have that

$$\begin{aligned}
\mathbb{E}_f[\mathrm{bias}_{\mathbf{X}}(f)^t] &= \mathbb{E}_{\{\alpha_I\}}\left[\prod_{j=1}^{t} \mathbb{E}_{x^{(j)}\sim\mathbf{X}}\left[(-1)^{\sum_{I\subseteq[n]:|I|\le d}\alpha_I x_I^{(j)}}\right]\right] \\
&= \mathbb{E}_{\{\alpha_I\}}\left[\mathbb{E}_{x^{(1)},\ldots,x^{(t)}\sim\mathbf{X}}\left[\prod_{j=1}^{t}(-1)^{\sum_{I\subseteq[n]:|I|\le d}\alpha_I x_I^{(j)}}\right]\right] \\
&= \mathbb{E}_{x^{(1)},\ldots,x^{(t)}\sim\mathbf{X}}\left[\mathbb{E}_{\{\alpha_I\}}\left[(-1)^{\sum_{I\subseteq[n]:|I|\le d}\alpha_I\left(\sum_{j=1}^{t} x_I^{(j)}\right)}\right]\right] \\
&= \mathbb{E}_{x^{(1)},\ldots,x^{(t)}\sim\mathbf{X}}\left[\prod_{I\subseteq[n]:|I|\le d}\mathbb{E}_{\alpha_I}\left[(-1)^{\alpha_I\left(\sum_{j=1}^{t} x_I^{(j)}\right)}\right]\right] \\
&= \mathbb{E}_{x^{(1)},\ldots,x^{(t)}\sim\mathbf{X}}\left[\prod_{I\subseteq[n]:|I|\le d}\mathbf{1}_{\left\{x_I^{(1)}+\cdots+x_I^{(t)}=0\right\}}\right] \\
&= \mathbb{E}_{x^{(1)},\ldots,x^{(t)}\sim\mathbf{X}}\left[\mathbf{1}_{\left\{\mathrm{eval}_d(x^{(1)})+\cdots+\mathrm{eval}_d(x^{(t)})=0\right\}}\right] \\
&= \Pr_{x^{(1)},\ldots,x^{(t)}\sim\mathbf{X}}\left[\mathrm{eval}_d(x^{(1)}) + \cdots + \mathrm{eval}_d(x^{(t)}) = 0\right],
\end{aligned}$$

as desired. □

We are now ready to prove Theorem 1.

Proof of Theorem 1. The proof follows along the same lines as the proof of [BHL12, Lemma 1.2], combined with a linear hashing trick. For an integer $t > 0$, we focus on bounding the t-th moment of $\mathrm{bias}_{\mathbf{X}}(f)$. By Lemma 6, we have that

$$\mathbb{E}_f\left[\mathrm{bias}_{\mathbf{X}}(f)^t\right] = \Pr_{x^{(1)},\ldots,x^{(t)}\sim\mathbf{X}}[\forall p : \mathbb{F}_2^n \to \mathbb{F}_2, \deg(p) \le d : p(x^{(1)}) + \cdots + p(x^{(t)}) = 0].$$

Fix any linear map $L : \mathbb{F}_2^n \to \mathbb{F}_2^m$. Observe that

$$\begin{aligned}\Pr_{x^{(1)},\ldots,x^{(t)}\sim\mathbf{X}}&[\forall p : \mathbb{F}_2^n \to \mathbb{F}_2, \deg(p) \le d : p(x^{(1)}) + \cdots + p(x^{(t)}) = 0] \\ &\le \Pr_{x^{(1)},\ldots,x^{(t)}\sim\mathbf{X}}[\forall p : \mathbb{F}_2^m \to \mathbb{F}_2, \deg(p) \le d : p(L(x^{(1)})) + \cdots + p(L(x^{(t)})) = 0]. \quad (4)\end{aligned}$$

This holds since $\deg(p \circ L) \le \deg(p) \le d$, where $p \circ L$ denotes the composition of the polynomial p and the linear map L. Then, Lemma 6 yields

$$\begin{aligned}&\Pr_{x^{(1)},\ldots,x^{(t)}\sim\mathbf{X}}[\forall p : \mathbb{F}_2^m \to \mathbb{F}_2, \deg(p) \le d : p(L(x^{(1)})) + \cdots + p(L(x^{(t)})) = 0] \\ &= \Pr_{w^{(1)},\ldots,w^{(t)}\sim L(\mathbf{X})}[\forall p : \mathbb{F}_2^m \to \mathbb{F}_2, \deg(p) \le d : p(w^{(1)}) + \cdots + p(w^{(t)}) = 0] \\ &= \mathbb{E}_g\left[\mathrm{bias}_{L(\mathbf{X})}(g)^t\right],\end{aligned}$$

where the expectation is taken over the choice of a random degree d polynomial $g : \mathbb{F}_2^m \to \mathbb{F}_2$. Therefore, we conclude that

$$\mathbb{E}_f\left[\mathrm{bias}_{\mathbf{X}}(f)^t\right] \le \mathbb{E}_g\left[\mathrm{bias}_{L(\mathbf{X})}(g)^t\right] \quad (5)$$

for all linear maps L.

Let $c > 0$ be the absolute constant from Theorem 0. Without loss of generality, we enforce that $c < \min(1/4, \delta/2)$ (if this does not hold for the choice of c from Theorem 0, take a smaller c). Since $H_\infty(\mathbf{X}) = k$, the leftover hash lemma (Lemma 3) guarantees the existence of a linear map $L : \mathbb{F}_2^n \to \mathbb{F}_2^m$ with $m = k(1 - 2c/d)$ such that

$$L(\mathbf{X}) \approx_{2^{-ck/d}} \mathbf{U}_m. \quad (6)$$

Note that, by our choice of c, we have that $m \ge (1-\delta)k \ge d$. Equation (6) implies that

$$\Pr_g\left[|\,\mathrm{bias}_{L(\mathbf{X})}(g)| > 2^{-\frac{ck}{2d}+1}\right] \le \Pr_g\left[|\,\mathrm{bias}(g)| > 2^{-\frac{ck}{2d}+1} - 2^{-\frac{ck}{d}}\right] \le 2^{-c\binom{m}{\le d}}. \quad (7)$$

The last inequality follows from Theorem 0 applied to $\mathbb{F}_2^m$ because

$$2^{-\frac{ck}{2d}+1} - 2^{-\frac{ck}{d}} \ge 2^{-\frac{ck}{2d}} \ge 2^{-cm/d},$$

since $m = k(1 - 2c/d) \ge k/2$ as we enforced that $c < 1/4$. Since

$$\binom{m}{\le d} = \binom{k(1-2c/d)}{\le d} \ge \alpha\binom{k}{\le d}$$

for some constant $\alpha > 0$ depending only on c and δ, we get from Equation (7) that

$$\Pr_g\left[|\,\mathrm{bias}_{L(\mathbf{X})}(g)| > 2^{-c_1 k/d}\right] \le 2^{-c_1\binom{k}{\le d}} \quad (8)$$

for some absolute constant $c_1 > 0$.

We now proceed to bound the t-th moment $\mathbb{E}_f\left[\mathrm{bias}_{\mathbf{X}}(f)^t\right]$ for an appropriate t. Set $t = c_2 \cdot \frac{d}{k} \cdot \binom{k}{\leq d}$ for a sufficiently large constant $c_2 > 0$ depending only on c_1. Using Equations (5) and (8), we have that

$$\begin{aligned}\mathbb{E}_f\left[\mathrm{bias}_{\mathbf{X}}(f)^t\right] &\leq \mathbb{E}_g\left[\mathrm{bias}_{L(\mathbf{X})}(g)^t\right] \\ &\leq 2^{-c_1 kt/d} + 2^{-c_1\binom{k}{\leq d}} \\ &\leq 2^{-c_3\binom{k}{\leq d}},\end{aligned}$$

where the last equality uses our choice of t and holds for a sufficiently small constant $c_3 > 0$. Let $c_4 = c_3/2$. Combining the bound above with Markov's inequality, we conclude that

$$\Pr_f\left[|\,\mathrm{bias}_{\mathbf{X}}(f)| > 2^{-c_4 k/d}\right] \leq \frac{\mathbb{E}_f\left[\mathrm{bias}_{\mathbf{X}}(f)^t\right]}{2^{-c_4 tk/d}} \leq 2^{-c_4\binom{k}{\leq d}},$$

which yields the desired lemma statement with absolute constant $c_4 > 0$. □

4.2 Low-degree polynomials extract from small families

We showcase applications of Theorem 1 to some important small families of sources.

Local sources

First, we consider the scenario of locally-samplable sources [DW12, Vio14, ACG+22]. A source $\mathbf{X} \sim \mathbb{F}_2^n$ is said to be *r-local* if $\mathbf{X} = g(\mathbf{U}_m)$, where $g : \mathbb{F}_2^m \to \mathbb{F}_2^n$ is some function such that each output bit depends on at most r input bits and $\mathbf{U}_m$ denotes the uniform distribution over $\mathbb{F}_2^m$. We can take $m \leq nr$ without loss of generality, in which case there are at most

$$\left(\binom{nr}{r}2^{2^r}\right)^n \leq (en)^{rn} \cdot 2^{2^r \cdot n} \leq 2^{2rn\log n} \cdot 2^{2^r \cdot n} = 2^{n(2r\log n + 2^r)}$$

r-local sources of length n, provided that $n \geq 4$. If $k = Cdn^{1/d}(r\log n + 2^r)^{1/d}$, we have that

$$\binom{k}{\leq d} \geq \binom{k}{d} \geq (k/d)^d = C^d n(r\log n + 2^r).$$

Let c be the absolute constant from Theorem 1. If we take C to be large enough depending only on c so that $c \cdot C^d n(r\log n + 2^r) > n(2r\log n + 2^r)$, we conclude that there are fewer than $2^{c\binom{k}{\leq d}}$ sources in the family. Combining a union bound over these sources with Theorem 1 immediately yields the following result.

Corollary 2 (Low-degree polynomials are good local source extractors)**.** *For every $\delta \in (0,1)$ there exist constants $c, C > 0$ such that the following holds for all n: For $k \geq Cdn^{1/d}(r\log n + 2^r)^{1/d}$ and $d \leq (1-\delta)k$, a random degree d polynomial is an $(\varepsilon = 2^{-ck/d})$-extractor for the family of length-n r-local sources of min-entropy k with probability at least $1 - 2^{-c\binom{k}{\leq d}}$ over the choice of f.*

Corollary 2 both improves on and simplifies the proof of [ACG+22, Theorem 1.1], which required min-entropy $k \geq C2^r r^2 d(2^r n\log n)^{1/d}$ and error $\varepsilon = 2^{-\frac{ck}{d^3 2^r r^2}}$ for some absolute constants $c, C > 0$ and an intricate initial reduction to local non-oblivious bit-fixing sources. In particular, observe that when $d > r$, we now get that the min-entropy is $O(d(n\log n)^{1/d})$. Previously, this only happened when $d > 2^r$.

It is also instructive to compare our improved result with the lower bound from [ACG+22, Theorem 1.2], which states that no degree-d polynomial extracts from length-n r-local sources of min-entropy $k = cd(rn\log n)^{1/d}$ for some absolute constant $c > 0$. For example, when the locality satisfies $r < \log\log n$, Corollary 2 is optimal up to the constant factor.

Polynomial sources

A random variable $\mathbf{X} \sim \mathbb{F}_2^n$ is a degree-r polynomial source if there exist $\mathbb{F}_2$-polynomials $p_1, \ldots, p_n : \mathbb{F}_2^m \to \mathbb{F}_2$ of degree at most r for some positive integer m (the *input length*) such that $\mathbf{X} = P(\mathbf{U}_m)$, where $P = (p_1, \ldots, p_n)$ and $\mathbf{U}_m$ is uniform over $\mathbb{F}_2^m$. This is a very challenging model to extract from, and several papers have attempted to do so [DGW09, BG13, GVJZ23, CGG24]. In the most recent work, Chattopadhyay, Goodman, and Gurumukhani [CGG24] established the following input reduction lemma.

Lemma 7 (Input reduction lemma for polynomial sources [CGG24, Lemma 1])**.** *There exists a constant $\alpha > 0$ such that the following holds for all positive integers n, k, r such that $k \le n$: Suppose that $\mathbf{X} \sim \mathbb{F}_2^n$ is a degree-r polynomial source with min-entropy k. Then, $\mathbf{X} \approx_{2^{-k}} \mathbf{Y}$, where $\mathbf{Y} \sim \mathbb{F}_2^n$ is a convex combination of degree-r polynomial sources with min-entropy $k' = k - 1$ and input length $m = \alpha(k-1)$.*

Given Lemma 7, in order to show that a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is an $(\varepsilon + 2^{-k})$-extractor for degree-r polynomial sources with min-entropy k it suffices to show that f is an ε-extractor for degree-r polynomial sources with min-entropy $k-1$ and input length $m = \alpha(k-1)$. There are at most

$$2^{\binom{m}{\le r} n} \le 2^{\left(\frac{em}{r}\right)^r n} = 2^{\left(\frac{\beta(k-1)}{r}\right)^r n}$$

such sources, where $\beta = e \cdot \alpha$ and we have used the inequality $\binom{m}{\le r} \le \left(\frac{e\cdot m}{r}\right)^r$ valid for all $m \ge 1$ and $1 \le r \le m$. Set $k = C\left(\frac{C^r d^d n}{r^r}\right)^{\frac{1}{d-r}}$. Then, if we take C to be a sufficiently large constant depending on β and the constant $c > 0$ guaranteed by Theorem 1, we get that

$$2^{\left(\frac{\beta(k-1)}{r}\right)^r n} < 2^{c\left(\frac{k-1}{d}\right)^d} \le 2^{c\binom{k-1}{\le d}}.$$

Therefore, we can apply Theorem 1 to the family of degree-r polynomial sources with min-entropy $k-1$ and input length $m = \alpha(k-1)$ to obtain the following result.

Corollary 3 (Low-degree polynomials extract from polynomial sources)**.** *For every $\delta \in (0,1)$ there exist constants $c, C > 0$ such that the following holds for all integers $n \ge 2$: For $k \ge 2$ satisfying $k \ge C\left(\frac{C^r d^d n}{r^r}\right)^{\frac{1}{d-r}}$ and $d \le (1-\delta)k$, a random polynomial of degree at most d is an $(\varepsilon = 2^{-ck/d})$-extractor for the family of length-n degree-r polynomial sources of min-entropy k with probability at least $1 - 2^{-c\binom{k}{\le d}}$ over the choice of f.*

In particular, if we take $d = 2r$ then the above holds for any $k \ge Cdn^{2/d}$ for a sufficiently large constant $C > 0$.

Variety sources

A random variable $\mathbf{X} \sim \mathbb{F}_2^n$ is a variety source of degree r and min-entropy k if for some t there exist polynomials $p_1, \ldots, p_t : \mathbb{F}_2^n \to \mathbb{F}_2$ of degree at most r such that $\mathbf{X}$ is uniformly distributed over the variety

$$V(p_1, \ldots, p_t) = \{x \in \mathbb{F}_2^n \mid p_1(x) = p_2(x) = \cdots = p_t(x) = 0\}.$$

This family of sources has received significant interest over the last decade, both over $\mathbb{F}_2$ and larger fields [Dvi09, Rem16, LZ19, GVJZ23]. Sufficiently strong explicit extractors for variety sources are known to imply breakthrough circuit lower bounds [GKW21]. In order to bound the number of variety sources we need to handle, we use the following structural result implicit in [CT15, Theorem 5.1], which states that we can assume the variety is defined by $\leq n+1$ polynomials. We provide the short proof for completeness.

Lemma 8 (Input reduction lemma for variety sources [CT15, Theorem 5.1])**.** *For any variety $V(p_1,\ldots,p_t)$ of degree r, there exists a variety $V(q_1,\ldots,q_\ell)$ of degree r such that $V(p_1,\ldots,p_t) = V(q_1,\ldots,q_\ell)$ and $\ell = n+1$.*

Proof. To define the polynomials q_i, first suppose that we sample coefficients α_{ij} uniformly at random from $\mathbb{F}_2$, and set $q_i(x) = \sum_{j=1}^{t} \alpha_{ij} p_j(x)$. Then, all the q_i's have degree at most r and $V(p_1,\ldots,p_t) \subseteq V(q_1,\ldots,q_\ell)$. Moreover, for each $x \notin V(p_1,\ldots,p_t)$ we have that $\Pr[q_i(x) = 0 \text{ for all } i \in [\ell]] = 2^{-\ell} = 2^{-(n+1)}$. This means the expected number of x's such that $x \in V(q_1,\ldots,q_\ell) \setminus V(p_1,\ldots,p_t)$ is at most $2^n \cdot 2^{-(n+1)} = 1/2$. Thus, there is a fixing of the coefficients (α_{ij}) such that $V(q_1,\ldots,q_\ell) = V(p_1,\ldots,p_t)$. □

With the help of Lemma 8, we conclude that the number of variety sources of degree r and min-entropy k is at most $2^{(n+1)\binom{n}{\leq r}} \leq 2^{(n+1)^{r+1}}$. If $k = Cd(n+1)^{\frac{r+1}{d}}$, we have that

$$\binom{k}{\leq d} \geq \binom{k}{d} \geq (k/d)^d = C^d (n+1)^{r+1}.$$

Let $c > 0$ be the absolute constant from Theorem 1. Then, if we take $C > 0$ to be sufficiently large depending only on c so that $c \cdot C^d n^{r+1} > (n+1)^{r+1}$, we get that the family of variety sources of degree r and min-entropy k has size smaller than $2^{c\binom{k}{\leq d}}$, and so combining Theorem 1 and a union bound immediately yields the following result.

Corollary 4 (Low-degree polynomials extract from variety sources)**.** *For every $\delta \in (0,1)$ there exist constants $c, C > 0$ such that the following holds for all n: For $k \geq Cdn^{\frac{r+1}{d}}$ and $d \leq (1-\delta)k$, a random polynomial of degree $\leq d$ is an $(\varepsilon = 2^{-ck/d})$-extractor for the family of length-n variety sources of degree r and min-entropy k with probability at least $1 - 2^{-c\binom{k}{\leq d}}$ over the choice of f.*

The best explicit extractors for variety sources over $\mathbb{F}_2$ either work for constant degree r and min-entropy $k = (1-c_r)n$ [LZ19], or large degree $r = n^\alpha$ and very high min-entropy $n - n^\beta$ with $\alpha + \beta < 1/2$ [Rem16]. By [GKW21], an explicit version of Corollary 4 would be more than enough to imply significantly improved circuit lower bounds.

4.3 Low-degree polynomials are good linear seeded extractors

In this section we study the parameters of strong linear seeded extractors computable by low-degree polynomials. Such objects are crucial in the construction of several affine extractors [Li11, LZ24]. We start with the following basic definitions.

Definition 12 (Almost balanced codes)**.** *We say that a string $x \in \mathbb{F}_2^n$ is ε-balanced if it has Hamming weight $\frac{1-\varepsilon}{2}n \leq \mathsf{wt}(x) \leq \frac{1+\varepsilon}{2}n$. We say that a code $C \subseteq \mathbb{F}_2^n$ is ε-balanced[14] if all nonzero $x \in C$ are ε-balanced. We say that a code $C \subseteq \mathbb{F}_2^n$ is δ-almost ε-balanced if $\Pr_{x \sim C}[x \text{ is not } \varepsilon\text{-balanced}] \leq \delta$.*

[14] Another standard name for this notion is "ε-biased code."

Definition 13 (Reed-Muller codes)**.** *We define the Reed-Muller code* $\mathsf{RM}(n,d)$ *as*

$$\mathsf{RM}(n,d) = \{(f(x))_{x\in\mathbb{F}_2^n} \mid f : \mathbb{F}_2^n \to \mathbb{F}_2, \deg f \le d\}.$$

We will need to import the following results. The first corollary about balancedness of $\mathsf{RM}(n,d)$ follows directly from [BHL12, Lemma 1.2] (Theorem 0), which corresponds to setting $k = n$ in Theorem 1.

Corollary 5. *For every* $\gamma \in (0,1)$ *there exists a constant* $c > 0$ *such that the Reed-Muller code* $\mathsf{RM}(n,d)$ *with* $d \le (1-\gamma)n$ *is* δ*-almost* ε*-balanced, where* $\delta = 2^{-c\binom{n}{\le d}}$ *and* $\varepsilon = 2^{-cn/d}$.

Lemma 9 (Johnson bound [GRS23, Section 7.3], adapted)**.** *Every linear* ε*-balanced code of length* T *has list-size at most* $2T$ *at radius* $\frac{1-\sqrt{\varepsilon}}{2}$.[15]

The following well-known lemma connects the list-decodability of a linear code and its performance as a seeded extractor. See [Sha11] for a nice exposition of this result.

Lemma 10 (Extractors from codes [Tre01])**.** *Let* $G \in \mathbb{F}_2^{T\times n}$ *be the generator matrix of a linear code with list size* $\le L$ *at radius* $\le \frac{1-\varepsilon}{2}$*, and let* $t = \log T$*. Then the function* $\mathsf{Ext}_G : \mathbb{F}_2^n \times \mathbb{F}_2^t \to \mathbb{F}_2$ *defined as* $\mathsf{Ext}_G(x,y) := (Gx)_y = \langle x, G_y\rangle$ *is a* (k,ε)*-strong seeded extractor for min-entropy* $k = \log L + \log(1/\varepsilon) + 1$.

Using these, we are ready to show that low-degree polynomials can function as strong linear seeded extractors. Our key tool is the following.

Lemma 11 (Random subcodes of almost-balanced codes are balanced)**.** *Let* $G \in \mathbb{F}_2^{n\times k}$ *be the generator matrix of a* δ*-almost* ε*-balanced code. Then, for a uniformly random* $H \in \mathbb{F}_2^{k\times\ell}$ *we have that*

$$\Pr_H[GH \text{ is not the generator matrix of an } \varepsilon\text{-balanced code}] \le \delta\cdot(2^\ell - 1).$$

Proof. When we refer to a matrix G as a code, we are referring to the code generated by G. We have

$$\begin{aligned}
\Pr_H[GH \text{ is not an } \varepsilon\text{-balanced code}] &= \Pr_H[\exists x \neq 0 : (GH)x \text{ is not } \varepsilon\text{-balanced}] \\
&= \Pr_H\left[\bigvee_{x\neq 0} (GH)x \text{ is not } \varepsilon\text{-balanced}\right] \\
&\le \sum_{x\neq 0} \Pr_H[(GH)x \text{ is not } \varepsilon\text{-balanced}] \\
&= \sum_{x\neq 0} \Pr_{y\sim\mathbb{F}_2^k}[Gy \text{ is not } \varepsilon\text{-balanced}] \\
&\le (2^\ell - 1)\cdot\delta.
\end{aligned}$$

□

Theorem 3 (Strong seeded extractors from random subcodes of linear almost-balanced codes)**.** *Let* $G \in \mathbb{F}_2^{T\times r}$ *be the generator matrix of a* δ*-almost* ε*-balanced code and let* $t = \log T$*. For any matrix* $H \in \mathbb{F}_2^{r\times n}$*, define the function* $\mathsf{Ext}_{GH} : \mathbb{F}_2^n \times \mathbb{F}_2^t \to \mathbb{F}_2$ *as* $\mathsf{Ext}_{GH}(x,y) := (GHx)_y$*. Then for* $k := \log(2T) + \log(1/\sqrt{\varepsilon}) + 1$ *and a uniformly random* H*, it holds that*

$$\Pr_H[\mathsf{Ext}_{GH} \text{ is not a } (k,\sqrt{\varepsilon})\text{-strong seeded extractor}] \le \delta\cdot 2^n.$$

[15] A code $C \subseteq \mathbb{F}_q^n$ has list-size at most L at radius ρ if $|\{c \in C : \mathsf{wt}(c-x) \le \rho n\}| \le L$ for all vectors $x \in \mathbb{F}_q^n$.

Proof. By Lemma 11, we know that GH is the generator of an ε-balanced code, except with probability at most $\delta \cdot 2^n$. By the Johnson bound (Lemma 9), this implies that GH is the generator of a code with list size at most $2T$ at radius at most $\frac{1-\sqrt{\varepsilon}}{2}$, except with probability at most $\delta \cdot 2^n$. This immediately implies via Lemma 10 that Ext_{GH} is a $(k, \sqrt{\varepsilon})$-strong seeded extractor, except with probability at most $\delta \cdot 2^n$. □

Using this, we obtain the following.

Theorem 4 (Strong seeded extractors from random subcodes of Reed-Muller codes)**.** *For any $\gamma > 0$ there exists a constant $c = c(\gamma) > 0$ such that for any $d \leq (1-\gamma)n$, the following holds. Let $G \in \mathbb{F}_2^{T \times \binom{t}{\leq d}}$ be a generator matrix of the Reed-Muller code $\mathsf{RM}(t, d)$. For any matrix $H \in \mathbb{F}_2^{\binom{t}{\leq d} \times n}$, define the function $\mathsf{Ext}_{GH} : \mathbb{F}_2^n \times \mathbb{F}_2^t \to \mathbb{F}_2$ as $\mathsf{Ext}_{GH}(x, y) := (GHx)_y$. Then for $k := \log(2T) + \log(1/\sqrt{\varepsilon}) + 1$ and $\varepsilon := 2^{-ct/d}$ and $\delta := 2^{-c\binom{t}{\leq d}}$ and a uniformly random H, it holds that*

$$\Pr_H \left[\mathsf{Ext}_{GH} \text{ is not a } (k, \sqrt{\varepsilon})\text{-strong seeded extractor} \right] \leq \delta \cdot 2^n.$$

Furthermore, for any H it holds that Ext_{GH} has left-degree at most 1 and right-degree at most d, i.e., any monomial in Ext_{GH} contains at most one variable from the left source and at most d variables from the right source.

Proof. The claim that Ext_{GH} is a strong seeded extractor follows immediately by combining the fact that the Reed-Muller code $\mathsf{RM}(t, d)$ is linear and δ-almost ε-balanced (Corollary 5) with the fact that random subcodes of linear δ-almost ε-balanced codes give strong seeded extractors (Theorem 3). All that remains is to show that Ext_{GH} will have left-degree at most 1 and right-degree at most d for any H. To see why, recall that, since $G \in \mathbb{F}_2^{T \times \binom{t}{\leq d}}$ is a generator of $\mathsf{RM}(t, d)$, there exist degree $\leq d$ polynomials $p_1, \ldots, p_{\binom{t}{\leq d}} : \mathbb{F}_2^t \to \mathbb{F}_2$ whose truth tables make up the columns of G. Thus, if we let $L_i : \mathbb{F}_2^n \to \mathbb{F}_2$ define the linear function $L_i(x) := \langle H_i, x \rangle$ for each row H_i of H, we can write $\mathsf{Ext}(x, y)$ as

$$\mathsf{Ext}_{GH}(x, y) := (GHx)_y = \sum_{i \in \binom{t}{\leq d}} p_i(y) \cdot L_i(x),$$

and thus it has left-degree at most 1 and right-degree at most d, as desired. □

This immediately gives the following.

Theorem 5. *For any $\gamma \in (0, 1)$ there exist constants $c, C > 0$ such that the following holds. For any $d \leq (1-\gamma)n$ and $k \geq Cdn^{1/(d-1)}$, there exists a linear (k, ε)-strong seeded extractor $\mathsf{Ext} : \mathbb{F}_2^n \times \mathbb{F}_2^k \to \mathbb{F}_2$ of degree $\leq d$ with error $\varepsilon = 2^{-ck/d}$.*

Proof. Set $t = k = O(dn^{1/(d-1)})$ in Theorem 4. □

5 Low-degree polynomials extract from sumset sources

5.1 Low-degree polynomials disperse from sumset sources

As a warmup, we prove a disperser version of Theorem 2 (with a slightly better bound on the probability). For simplicity, we focus here on the case of even degree d.

Theorem 6. *There exist constants $C, c > 0$ such that for any $n \geq k \geq d \in \mathbb{N}$ (with d even) satisfying $d \leq \frac{c \log n}{\log \log n}$ and $k \geq Cdn^{2/d}$, a random degree d polynomial $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is a disperser for (n, k)-sumset sources, with probability at least $1 - 2^{-c\binom{k}{\leq d}}$.*

Our key lemma (recall the informal Claim 1) states that for any two large subsets A and B we can find appropriately small subsets $A' \subseteq A$ and $B' \subseteq B$ such that $\mathrm{rank}(\mathrm{eval}_d(A' + B'))$ is large.

Lemma 12. *There exists a constant $C > 0$ such that for all $n \geq k \geq d \in \mathbb{N}$ (with d even) satisfying $k \geq C(1 + \log n)$, the following holds. Let $A, B \subseteq \mathbb{F}_2^n$ be sets of size 2^k. Then, there exist subsets $A' \subseteq A$ and $B' \subseteq B$ such that $|A'|, |B'| = \binom{k - C \log n}{\leq d/2}$ and $\mathrm{rank}(\mathrm{eval}_d(A' + B')) \geq \binom{k - C \log n}{\leq d}$.*

Before we proceed to the (simple) proof of Lemma 12, we note some basic properties of evaluation vectors. First, it is well known that $\mathrm{rank}(\mathrm{eval}_d(\mathcal{B}_d^n(0))) = \binom{n}{\leq d}$, as $\mathcal{B}_d^n(0)$ is an interpolating set for degree $\leq d$ polynomials (see, e.g., [O'D14, Proposition 6.21]). Moreover, we have the following property.

Claim 2. *If $L : \mathbb{F}_2^n \to \mathbb{F}_2^m$ is a linear map, then*

$$\mathrm{rank}(\mathrm{eval}_d(S)) \geq \mathrm{rank}(\mathrm{eval}_d(L(S)))$$

for every set $S \subseteq \mathbb{F}_2^n$.

Proof. Suppose that $x_1, \ldots, x_r \in S$ are distinct vectors such that the associated evaluation vectors $\mathrm{eval}_d(x_1), \ldots, \mathrm{eval}_d(x_r)$ satisfy $\sum_{i=1}^r \mathrm{eval}_d(x_i) = 0$. Equivalently, for every polynomial f of degree at most d it holds that

$$\sum_{i=1}^r f(x_i) = 0.$$

It suffices to show that the vectors $\mathrm{eval}_d(L(x_1)), \ldots, \mathrm{eval}_d(L(x_r))$ also satisfy

$$\sum_{i=1}^r \mathrm{eval}_d(L(x_i)) = 0. \tag{9}$$

Fix a set $U \subseteq [m]$ such that $|U| \leq d$. Note that $L(x)^U = \prod_{j \in U} L(x)_j$ can be written as $f_{L,U}(x)$ for some polynomial $f_{L,U}$ of degree at most $|U| \leq d$ which depends only on L and U. Therefore, we have

$$\sum_{i=1}^r L(x_i)^U = \sum_{i=1}^r f_{L,U}(x_i) = 0,$$

which establishes Equation (9). □

By combining the above property with the leftover hash lemma, we can now prove Lemma 12.

Proof of Lemma 12. Fix arbitrary sets $A, B \subseteq \mathbb{F}_2^n$ of size 2^k. By Lemma 5 and a union bound, there exists a linear map $L : \mathbb{F}_2^n \to \mathbb{F}_2^t$ with output length $t = k - C(1 + \log n)$ that is surjective on both A and B. In particular, there are subsets $A' \subseteq A$ and $B' \subseteq B$ of size $\binom{t}{\leq d/2}$ such that $L(A'), L(B') = \mathcal{B}_{d/2}^t(0)$. Since L is linear, we have that $L(A' + B') = L(A') + L(B') = \mathcal{B}_d^t(0)$, and so

$$\mathrm{rank}(\mathrm{eval}_d(A' + B')) \geq \mathrm{rank}(\mathrm{eval}_d(L(A' + B'))) = \mathrm{rank}(\mathrm{eval}_d(\mathcal{B}_d^t(0))) = \binom{t}{\leq d},$$

where we have invoked Claim 2. □

Next, since it will be useful in Section 6, we state here an alternative version of Lemma 12, which is more helpful when k is very small. Its proof follows in exactly the same manner as that of Lemma 12, with the exception that we invoke Lemma 4 instead of Lemma 5.

Lemma 13. *For all $n \geq k \geq d \in \mathbb{N}$ (with d even) satisfying $k > 100$, the following holds. Let $A, B \subseteq \mathbb{F}_2^n$ be sets of size 2^k. Then, there exist subsets $A' \subseteq A$ and $B' \subseteq B$ such that $|A'|, |B'| = \binom{k/100}{\leq d/2}$ and $\operatorname{rank}(\operatorname{eval}_d(A' + B')) \geq \binom{k/100}{\leq d}$.*

We now use Lemma 12 to prove Theorem 6.

Proof of Theorem 6. Without loss of generality, we may assume that $\mathbf{X}$ and $\mathbf{Y}$ are uniformly distributed over sets $A, B \subseteq \mathbb{F}_2^n$, respectively, each of size 2^k. By Lemma 12, there exist subsets $A' \subseteq A$ and $B' \subseteq B$ of size $\binom{t}{\leq d/2}$ such that

$$\operatorname{rank}(\operatorname{eval}_d(A' + B')) \geq \binom{t}{\leq d},$$

with $t = k - C_0(1 + \log n)$ for an absolute constant $C_0 > 0$. Since $A' + B' \subseteq A + B$, it follows that f is constant on $A + B$ with probability at most

$$2 \cdot 2^{-\operatorname{rank}(\operatorname{eval}_d(A'+B'))} \leq 2^{-\binom{t}{\leq d}+1}.$$

By taking a union bound over all the at most

$$\binom{2^n}{\binom{t}{\leq d/2}}^2 \leq 2^{2n\binom{t}{\leq d/2}}$$

choices of A' and B', we conclude that the probability that f is constant on some set $A + B$ is at most

$$2^{2n\binom{t}{\leq d/2}} \cdot 2^{-\binom{t}{\leq d}+1} \leq 2^{2n\binom{k}{\leq d/2}+1} \cdot 2^{-\binom{t}{\leq d}}. \tag{10}$$

We will now invoke Fact 1 with $\beta = \delta = 1/2$, k in place of n, and $t = k - C_0(1 + \log n)$. Let $\gamma > 0$ be the absolute constant correspoding to $\beta = \delta = 1/2$ in Fact 1. By hypothesis, we have $k \geq Cdn^{2/d}$ and $d \leq \frac{c \log n}{\log \log n} < k/2 = \delta k$ for $C > 0$ an appropriately large constant and $c > 0$ an appropriately small constant. In particular, this means that we can enforce $0 < c < 1$ and choose $C > 0$ large enough depending on γ and C_0 so that $k \geq Cdn^{\frac{2 \log \log n}{\log n}} \geq Cd \log n$ satisfies $t = k - C_0(1 + \log n) \geq d$ and $t \geq (1 - \gamma/d)k$. By Fact 1, we then get that $\binom{t}{\leq d} \geq \frac{1}{2}\binom{k}{\leq d}$. Plugging this into Equation (10), we conclude that the failure probability is at most

$$2^{2n\binom{k}{\leq d/2}+1} \cdot 2^{-\frac{1}{2}\binom{k}{\leq d}}. \tag{11}$$

We also have

$$2n\binom{k}{\leq d/2} \leq 2dn\left(\frac{2ek}{d}\right)^{d/2} \leq \frac{1}{4}(k/d)^d \leq \frac{1}{4}\binom{k}{\leq d},$$

where the second inequality uses that $k \geq Cdn^{2/d}$ for a sufficiently large constant $C > 0$. Plugging this into Equation (11), we conclude that the probability that f is constant on $A + B$ is at most

$$2^{2n\binom{k}{\leq d/2}+1} \cdot 2^{-\frac{1}{2}\binom{k}{\leq d}} \leq 2^{-\frac{1}{4}\binom{k}{\leq d}+1}.$$

□

5.2 Low-degree polynomials extract from sumset sources

In this section, we prove Theorem 2 in full generality. We begin with the following definition of a special type of sumset sources, for which it will be easy to show that random low-degree polynomials extract.

Definition 14. *We say that a sumset source $\mathbf{W} = \mathbf{X} + \mathbf{Y}$ has* full eval_d-rank *if the set $\mathrm{eval}_d(\mathsf{supp}(\mathbf{X}) + \mathsf{supp}(\mathbf{Y}))$ is a collection of $|\mathsf{supp}(\mathbf{X})| \cdot |\mathsf{supp}(\mathbf{Y})|$ linearly independent vectors.*

We proceed with a simple proof that random low-degree polynomials extract from sumset sources with full eval-rank. Then, for the main part of the proof, we show that every sumset source is close to a convex combination of sumset sources with full eval-rank.

Lemma 14. *For any $t, n \in \mathbb{N}$ and $\varepsilon > 0$ such that $t \geq 64n/\varepsilon^2$, the following holds with probability at least $1 - 2 \cdot 2^{-\varepsilon^2 t^2/32}$ over the selection of a random degree d polynomial $f : \mathbb{F}_2^n \to \mathbb{F}_2$. For any sets $X, Y \subseteq \mathbb{F}_2^n$ of size t satisfying $\mathrm{rank}(\mathrm{eval}_d(X + Y)) = t^2$, we have that*

$$|\operatorname{bias}_{\mathbf{X}+\mathbf{Y}}(f)| \leq \varepsilon,$$

where $\mathbf{X}$ and $\mathbf{Y}$ are uniformly distributed over X and Y, respectively.

Proof. Fix any sets $X, Y \subseteq \mathbb{F}_2^n$ of size t satisfying $\mathrm{rank}(\mathrm{eval}_d(X + Y)) = t^2$. By this rank condition, we know that the t^2 random variables $\{f(x+y)\}_{(x,y)\in X\times Y}$ are independent and uniform over $\mathbb{F}_2$. Thus, if we define for every $(x, y) \in X \times Y$ the indicator random variable $\mathbf{Z}_{x,y} := 1[f(x + y) = 0]$, then

$$\operatorname{bias}_{\mathbf{X}+\mathbf{Y}}(f) = 2\left(\frac{1}{t^2}\sum_{x,y}\mathbf{Z}_{x,y} - \frac{1}{2}\right) = 2\left(\frac{1}{t^2}\sum_{x,y}\mathbf{Z}_{x,y} - \mathbb{E}\left[\frac{1}{t^2}\sum_{x,y}\mathbf{Z}_{x,y}\right]\right),$$

and a direct application of the Chernoff bound (Lemma 1) yields

$$\Pr_f\left[|\operatorname{bias}_{\mathbf{X}+\mathbf{Y}}| > \varepsilon\right] \leq 2e^{-\varepsilon^2 t^2/16}.$$

Finally, by taking a union bound over all the at most $\binom{2^n}{t}^2 \leq 2^{2nt}$ choices of X and Y, we conclude that the probability that $|\operatorname{bias}_{\mathbf{X}+\mathbf{Y}}(f)| > \varepsilon$ for some sets X, Y is at most

$$2^{2nt} \cdot 2e^{-\varepsilon^2 t^2/16},$$

which is at most $2 \cdot 2^{-\varepsilon^2 t^2/32}$ when $t \geq 64n/\varepsilon^2$, as desired. □

Next, we show every sumset source is close to a convex combination of sumset sources with full eval_d-rank.

Lemma 15. *There exists a constant $c > 0$ such that for all $n \geq k \geq d \in \mathbb{N}$, the following holds. Let $\mathbf{W} = \mathbf{X} + \mathbf{Y}$ be an (n, k)-sumset source. Then $\mathbf{W}$ is 2^{-ck}-close to a convex combination of flat sumset sources $\mathbf{W}^\star = \mathbf{X}^\star + \mathbf{Y}^\star$ with full eval_d-rank and such that $|\mathsf{supp}(\mathbf{X}^\star)| = |\mathsf{supp}(\mathbf{Y}^\star)| = \binom{k/6}{\lfloor d/2 \rfloor}$.*

Proof. Let $\mathsf{Ext} : \mathbb{F}_2^n \to \mathbb{F}_2^m$ be a linear map such that $m = k/2$ and

$$\mathsf{Ext}(\mathbf{X}) \approx_{2\varepsilon} \mathbf{U}_m \tag{12}$$
$$\mathsf{Ext}(\mathbf{Y}) \approx_{2\varepsilon} \mathbf{U}_m. \tag{13}$$

This map is guaranteed to exist by the leftover hash lemma (Lemma 3) with $\varepsilon = 2^{-k/4}$ and a union bound. Using the dependency reversal lemma (Lemma 2), there exist functions Ext_X^{-1} and Ext_Y^{-1} and random variables $\mathbf{A} = (\mathbf{A}_0, \ldots, \mathbf{A}_t)$ and $\mathbf{B} = (\mathbf{B}_0, \ldots, \mathbf{B}_t)$ such that

$$\begin{aligned}\mathbf{X} &\equiv \mathsf{Pick}\left(\mathsf{Ext}_X^{-1}(\mathsf{Ext}(\mathbf{X}), \mathbf{A}_0), \ldots, \mathsf{Ext}_X^{-1}(\mathsf{Ext}(\mathbf{X}), \mathbf{A}_t); \mathbf{U}\right)\\ \mathbf{Y} &\equiv \mathsf{Pick}\left(\mathsf{Ext}_Y^{-1}(\mathsf{Ext}(\mathbf{Y}), \mathbf{B}_0), \ldots, \mathsf{Ext}_Y^{-1}(\mathsf{Ext}(\mathbf{Y}), \mathbf{B}_t); \mathbf{U}'\right),\end{aligned}$$

where $\mathbf{A}_0, \ldots, \mathbf{A}_t, \mathbf{B}_0, \ldots, \mathbf{B}_t, \mathbf{U}, \mathbf{U}', \mathbf{X}, \mathbf{Y}$ are mutually independent, $\mathbf{U}, \mathbf{U}' \sim \{0, \ldots, t\}$ and Pick uses its last argument to pick among its first $t+1$ arguments (i.e., $\mathsf{Pick}(v_0, v_1, \ldots, v_t; i)$ outputs v_i). To see why the above is true, we consider an arbitrary fixing of $\mathbf{U}, \mathbf{U}'$ and simply apply the dependency reversal lemma.

Now, by Equation (13) and the independence of $\mathbf{X}$ and $\mathbf{Y}$, there are random variables $\mathbf{R}, \mathbf{R}' \sim \mathbb{F}_2^m$ independent of each other and the rest that are both uniformly random over $\mathbb{F}_2^m$ and such that we can replace $\mathsf{Ext}(\mathbf{X}), \mathsf{Ext}(\mathbf{Y})$ with them. Recalling that $\mathbf{X}, \mathbf{Y}$ are independent, we get from an application of the data-processing inequality and Equation (13) that

$$\begin{aligned}&\mathsf{Pick}\left(\mathsf{Ext}_X^{-1}(\mathsf{Ext}(\mathbf{X}), \mathbf{A}_0), \ldots, \mathsf{Ext}_X^{-1}(\mathsf{Ext}(\mathbf{X}), \mathbf{A}_t); \mathbf{U}\right) \approx_{2\varepsilon} \mathsf{Pick}\left(\mathsf{Ext}_X^{-1}(\mathbf{R}, \mathbf{A}_0), \ldots, \mathsf{Ext}_X^{-1}(\mathbf{R}, \mathbf{A}_t); \mathbf{U}\right),\\ &\mathsf{Pick}\left(\mathsf{Ext}_Y^{-1}(\mathsf{Ext}(\mathbf{Y}), \mathbf{B}_0), \ldots, \mathsf{Ext}_Y^{-1}(\mathsf{Ext}(\mathbf{Y}), \mathbf{B}_t); \mathbf{U}'\right) \approx_{2\varepsilon} \mathsf{Pick}\left(\mathsf{Ext}_Y^{-1}(\mathbf{R}', \mathbf{B}_0), \ldots, \mathsf{Ext}_Y^{-1}(\mathbf{R}', \mathbf{B}_t); \mathbf{U}'\right).\end{aligned} \tag{14}$$

We now couple the randomness of $\mathbf{R}, \mathbf{R}'$ in a specific way. Let $\mathbf{L} \sim \mathbb{F}_2^{m \times m}$ be a uniformly random invertible matrix ($\mathbf{L}$ is obtained by sampling its i-th column uniformly at random from $\mathbb{F}_2^m$ conditioned on it being linearly independent of the previous $i-1$ columns). Intuitively, we take appropriate disjoint subsets $\mathcal{B}_0$ and $\mathcal{B}_1$ of the radius-$d/2$ Hamming ball, and replace $\mathbf{R}$ and $\mathbf{R}'$ by applications of $\mathbf{L}$ to vectors in these sets. More precisely, consider the sets $\mathcal{B}_0, \mathcal{B}_1 \subseteq \mathbb{F}_2^m$ defined as

$$\begin{aligned}\mathcal{B}_0 &= \{u \in \mathbb{F}_2^m : \mathsf{wt}(u) = \lfloor d/2 \rfloor, \mathsf{supp}(u) \subseteq \{1, \ldots, m/3\}\},\\ \mathcal{B}_1 &= \{v \in \mathbb{F}_2^m : \mathsf{wt}(v) = \lfloor d/2 \rfloor, \mathsf{supp}(v) \subseteq \{2m/3+1, \ldots, m\}\}.\end{aligned}$$

Note that vectors in $\mathcal{B}_0$ and $\mathcal{B}_1$ are nonzero and have disjoint supports. Moreover, $\mathcal{B}_0 + \mathcal{B}_1$ is a subset of the radius-d Hamming ball, and so $\mathrm{rank}(\mathrm{eval}_d(\mathcal{B}_0 + \mathcal{B}_1)) = |\mathcal{B}_0| \cdot |\mathcal{B}_1|$.

Let $u_0, \ldots, u_t$ be the elements in $\mathcal{B}_0$, and let $v_0, \ldots, v_t$ be the elements in $\mathcal{B}_1$. We argue that

$$\begin{aligned}&\left(\mathsf{Pick}\left(\mathsf{Ext}_X^{-1}(\mathbf{R}, \mathbf{A}_0), \ldots, \mathsf{Ext}_X^{-1}(\mathbf{R}, \mathbf{A}_t); \mathbf{U}\right), \mathsf{Pick}\left(\mathsf{Ext}_Y^{-1}(\mathbf{R}', \mathbf{B}_0), \ldots, \mathsf{Ext}_Y^{-1}(\mathbf{R}', \mathbf{B}_t); \mathbf{U}'\right)\right)\\ &\approx_{m2^{-m/3}} \left(\mathsf{Pick}\left(\mathsf{Ext}_X^{-1}(\mathbf{L}u_0, \mathbf{A}_0), \ldots, \mathsf{Ext}_X^{-1}(\mathbf{L}u_t, \mathbf{A}_t); \mathbf{U}\right), \mathsf{Pick}\left(\mathsf{Ext}_Y^{-1}(\mathbf{L}v_0, \mathbf{B}_0), \ldots, \mathsf{Ext}_Y^{-1}(\mathbf{L}v_t, \mathbf{B}_t); \mathbf{U}'\right)\right)\\ &:= (\mathbf{X}^*, \mathbf{Y}^*).\end{aligned} \tag{15}$$

To see why this holds, consider an arbitrary fixing of $(\mathbf{U}, \mathbf{U}') = (i, j)$. Then, by an application of the data-processing inequality, it suffices to show that $(\mathbf{L}u_i, \mathbf{L}v_j) \approx_{m2^{-m/3}} (\mathbf{R}, \mathbf{R}')$. Towards this end, recall that for any $i, j \in [t]$, the vectors u_i and v_j are nonzero with disjoint supports of size $m/3$ each. Let $\mathbf{L}'$ denote the $m \times (2m/3)$ matrix obtained by selecting columns of $\mathbf{L}$ indexed by the supports of u_i and v_j. Then, we have that $\mathbf{L}' \approx_{m2^{-m/3}} \mathbf{M}'$, where $\mathbf{M}'$ is a uniformly random $m \times (2m/3)$ matrix. To see this, note that a uniformly random vector in $\mathbb{F}_2^m$ will be linearly independent from any given collection of $2m/3$ vectors with probability at least $1 - 2^{-m/3}$, and then apply a union bound over all the $2m/3 < m$ columns of $\mathbf{L}'$. Therefore, letting u_i' and v_j' denote the restrictions of u_i and v_j to the coordinates in $\mathsf{supp}(u_i) \cup \mathsf{supp}(v_j)$, we

have that $(\mathbf{L}u_i, \mathbf{L}v_j) \approx_{m2^{-m/3}} (\mathbf{M}'u'_i, \mathbf{M}'v'_j) \equiv (\mathbf{R}, \mathbf{R}')$. The last step holds because u'_i and v'_j are linearly independent, so the random variables $\mathbf{M}'u'_i$ and $\mathbf{M}'v'_j$ are independent and uniformly distributed over $\mathbb{F}_2^m$.

We now analyze the eval_d-rank of $\mathbf{X}^* + \mathbf{Y}^*$. Consider any fixing of the random variables $\mathbf{L}$ and $\mathbf{A}_0, \ldots, \mathbf{A}_t, \mathbf{B}_0, \ldots, \mathbf{B}_t$. Upon such a fixing, $(\mathbf{X}^*, \mathbf{Y}^*)$ becomes of the form $(\mathbf{X}^\star, \mathbf{Y}^\star)$, where $\mathbf{X}^\star, \mathbf{Y}^\star$ are independent and uniform over the sets

$$X^\star := \{\mathsf{Ext}_X^{-1}(Lu_0, a_0), \ldots, \mathsf{Ext}_X^{-1}(Lu_t, a_t)\},$$
$$Y^\star := \{\mathsf{Ext}_Y^{-1}(Lv_0, b_0), \ldots, \mathsf{Ext}_Y^{-1}(Lv_t, b_t)\},$$

respectively. Then, notice that the support of $\mathbf{X}^\star + \mathbf{Y}^\star$ is exactly

$$S^\star := \{\mathsf{Ext}_X^{-1}(Lu_i, a_i) + \mathsf{Ext}_Y^{-1}(Lv_j, b_j)\}_{i,j\in[t]}.$$

To analyze the eval_d-rank of $S^\star$, recall from Claim 2 that applying linear transformations can only decrease the eval_d-rank. Furthermore, note that for any $i, j \in [t]$ it holds that

$$L^{-1}(\mathsf{Ext}(\mathsf{Ext}_X^{-1}(Lu_i, a_i) + \mathsf{Ext}_Y^{-1}(Lv_j, b_j))) = u_i + v_j.$$

Furthermore, the composition $L^{-1} \circ \mathsf{Ext}$ is linear, since L was linear (and invertible) and Ext is also linear (since it comes from the leftover hash lemma). Thus,

$$\mathrm{rank}(\mathrm{eval}_d(S^\star)) \geq \mathrm{rank}(\mathrm{eval}_d(L^{-1}(\mathsf{Ext}(S^\star)))) = \mathrm{rank}(\mathrm{eval}_d(\{u_i + v_j\}_{i,j})) = \mathrm{rank}(\mathrm{eval}_d(\mathcal{B}_0 + \mathcal{B}_1)).$$

Since $|S^\star| \leq |\mathcal{B}_0| \cdot |\mathcal{B}_1|$, we get that $\mathrm{rank}(\mathrm{eval}_d(S^\star)) = |\mathcal{B}_0| \cdot |\mathcal{B}_1|$, and so $\mathbf{X}^\star + \mathbf{Y}^\star$ has full eval_d-rank. Recalling Equations (14) and (15) and the parameter settings $m = k/2$ and $\varepsilon = 2^{-k/4}$, this means that the sumset source $\mathbf{X} + \mathbf{Y}$ is $\varepsilon^\star$-close to a convex combination of flat sumset sources $\mathbf{X}^\star + \mathbf{Y}^\star$ with full eval_d-rank and support sizes $|\mathsf{supp}(\mathbf{X}^\star)|, |\mathsf{supp}(\mathbf{Y}^\star)| = \binom{m/3}{\lfloor d/2 \rfloor} = \binom{k/6}{\lfloor d/2 \rfloor}$, where $\varepsilon^\star = 4\varepsilon + m2^{-m/3} \leq 4 \cdot 2^{-k/4} + (k/2) \cdot 2^{-k/6} \leq 5k \cdot 2^{-k/6}$. This is at most $2^{-k/7}$ as long as k exceeds a big enough constant C. Since the lemma is straightforward to obtain whenever $k \leq C$,[16] this completes the proof. □

Finally, we show how to combine Lemma 14 and Lemma 15 to get Theorem 2, restated here for convenience.

Theorem 2 (Low-degree polynomials extract from sumset sources)**.** *There exists a constant $C > 0$ such that for any $n \geq k \geq d \in \mathbb{N}$ and $\varepsilon > 0$ such that $k \geq Cd(n/\varepsilon^2)^{1/\lfloor d/2 \rfloor}$, a random degree d polynomial $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is an ε-extractor for (n, k)-sumset sources, with probability at least $1 - 2^{-\varepsilon^2 \binom{k/C}{2\lfloor d/2 \rfloor}} \geq 1 - 2^{-n^2/\varepsilon^2}$.*

Proof. Fix any $n \geq k \geq d \in \mathbb{N}$ and $\varepsilon > 0$ such that $k \geq Cd(n/\varepsilon^2)^{1/\lfloor d/2 \rfloor}$, where C is a sufficiently large constant,[17] and define $\varepsilon_0 := \varepsilon/2$. By Lemma 15, every (n, k)-sumset source $\mathbf{W} = \mathbf{X} + \mathbf{Y}$ is $2^{-c_0 k}$-close to a convex combination of sumset sources $\mathbf{W}^\star = \mathbf{X}^\star + \mathbf{Y}^\star$ with full eval_d-rank and

$$t := |\mathsf{supp}(\mathbf{X}^\star)| = |\mathsf{supp}(\mathbf{Y}^\star)| = \binom{k/6}{\lfloor d/2 \rfloor},$$

where $c_0 > 0$ is some constant. By the lower bound on k and the fact that C is a sufficiently large constant,

$$t = \binom{k/6}{\lfloor d/2 \rfloor} \geq \left(\frac{k/6}{\lfloor d/2 \rfloor}\right)^{\lfloor d/2 \rfloor} \geq (C/3)^{\lfloor d/2 \rfloor} n/\varepsilon^2 = (C/3)^{\lfloor d/2 \rfloor} n/(4\varepsilon_0^2) \geq 64n/\varepsilon_0^2.$$

[16] This is because $1 - 2^{-2k} \leq 2^{-ck}$ whenever $k \leq C$, provided that $c > 0$ is a small enough constant that depends on C. Thus, in this regime, the lemma amounts to simply finding a sumset source $\mathbf{W}^\star$ with full eval_d-rank that shares at least one support element with $\mathbf{W}$, which is straightforward.

[17] What exactly "sufficiently large" means will become clear later in the proof.

Thus, Lemma 14 tells us that a random degree d polynomial $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is an $(\varepsilon_0 + 2^{-c_0 k})$-extractor for (n,k)-sumset sources with $k \geq Cd(n/\varepsilon^2)^{1/\lfloor d/2 \rfloor}$, except with probability at most $2 \cdot 2^{-\varepsilon_0^2 t^2/32}$.

We now argue that $\varepsilon_0 + 2^{-c_0 k} \leq \varepsilon$. For this, it suffices to argue that $2^{-c_0 k} \leq \varepsilon_0$ (which was set to $\varepsilon/2$). Thus, given the lower bound on k, we just need $2^{-c_0 Cd(n/\varepsilon^2)^{1/\lfloor d/2 \rfloor}} \leq \varepsilon$. For this, it suffices to show $2^{-c_0 Cd(1/\varepsilon)^{1/\lfloor d/2 \rfloor}} \leq \varepsilon$. And if C is a sufficiently large constant $C \geq (\log_2 e)/(2c_0)$, then it suffices to show $2^{-(\log_2 e)(d/2)(1/\varepsilon)^{1/\lfloor d/2 \rfloor}} \leq \varepsilon$, or $e^{\lfloor d/2 \rfloor (1/\varepsilon)^{1/\lfloor d/2 \rfloor}} \geq 1/\varepsilon$, which is $e^{(1/\varepsilon)^{1/\lfloor d/2 \rfloor}} \geq (1/\varepsilon)^{1/\lfloor d/2 \rfloor}$. And this follows from the standard inequality $x \leq e^x$ for all real x.

Finally, it is easy to verify that $2 \cdot 2^{-\varepsilon_0^2 t^2/32} \leq 2^{-\varepsilon^2 \binom{k/C}{2\lfloor d/2 \rfloor}}$, using our definitions of ε_0 and t above, and the fact that $k \geq Cd(n/\varepsilon^2)^{1/\lfloor d/2 \rfloor}$ for a sufficiently large constant C. □

5.3 Impossibility results

Above, we proved that random low-degree polynomials are good extractors for sumset sources (Theorem 2). Given this result, it is natural to ask just how tight it is: in particular, can one show that every low-degree polynomial is nevertheless constant on some relatively large sumset source? As it turns out, the work of Cohen and Tal [CT15] already gives an answer to this question. Indeed, they show that every low-degree polynomial is constant on some relatively large *affine source*, which is a special case of a sumset source. Thus, their result already implies that every degree d polynomial $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is constant on some (n,k)-sumset source $\mathbf{W}$ with $k \geq \Omega(dn^{\frac{1}{d-1}})$, proving that Theorem 2 is tight up to the constant 2 in the exponent.

In this section, we instead ask whether such a result can be obtained for *two independent sources* – which can be viewed as the *other* well-known specialization of sumset sources. We prove such a result below, and show that it holds even if we consider the following (more general) family of functions $f : \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{F}_2$.

Definition 15 (Left-degree)**.** *We say that a function $f : \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{F}_2$ has* left-degree d *if for any fixed $y \in \mathbb{F}_2^n$, the restricted function $f(\cdot, y)$ can be written as an $\mathbb{F}_2$-polynomial of degree $\leq d$.*

Note that every polynomial of degree $\leq d$ also has left degree $\leq d$, but a function with left degree $\leq d$ can be much more complex than a degree $\leq d$ polynomial. Now, given this definition, we prove the following impossibility result for two-source dispersers that are computed by functions with low left-degree.

Theorem 7. *There is a constant $c > 0$ such that for every positive integers n and $d \leq c \log n$, the following holds. For any function $f : \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{F}_2$ with left-degree $\leq d$, there exist independent (n,k)-sources $\mathbf{X}, \mathbf{Y}$ with $k \geq cdn^{1/d}$ such that $f(\mathbf{X}, \mathbf{Y})$ is constant.*

Before we prove Theorem 7, we present a corollary for degree $\leq d$ polynomials.

Corollary 6. *There is a constant $c > 0$ such that for every positive integers n and $d \leq c \log n$, the following holds. For every polynomial $f : \mathbb{F}_2^{2n} \to \mathbb{F}_2$ with degree $\leq d$, there exist independent (n,k)-sources $\mathbf{X}, \mathbf{Y}$ with $k \geq cdn^{\frac{1}{d-1}}$ such that $f(\mathbf{X}, \mathbf{Y})$ is constant.*

Proof. Every polynomial $f : \mathbb{F}_2^{2n} \to \mathbb{F}_2$ of degree at most d can be written as

$$f(x,y) = g(x,y) + h(x),$$

such that g has degree at most d and does not feature monomials of f containing only variables from $\{x_1, \ldots, x_n\}$, and h collects all remaining monomials. Note that g has left-degree at most $d-1$. Thus, by Theorem 7, there exist independent (n,k)-sources $\mathbf{X}, \mathbf{Y}$ with $k \geq cdn^{\frac{1}{d-1}}$ such that $g(\mathbf{X}, \mathbf{Y})$ is constant.

Next, without loss of generality, assume that $\Pr[h(\mathbf{X}) = 0] \geq 1/2$. In this case, the random variable $\mathbf{X}' = (\mathbf{X} \mid h(\mathbf{X}) = 0)$ satisfies

$$H_\infty(\mathbf{X}') \geq H_\infty(\mathbf{X}) - 1 \geq k - 1 \geq c' d n^{\frac{1}{d-1}}$$

for some absolute constant $c' > 0$. Since $f(\mathbf{X}', \mathbf{Y}) = g(\mathbf{X}', \mathbf{Y})$ is constant, the desired result follows. □

Now, in order to prove Theorem 7, we revisit the argument used by Cohen and Tal [CT15] in the context of affine sources. In particular, we will use the following result from [CT15, Appendix B], which is a consequence of the Chevalley-Warning theorem. It says that if a low-degree polynomial f vanishes on a subspace V, then it also vanishes on many shifts of that subspace.

Lemma 16 ([CT15, Appendix B])**.** *Consider a subspace $V \subseteq \mathbb{F}_2^n$ of dimension t and a polynomial $f : \mathbb{F}_2^n \to \mathbb{F}_2$ of degree at most d such that $f(v) = 0$ for all $v \in V$. Then, it holds that*

$$|\{x \in \mathbb{F}_2^n : \forall v \in V, f(x + v) = 0\}| \geq 2^{n - \sum_{j=0}^{d-1}(d-j)\binom{t}{j}}.$$

With this tool in hand, we are ready to prove Theorem 7.

Proof of Theorem 7. We may alternatively see a function $f : \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{F}_2$ of left-degree at most $\leq d$ as a collection of 2^n polynomials $\{f_y : \mathbb{F}_2^n \to \mathbb{F}_2\}_{y \in \mathbb{F}_2^n}$, each of degree $\leq d$, where $f_y(x) := f(x, y)$. Without loss of generality, we assume that $f_y(0) = 0$ for at least 2^{n-1} choices of y. The desired result follows if we can find sets $X, Y \subseteq \mathbb{F}_2^n$, each of size at least $2^{cdn^{1/d}}$, such that $f_y(x) = 0$ for every $y \in Y$ and $x \in X$. Indeed, we can then take $\mathbf{X}$ and $\mathbf{Y}$ to be uniformly distributed over X and Y, respectively.

Towards this end, consider sampling a subspace $V \subseteq \mathbb{F}_2^n$ of dimension

$$t = \left\lfloor \frac{d}{e} \left(\frac{en}{4d^3} - 1 \right)^{1/d} \right\rfloor$$

by iteratively sampling v_i uniformly at random from $\mathbb{F}_2^n \setminus \operatorname{span}(v_1, \ldots, v_{i-1})$ for $i = 1, \ldots, t$. Looking ahead, we will choose $c > 0$ small enough (i.e., smaller than $\frac{1}{4\log(2e)}$) so that $d \leq \frac{t-1}{2}$ for $d \leq c \log n$.

Using Lemma 16, the probability that some fixed polynomial $f : \mathbb{F}_2^n \to \mathbb{F}_2$ of degree at most d with $f(0) = 0$ satisfies $f(x) = 0$ for all $x \in V$ is at least

$$\begin{aligned}
\prod_{i=0}^{t-1} \frac{2^{n - \sum_{j=0}^{d-1}(d-j)\binom{i}{j}} - 2^i}{2^n} &\geq \prod_{i=0}^{t-1} \frac{2^{n-1-\sum_{j=0}^{d-1}(d-j)\binom{i}{j}}}{2^n} && (16) \\
&= 2^{-t - \sum_{i=0}^{t-1}\sum_{j=0}^{d-1}(d-j)\binom{i}{j}} \\
&\geq 2^{-t - d^2 t (et/d)^{d-1}} && (17) \\
&\geq 2^{-n/2}. && (18)
\end{aligned}$$

Equation (16) holds because

$$n - \sum_{j=0}^{d-1}(d-j)\binom{t-1}{j} \geq n - d^2 \left(\frac{et}{d}\right)^{d-1} \geq 3n/4 \geq 2t,$$

which follows from the fact that $d \le \frac{t-1}{2}$ and by our choice of t. Equation (17) uses the fact that

$$\sum_{i=0}^{t-1}\sum_{j=0}^{d-1}(d-j)\binom{i}{j} \le d^2 t\left(\frac{et}{d}\right)^{d-1} < n/4,$$

again by our choice of t.

Now, if $\mathbf{A} \sim [2^n]$ denotes the number of polynomials in the collection $\{f_y : y \in \mathbb{F}_2^n, f_y(0) = 0\}$ such that $f_y(v) = 0$ for all $v \in V$, then by linearity of expectation and Equation (18), we have

$$\mathbb{E}[\mathbf{A}] \ge 2^{n-1} \cdot 2^{-n/2} = 2^{n/2-1} \ge 2^t.$$

Therefore, by an averaging argument, there exist a t-dimensional subspace V and a set $Y \subseteq \mathbb{F}_2^n$ of size at least 2^t such that $f_y(V) \equiv 0$ for all $y \in Y$. Taking $X = V$ and this Y (and $\mathbf{X}$ and $\mathbf{Y}$ uniformly distributed over these subsets, respectively) concludes the argument. □

6 Low-degree polynomials yield evasive sets

6.1 Low-degree, low-error two-source extractors

Recall that in Theorem 2, we showed that a random degree d polynomial $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is an ε-extractor for n-bit sumset sources with min-entropy $k = O(d(n/\varepsilon^2)^{1/\lfloor d/2 \rfloor})$. While this can handle any range of parameters, notice that in order to achieve "low error" $\varepsilon = k^{-\omega(1)}$, it requires superconstant degree $d = \omega(1)$. It is natural to ask whether low error can still be achieved with polynomials of constant degree $d = O(1)$.

As a step in this direction, in this section, we show the existence of a low-error two-source extractor for arbitrary linear min-entropy $k = \delta n$, which is computable by a degree ≤ 4 polynomial. In contrast, the best known explicit constructions of low-error two-source extractors require min-entropy $k \approx 0.44n$ [Bou05, Lew19]. More precisely, we prove the following.

Theorem 8 (Low-degree low-error two-source extractor)**.** *For every $\delta > 0$ there is some $\zeta > 0$ such that there exists a $(k = \delta n, \varepsilon = 2^{-\zeta n})$-two-source extractor $\mathsf{Ext} : \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{F}_2$ that is computable by an $\mathbb{F}_2$-polynomial of degree ≤ 4. Moreover, such a function Ext can be sampled with probability 0.99 in time $O(n^3)$, and $\mathsf{Ext}(x, y)$ can be evaluated in time $O(n^3)$ for every input (x, y).*

We prove Theorem 8 by combining a previously known connection between subspace-evasive sets (over $\mathbb{F}_2$) and low-error two-source extractors for linear min-entropy, and a construction of such subspace-evasive sets from degree-2 polynomials. This connection was originally conditional on (a weak form of) the *approximate duality conjecture* [BR15]. However, this conjecture is now a theorem since it is implied by the Polynomial Freiman-Ruzsa theorem, which was recently proved by Gowers, Green, Manners, and Tao [GGMT23].

A known reduction to low-degree subspace-evasive sets

We make the connection clear here for completeness. We need the following definition and previously known results.

Definition 16 (Rank of a function)**.** *We say that a function $f : \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{F}_2$ has* rank *at most r if there exist functions $g_1, g_2 : \mathbb{F}_2^n \to \mathbb{F}_2^r$ such that*

$$f(x, y) = \langle g_1(x), g_2(y) \rangle.$$

Equivalently, f has rank at most r if and only if its $2^n \times 2^n$ truth table has rank at most r over $\mathbb{F}_2$.

A uniformly random function will have exponentially large rank with high probability. The following result states that every *linear-rank* two-source disperser for linear min-entropy is also a low-error two-source extractor for essentially the same min-entropy.

Theorem 9 ([BR15, Lemma 2.17] and Polynomial Freiman-Ruzsa [GGMT23])**.** *For every $C, \gamma, \rho > 0$ there exists $\eta > 0$ such that the following holds for all n: Let $f : \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{F}_2$ be a $(k = \rho n)$-two-source disperser with rank at most Cn. Then, f is also a $(k' = (\rho + \gamma)n, \varepsilon = 2^{-\eta n})$-two-source extractor.*

Ben-Sasson, Lovett, and Ron-Zewi [BLR12] obtained an entropy-error tradeoff for this result, which allows one to drop the min-entropy k' to $O(n/\log n)$ while having error $2^{-\Omega(\sqrt{n})}$. For the sake of exposition, we do not use this alternative version, but remark that it can be applied here as well.

Based on the theorem above, our goal is to construct a linear-rank two-source disperser for any sublinear min-entropy. We use a simple observation already present in [PR04, BR15], whose short proof we include for completeness.

Lemma 17 ([PR04, BR15])**.** *If $A, B \subseteq \mathbb{F}_2^n$ are sets such that*[18] $\dim A + \dim B > n + 1$*, then*

$$\langle A, B \rangle := \{\langle a, b \rangle : a \in A, b \in B\} = \{0, 1\}.$$

Proof. First, suppose that $\langle A, B \rangle = \{0\}$. This means that B lies in the subspace orthogonal to $\operatorname{span}(A)$, and so $\dim A + \dim B \leq n$. On the other hand, if $\langle A, B \rangle = \{1\}$, then fix any $a \in A$ and consider instead the set $A' = A - a$. Note that $\langle A', B \rangle = \{0\}$ and that $\dim A' \geq \dim A - 1$. By the argument above, we must have $\dim A' + \dim B \leq n$, and so also $\dim A + \dim B \leq n + 1$. □

Given an input length n and min-entropy requirement $k \leq n$, Lemma 17 suggests building an encoding function $h : \mathbb{F}_2^n \to \mathbb{F}_2^r$ that is subspace-evasive in the sense of Pudlák and Rödl [PR04].

Definition 17 (Subspace-evasive sets and functions)**.** *A set $S \subseteq \mathbb{F}_2^r$ is said to be $(\ell, 2^k)$-*subspace-evasive *if $|S \cap V| < 2^k$ for any dimension-ℓ subspace V of $\mathbb{F}_2^r$. An injective function $h : \mathbb{F}_2^n \to \mathbb{F}_2^r$ is $(\ell, 2^k)$-*subspace-evasive *if $h(\mathbb{F}_2^n)$ is $(\ell, 2^k)$-subspace-evasive.*

Given a function $h : \mathbb{F}_2^n \to \mathbb{F}_2^r$, we consider the two-source extractor

$$\mathsf{Ext}_h(x, y) = \langle h(x), h(y) \rangle.$$

The following lemma relates the subspace-evasiveness of h to the dispersion and rank of Ext_h.

Lemma 18. *Given an injective $(\ell = r/2, 2^k)$-subspace-evasive function $h : \mathbb{F}_2^n \to \mathbb{F}_2^r$, it follows that Ext_h is a k-two-source disperser of rank at most r.*

Proof. Fix any two sets $A, B \subseteq \mathbb{F}_2^n$ of size 2^k each. We have that $\dim h(A), \dim h(B) \geq r/2 + 1$, since h is $(r/2, 2^k)$-subspace-evasive. In particular, this means that $\dim h(A) + \dim h(B) > r + 1$. By Lemma 17, we conclude that $\mathsf{Ext}_h(A, B) = \langle h(A), h(B) \rangle = \{0, 1\}$. □

Combining Lemma 18 with Theorem 9, we immediately obtain the following theorem, which reduces constructing a low-error two-source extractor for any linear min-entropy to constructing an appropriate subspace evasive set.

Theorem 10. *For every $C, \gamma, \rho > 0$ there exists $\eta > 0$ such that the following holds for all n: Let $h : \mathbb{F}_2^n \to \mathbb{F}_2^r$ be an injective $(\ell = r/2, 2^k = 2^{\rho n})$-subspace-evasive function with $r \leq Cn$. Then, $\mathsf{Ext}_h(x, y) = \langle h(x), h(y) \rangle$ is a $(k' = (\rho + \gamma)n, \varepsilon = 2^{-\eta n})$-two-source extractor. Moreover, if h is explicit, then so is Ext_h.*

[18] We take $\dim A$ to mean $\dim \operatorname{span}(A)$.

A random construction of low-degree subspace-evasive sets

With Theorem 10 in mind, it remains to construct an $(\ell, 2^k)$-subspace-evasive function $h : \mathbb{F}_2^n \to \mathbb{F}_2^{n+r}$ with $r = O(n)$, $k = o(n)$, and $\ell = \frac{n+r}{2}$. Furthermore, we would like to construct h in a way that Ext_h is computable by a low-degree $\mathbb{F}_2$-polynomial. Towards this end, we consider the function

$$h(x) = (x_1, \ldots, x_n, f_1(x), f_2(x), \ldots, f_r(x)),$$

where $f_1, \ldots, f_r : \mathbb{F}_2^n \to \mathbb{F}_2$ are random degree d polynomials.

Theorem 11. *Let $k = 10dn^{1/d}$ and $r = 11n$. Then, with probability at least $1 - 2^{-3n^2}$ over the sampling of $f_1, \ldots, f_r$ we have that h is an injective $(\ell = \frac{n+r}{2}, 2^k)$-subspace-evasive function.*

Sudakov and Tomon [ST23] also consider subspace-evasive sets induced by collections of multivariate polynomials, but they analyze instead the setting where the field is large and both the dimension and intersection size are small. Next, we need the following lemma.

Lemma 19 ([BHL12, Lemma 1.4] and [KS05, Theorem 1.5], adapted)**.** *For any set $S \subseteq \mathbb{F}_2^n$ of size 2^k we have that $\operatorname{rank}(\operatorname{eval}_d(S)) \geq \binom{k}{\leq d}$.*

With this lemma in hand, we are ready to prove Theorem 11.

Proof of Theorem 11. The injectivity of h holds by construction, so we focus on the subspace-evasiveness of h. We need to show that with high probability it holds that for any set $A \subseteq \mathbb{F}_2^n$ of size 2^k we have $\dim h(A) > \frac{r+n}{2}$. We may write $h(x) = (x, M \operatorname{eval}_d(x))$, where M is an $r \times \binom{n}{\leq d}$ matrix with i.i.d. uniformly random $\mathbb{F}_2$ entries.

Fix such a set A. Then, by Lemma 19, there exists a subset $S \subseteq A$ of size $r = 10n \leq \binom{k}{\leq d}$ such that the vectors in $\operatorname{eval}_d(S)$ are all linearly independent. It suffices to show that $\dim(M \operatorname{eval}_d(S)) > \frac{r+n}{2} = 5.5n$. Choose $\frac{n+r}{2}$ vectors in $\operatorname{eval}_d(S) = \{v_1, v_2, \ldots, v_r\}$. Without loss of generality, we may take these vectors to be $v_1, \ldots, v_{\frac{n+r}{2}}$. Since the v_i's are linearly independent, the random variables $(Mv_i)_{i \in [r]}$ are also independent. For a fixed $i > \frac{n+r}{2}$, the probability that Mv_i lies in $\operatorname{span}(Mv_1, \ldots, Mv_{\frac{n+r}{2}})$ is $2^{-\frac{r-n}{2}}$. Therefore, the probability that $Mv_{\frac{n+r}{2}+1}, \ldots, Mv_r$ all lie in this subspace is

$$(2^{-\frac{r-n}{2}})^{r-\frac{n+r}{2}} = 2^{-25n^2}.$$

Then, a union bound over the $\binom{|S|}{\frac{n+r}{2}} = \binom{r}{\frac{n+r}{2}}$ choices for the initial set of $\frac{n+r}{2}$ vectors from $\operatorname{eval}_d(S)$ shows that the probability that $\dim h(S) \leq \frac{n+r}{2}$ is at most

$$\binom{r}{\frac{n+r}{2}} \cdot 2^{-25n^2} \leq 2^r \cdot 2^{-25n^2} = 2^{11n-25n^2} \leq 2^{-14n^2}.$$

Finally, we take a union bound over all the $\binom{2^n}{r}$ choices of S to conclude that the probability that h is not subspace-evasive is at most

$$\binom{2^n}{r} \cdot 2^{-14n^2} \leq 2^{nr-14n^2} = 2^{-3n^2}.$$

□

We can now combine Theorem 11 for degree $d = 2$ with Theorem 10 to immediately obtain a low-error two-source extractor for arbitrary linear min-entropy, which in particular yields Theorem 8.

Corollary 7. *Set $r = 11n$. Sample r random degree 2 polynomials $f_1, \ldots, f_r : \mathbb{F}_2^n \to \mathbb{F}_2$ and define $h : \mathbb{F}_2^n \to \mathbb{F}_2^{n+r}$ as $h(x) = (x_1, \ldots, x_n, f_1(x), \ldots, f_r(x))$. Then, with probability at least $1 - 2^{-3n^2}$ over the sampling of $f_1, \ldots, f_r$ the following holds: For any $\delta > 0$ there exists $c > 0$ such that the function $\mathsf{Ext}_h : \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{F}_2$ given by*

$$\mathsf{Ext}_h(x, y) = \langle h(x), h(y) \rangle$$

is a $(k' = \delta n, \varepsilon = 2^{-cn})$-two-source extractor. Furthermore, Ext_h is computable by a $2n$-variate $\mathbb{F}_2$-polynomial of degree at most 4.

6.2 Improved impossibility results for dispersing from polynomial and variety sources

In this section, we show how to construct what we call *sumset-evasive sets* using low-degree polynomials. Then, we show how our constructions yield new impossibility results for dispersing from polynomial and variety sources using sumset dispersers, improving and simplifying results of Chattopadhyay, Goodman, and Gurumukhani [CGG24].

Definition 18 (Sumset-evasive sets and functions)**.** *A set $S \subseteq \mathbb{F}_2^r$ is said to be* ℓ-sumset-evasive *if for any sets $A, B \subseteq \mathbb{F}_2^r$ each of size 2^ℓ it holds that $A + B \not\subseteq S$. We say that a function $f : \mathbb{F}_2^n \to \mathbb{F}_2^r$ is ℓ-sumset-evasive if the image $f(\mathbb{F}_2^n) \subseteq \mathbb{F}_2^r$ is ℓ-sumset-evasive.*

We show that the same construction we used to obtain low-degree subspace-evasive sets above also works for sumset-evasiveness, if we use our new Lemma 13 instead of Lemma 19. To keep things simple, we focus here on even degree d, and note that it is straightforward to extend our results to hold for all degrees d, at a slight loss in parameters (as was done for our fully general result on sumset extractors in Section 5.2).

Theorem 12. *There exist constants $C, c > 0$ such that the following holds for all positive integers $k \geq t \geq 200d$ with d even and $\binom{t/100}{\leq d/2} \geq 4d^2(2e)^d$. Let $f_1, \ldots, f_r : \mathbb{F}_2^k \to \mathbb{F}_2$ be independent random degree d polynomials with*

$$r \geq \frac{8d^2(2e)^d k}{\binom{t/100}{\leq d/2}}.$$

Consider the function $h : \mathbb{F}_2^k \to \mathbb{F}_2^{k+r}$ defined as $h(x) = (x, f_1(x), \ldots, f_r(x))$. Then

$$\Pr_{f_1, \ldots, f_r}\left[h(\mathbb{F}_2^k) \text{ is not } t\text{-sumset-evasive}\right] \leq 2^{-\frac{r}{2} \cdot \binom{t/100}{\leq d}}.$$

Proof. Let $S \subseteq \mathbb{F}_2^{k+r}$ denote the image of h, and recall it is of the form

$$S = \{(x, f(x)) : x \in \mathbb{F}_2^k\},$$

where $f : \mathbb{F}_2^k \to \mathbb{F}_2^r$ is defined as $f(x) := (f_1(x), \ldots, f_r(x))$. First, suppose there exist $A, B \subseteq \mathbb{F}_2^n$ each of size at least 2^t such that $A + B \subseteq S$. Since the first k bits of S determine the rest, the same is true of both A and B. In particular, there must exist some $A', B' \subseteq \mathbb{F}_2^k$ each of size at least 2^t and functions $\phi, \psi : \mathbb{F}_2^k \to \mathbb{F}_2^r$ such that

$$A = \{(a, \phi(a)) : a \in A'\},$$
$$B = \{(b, \psi(b)) : b \in B'\}.$$

Furthermore, again because $A + B \subseteq S$, we know that for every $a \in A', b \in B'$,

$$(a, \phi(a)) + (b, \psi(b)) = (a + b, \phi(a) + \psi(b)) = (a + b, f(a + b)).$$

In particular, this implies that $f(a + b) = \phi(a) + \psi(b)$ for every $a \in A', b \in B'$.

Next, since A', B' each have size at least 2^t, we know by Lemma 13 that there must exist some $A'' \subseteq A', B'' \subseteq B'$ each of size $\tau = \binom{t/100}{\leq d/2}$ such that

$$\operatorname{rank}(\operatorname{eval}_d(A'' + B'')) \geq \binom{t/100}{\leq d} \geq (2e)^{-d} d^{-2} \tau^2 =: \tau'. \tag{19}$$

The rightmost inequality uses the fact that $\binom{t/100}{\leq d} \geq \left(\frac{t}{100d}\right)^d$ and that $\tau^2 = \binom{t/100}{\leq d/2}^2 \leq d^2 \left(\frac{2et}{100d}\right)^d$, which in turn use the standard inequalities $\binom{u}{\leq v} \geq \binom{u}{v} \geq (u/v)^v$ and $\binom{u}{\leq v} \leq v\binom{u}{v} \leq v\,(eu/v)^v$, valid for $v \leq u/2$.

Furthermore, if we define $\phi'' : A'' \to \mathbb{F}_2^r, \psi'' : B'' \to \mathbb{F}_2^r$ to denote the restrictions of ϕ, ψ to A'', B'' respectively, it still holds that $f(a + b) = \phi''(a) + \psi''(b)$ for all $a \in A'', b \in B''$. Now, note that by Equation (19) there are at least τ' choices of pairs $a \in A'', b \in B''$ such that the corresponding random variables $f(a + b)$ are independent and uniformly distributed over $\mathbb{F}_2^r$, while $\phi''(a) + \psi''(b)$ is a fixed value. Thus, the probability that $f(a + b) = \phi''(a) + \psi''(b)$ is precisely 2^{-r}. Therefore, by the independence of $f(a + b)$ for the choices of $a \in A''$ and $b \in B''$ above, the probability they are equal for *every* fixed pair $a \in A'', b \in B''$ above is precisely $2^{-r \cdot \tau'}$. Applying a union bound over all possible choices of A'', B'', ϕ'', ψ'' yields that the probability that such objects exist is at most

$$2^{-r\tau'} \cdot \binom{2^k}{\tau}^2 \cdot ((2^r)^\tau)^2 \leq 2^{-r\tau' + 2k\tau + 2r\tau},$$

which is at most $2^{-r\tau'/2}$ provided that $r \geq \frac{4\tau k}{\tau' - 4\tau} = \frac{4k}{(2e)^{-d} d^{-2} \tau - 4}$, the result follows by noting that our hypotheses implies $(2e)^{-d} d^{-2} \tau - 4 \geq (2e)^{-d} d^{-2} \tau / 2$, so it suffices to choose any $r \geq 8d^2 (2e)^d k / \tau$. □

Note that for the functions h from Theorem 12 we have that $\mathbf{Z} = h(\mathbf{U}_k)$ is a degree $\leq d$ polynomial source of length $n = k + r$ and min-entropy k such that, with high probability over the choice of h, it holds that $\mathsf{supp}(\mathbf{Z})$ does not contain any sumset $A + B$ with $|A|, |B| \geq 2^t$. Moreover, the *entropy gap* of $\mathbf{Z}$ is

$$n - k = r = O\left(\frac{8d^2 (2e)^d k}{\binom{t/100}{\leq d/2}}\right) = O\left(8d^2 (2e)^d n \left(\frac{100d}{2t}\right)^{d/2}\right) \leq n \left(\frac{Cd}{t}\right)^{d/2},$$

for a sufficiently large constant $C > 0$. The third equality uses the fact that $k \leq n$ and $\binom{a}{\leq b} \geq \binom{a}{b} \geq \left(\frac{a}{b}\right)^b$. Using this and setting $t = \alpha \log n$ for a sufficiently small constant $\alpha > 0$ yields the following corollary of Theorem 12, which states that there exist polynomial NOBF sources[19] with extremely high min-entropy that avoid even very tiny sumsets.

Corollary 8. *For any constant $c > 0$ there exists a constant $C > 0$ such that the following holds for all positive integers $n \geq C$ and even degrees d. There exists a degree d polynomial NOBF source $\mathbf{Z} \sim \mathbb{F}_2^n$ with min-entropy $H_\infty(\mathbf{Z}) \geq n - n\left(\frac{Cd}{\log n}\right)^{d/2}$ such that $\mathsf{supp}(\mathbf{Z})$ does not contain any sumset $A + B$ with $|A|, |B| \geq n^c$.*

[19] Following [CGG24, Definition 4], a degree d polynomial NOBF source $\mathbf{Z} \sim \mathbb{F}_2^n$ with min-entropy k is a degree d polynomial source with the following structure: There exists a set $G \subseteq [n]$ of k good coordinates which are independent and uniformly distributed, and each coordinate outside G is computed by a degree $\leq d$ polynomial of the coordinates in G. Polynomial NOBF sources are a special case of both polynomial sources and variety sources.

Corollary 8 generalizes [CGG24, Theorem 5.11], which worked for $d = 2$ only, to all even degrees d, and it has a simpler proof. It also yields the following corollary, extending the impossibility result from [CGG24, Section 5.3] on using sumset dispersers to disperse from polynomial NOBF sources in a black-box manner to larger degrees $d > 2$.

Corollary 9 (Sumset dispersers cannot disperse from high-entropy polynomial NOBF sources and variety sources, any even degree d)**.** *Fix any even degree $d \in \mathbb{N}$. Then, there exists a constant $C > 0$ such that for any integer $n \geq C$ the following holds. Sumset dispersers cannot be used to disperse from degree d polynomial NOBF sources over $\mathbb{F}_2^n$ with min-entropy $k = n - n\left(\frac{Cd}{\log n}\right)^{d/2}$ in a black-box manner. In particular, this also means that sumset dispersers cannot be used to disperse from degree d variety sources with the same min-entropy k in a black-box manner.*

Proof. In order to be able to conclude that k'-sumset disperser is also a disperser for another class of sources $\mathcal{C}$ using only its black-box property (i.e., that it is not constant on any sumset $A + B$ with $|A|, |B| \geq 2^{k'}$), one needs that for every $\mathbf{X} \in \mathcal{C}$, $\mathsf{supp}(\mathbf{X})$ contains a sumset $A + B$ with $|A|, |B| \geq 2^{k'}$. Since we can take $k' = \alpha \log n$ for an arbitrarily small constant $\alpha > 0$ and there are no sumset dispersers for such k' (because every function is constant on an $\Omega(\log n)$-dimensional subspace, and affine sources are sumset sources), the result follows. To see the "In particular" part, it suffices to note that degree d polynomial NOBF sources with min-entropy k are also degree d variety sources with min-entropy k [CGG24, Claim 1]. □

7 Open problems

We list here some of our favorite directions for future research:

- In Theorem 2, we showed that most degree $\leq d$ polynomials are sumset dispersers (in fact, extractors) for min-entropy $k = O(dn^{1/\lfloor d/2 \rfloor})$. On the other hand, we also know that no degree $\leq d$ polynomial is a sumset disperser for min-entropy $k = c \cdot dn^{1/(d-1)}$, where $c > 0$ is some constant. Can we narrow this gap?

- We conjecture that most degree $\leq d$ polynomials are sumset extractors with *exponentially small error* for min-entropy $k = Cdn^{C/d}$ for some constant $C > 0$, even when d is a constant.[20] We think that even showing this for small linear min-entropy would already be quite interesting.

- In Theorem 8, we showed that there exist degree ≤ 4 low-error two-source extractors for any linear min-entropy via approximate duality. This approach, however, provably cannot go below min-entropy $\sqrt{n}$ [BLR12]. Can we show the existence of low-degree low-error two-source extractors for min-entropy below $\sqrt{n}$?

Acknowledgements

We thank Alexander Golovnev, Zeyu Guo, Pooya Hatami, Satyajeet Nagargoje, and Chao Yan for sharing with us an early draft of their work. We also thank Dean Doron for insightful discussions and feedback, and Mohit Gurumukhani for helpful pointers to facts about quadratic forms. Part of this work was carried out while the authors were visiting the Simons Institute for the Theory of Computing at UC Berkeley, and while the second and fourth author were visiting NTT Research in Sunnyvale, CA.

[20] Note that Theorem 2 can handle min-entropy $k = O(dn^{1/\lfloor d/2 \rfloor})$ if $d \geq O(\log(1/\varepsilon))$.

References

[ABK08] Noga Alon, Ido Ben-Eliezer, and Michael Krivelevich. Small sample spaces cannot fool low degree polynomials. In *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2008)*, pages 266–275. Springer Berlin Heidelberg, 2008.

[ACG+22] Omar Alrabiah, Eshan Chattopadhyay, Jesse Goodman, Xin Li, and João Ribeiro. Low-degree polynomials extract from local sources. In *49th International Colloquium on Automata, Languages, and Programming (ICALP 2022)*, pages 10:1–10:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022.

[ASW15] Emmanuel Abbe, Amir Shpilka, and Avi Wigderson. Reed–Muller codes for random erasures and errors. *IEEE Transactions on Information Theory*, 61(10):5229–5252, 2015.

[ASY20] Emmanuel Abbe, Amir Shpilka, and Min Ye. Reed–muller codes: Theory and algorithms. *IEEE Transactions on Information Theory*, 67(6):3251–3277, 2020.

[BG13] Andrej Bogdanov and Siyao Guo. Sparse extractor families for all the entropy. In *Innovations in Theoretical Computer Science (ITCS 2013)*, pages 553–560. ACM, 2013.

[BHL12] Ido Ben-Eliezer, Rani Hod, and Shachar Lovett. Random low-degree polynomials are hard to approximate. *Comput. Complex.*, 21(1):63–81, 2012.

[BLR12] Eli Ben-Sasson, Shachar Lovett, and Noga Ron-Zewi. An additive combinatorics approach relating rank to communication complexity. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science (FOCS 2012)*, pages 177–186, 2012.

[BNS92] László Babai, Noam Nisant, and Márió Szegedy. Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *Journal of Computer and System Sciences*, 45(2):204–232, 1992.

[Bou05] Jean Bourgain. More on the sum-product phenomenon in prime fields and its applications. *International Journal of Number Theory*, 01(01):1–32, 2005.

[BOY20] Paul Beame, Shayan Oveis Gharan, and Xin Yang. On the bias of Reed-Muller codes over odd prime fields. *SIAM Journal on Discrete Mathematics*, 34(2):1232–1247, 2020.

[BR15] Eli Ben-Sasson and Noga Ron-Zewi. From affine to two-source extractors via approximate duality. *SIAM Journal on Computing*, 44(6):1670–1697, 2015. Preliminary version in STOC 2011.

[BSHR+01] Eli Ben-Sasson, Shlomo Hoory, Eyal Rozenman, Salil Vadhan, and Avi Wigderson. Extractors for affine sources. Unpublished manuscript, 2001.

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

[CFPY24] David Conlon, Jacob Fox, Huy Tuan Pham, and Liana Yepremyan. On the clique number of random Cayley graphs and related topics, 2024. https://arxiv.org/abs/2412.21194.

[CG88] Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing*, 17(2):230–261, 1988.

[CG22] Eshan Chattopadhyay and Jesse Goodman. Improved extractors for small-space sources. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS 2022)*, pages 610–621. IEEE, 2022.

[CG25] Eshan Chattopadhyay and Jesse Goodman. Leakage-resilient extractors against number-on-forehead protocols. In *Proceedings of the 57th Annual ACM SIGACT Symposium on Theory of Computing (STOC 2025, to appear)*. ACM, 2025.

[CGG24] Eshan Chattopadhyay, Jesse Goodman, and Mohit Gurumukhani. Extractors for polynomial sources over $\mathbb{F}_2$. In *15th Innovations in Theoretical Computer Science Conference (ITCS 2024)*, pages 28:1–28:24. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024.

[CGZ22] Eshan Chattopadhyay, Jesse Goodman, and David Zuckerman. The space complexity of sampling. In *13th Innovations in Theoretical Computer Science Conference (ITCS 2022)*, volume 215 of *LIPIcs*, pages 40:1–40:23. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022.

[CL16] Eshan Chattopadhyay and Xin Li. Extractors for sumset sources. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing (STOC 2016)*, page 299–311, 2016.

[CL18] Kuan Cheng and Xin Li. Randomness extraction in AC0 and with small locality. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018)*, pages 37:1–37:20. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2018.

[CL22] Eshan Chattopadhyay and Jyun-Jie Liao. Extractors for sum of two sources. In *54th Annual ACM SIGACT Symposium on Theory of Computing (STOC 2022)*, pages 1584–1597. Association for Computing Machinery, 2022.

[CL23] Eshan Chattopadhyay and Jyun-Jie Liao. Hardness against linear branching programs and more. In *38th Computational Complexity Conference (CCC 2023)*, pages 9:1–9:27, 2023.

[CS16] Gil Cohen and Igor Shinkar. The complexity of DNF of parities. In *2016 ACM Conference on Innovations in Theoretical Computer Science*, pages 47–58, 2016.

[CT15] Gil Cohen and Avishay Tal. Two structural results for low degree polynomials and applications. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2015)*, pages 680–709, 2015.

[CW24] Kuan Cheng and Ruiyang Wu. Randomness extractors in AC^0 and NC^1: Optimal up to constant factors. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2024)*, pages 69:1–69:22, 2024.

[DD22] Manik Dhar and Zeev Dvir. Linear hashing with ℓ_∞ guarantees and two-sided Kakeya bounds. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS 2022)*, pages 419–428, 2022.

[DGW09] Zeev Dvir, Ariel Gabizon, and Avi Wigderson. Extractors and rank extractors for polynomial sources. *Comput. Complex.*, 18(1):1–58, 2009.

[DK11] Evgeny Demenkov and Alexander Kulikov. An elementary proof of a $3n - o(n)$ lower bound on the circuit complexity of affine dispersers. In *International Symposium on Mathematical Foundations of Computer Science*, pages 256–265. Springer, 2011.

[DLWZ14] Yevgeniy Dodis, Xin Li, Trevor D Wooley, and David Zuckerman. Privacy amplification and nonmalleable extractors via character sums. *SIAM Journal on Computing*, 43(2):800–830, 2014.

[DM02] Stefan Dziembowski and Ueli Maurer. Tight security proofs for the bounded-storage model. In *Thiry-Fourth Annual ACM Symposium on Theory of Computing (STOC 2002)*, page 341–350. Association for Computing Machinery, 2002.

[DT09] Anindya De and Luca Trevisan. Extractors using hardness amplification. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2009)*, pages 462–475. Springer Berlin Heidelberg, 2009.

[Dvi09] Zeev Dvir. Extractors for varieties. In *2009 24th Annual IEEE Conference on Computational Complexity (CCC 2009)*, pages 102–113, 2009.

[Dvi12] Zeev Dvir. Incidence theorems and their applications. *Foundations and Trends® in Theoretical Computer Science*, 6(4):257–393, 2012.

[DW12] Anindya De and Thomas Watson. Extractors and lower bounds for locally samplable sources. *ACM Trans. Comput. Theory*, 4(1), March 2012.

[FGHK16] Magnus Gausdal Find, Alexander Golovnev, Edward Hirsch, and Alexander S Kulikov. A better-than-$3n$ lower bound for the circuit complexity of an explicit function. In *57th Annual Symposium on Foundations of Computer Science (FOCS 2016)*, pages 89–98. IEEE, 2016.

[GG25] Jesse Goodman and Vipul Goyal. Two-source extractors don't shrink. Unpublished manuscript, 2025.

[GGH+24] Alexander Golovnev, Zeyu Guo, Pooya Hatami, Satyajeet Nagargoje, and Chao Yan. Hilbert functions and low-degree randomness extractors. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2024)*, pages 41:1–41:24, 2024.

[GGMT23] W. T. Gowers, Ben Green, Freddie Manners, and Terence Tao. On a conjecture of Marton, 2023. https://arxiv.org/abs/2311.05762.

[GK16] Alexander Golovnev and Alexander Kulikov. Weighted gate elimination: Boolean dispersers for quadratic varieties imply improved circuit lower bounds. In *7th Conference on Innovations in Theoretical Computer Science (ITCS 2016)*, pages 405–411, 2016.

[GKW21] Alexander Golovnev, Alexander S. Kulikov, and R. Ryan Williams. Circuit depth reductions. In *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*, pages 24:1–24:20. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2021.

[Goo23] Jesse Goodman. *Seedless Extractors*. PhD thesis, Cornell University, 2023.

[GPT22] Svyatoslav Gryaznov, Pavel Pudlák, and Navid Talebanfard. Linear branching programs and directional affine extractors. In *37th Computational Complexity Conference (CCC 2022)*, pages 4:1–4:16. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022.

[GRS23] Venkatesan Guruswami, Atri Rudra, and Madhu Sudan. *Essential Coding Theory*. 2023. Draft available at https://cse.buffalo.edu/faculty/atri/courses/coding-theory/book.

[GVJZ23] Zeyu Guo, Ben Lee Volk, Akhil Jalan, and David Zuckerman. Extractors for images of varieties. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, STOC 2023, page 46–59, New York, NY, USA, 2023. Association for Computing Machinery.

[GVW15] Oded Goldreich, Emanuele Viola, and Avi Wigderson. On randomness extraction in AC0. In *Proceedings of the 30th Conference on Computational Complexity (CCC 2015)*, pages 601–668. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2015.

[HHTT23] Pooya Hatami, William M Hoza, Avishay Tal, and Roei Tell. Depth-d threshold circuits vs. depth-$(d+1)$ AND-OR trees. In *55th Annual ACM Symposium on Theory of Computing (STOC 2023)*, pages 895–904, 2023.

[HILL99] Johan Håstad, Russell Impagliazzo, Leonid A Levin, and Michael Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28(4):1364–1396, 1999.

[HIV22] Xuangui Huang, Peter Ivanov, and Emanuele Viola. Affine extractors and AC0-parity. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2022)*, pages 9:1–9:14, 2022.

[HR15] Pavel Hrubes and Anup Rao. Circuits with medium fan-in. In *30th Conference on Computational Complexity (CCC 2015)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2015.

[KLP12] Tali Kaufman, Shachar Lovett, and Ely Porat. Weight distribution and list-decoding size of Reed–Muller codes. *IEEE Transactions on Information Theory*, 58(5):2689–2696, 2012.

[Kop10] Swastik Kopparty. *Algebraic methods in randomness and pseudorandomness*. PhD thesis, Massachusetts Institute of Technology, 2010.

[KRVZ11] Jesse Kamp, Anup Rao, Salil Vadhan, and David Zuckerman. Deterministic extractors for small-space sources. *J. Comput. Syst. Sci.*, 77(1):191–220, 2011.

[KS05] Peter Keevash and Benny Sudakov. Set systems with restricted cross-intersections and the minimum rank of inclusion matrices. *SIAM J. Discret. Math.*, 18(4):713–727, 2005.

[KS18] S.V. Konyagin and I.D. Shkredov. On subgraphs of random Cayley sum graphs. *European Journal of Combinatorics*, 70:61–74, 2018.

[Lew19] Mark Lewko. An explicit two-source extractor with min-entropy rate near $4/9$. *Mathematika*, 65(4):950–957, 2019.

[Li11] Xin Li. A new approach to affine extractors and dispersers. In *2011 IEEE 26th Annual Conference on Computational Complexity (CCC 2011)*, pages 137–147, 2011.

[LN97] Rudolf Lidl and Harald Niederreiter. *Finite Fields*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2 edition, 1997.

[Lu02] Chi-Jen Lu. Hyper-encryption against space-bounded adversaries from on-line strong extractors. In Moti Yung, editor, *Advances in Cryptology — CRYPTO 2002*, pages 257–271, Berlin, Heidelberg, 2002. Springer Berlin Heidelberg.

[LY22] Jiatu Li and Tianqi Yang. $3.1n - o(n)$ circuit lower bounds for explicit functions. In *54th Annual ACM Symposium on Theory of Computing (STOC 2022)*, pages 1180–1193, 2022.

[LZ19] Fu Li and David Zuckerman. Improved extractors for recognizable and algebraic sources. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019)*, pages 72:1–72:22. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2019.

[LZ24] Xin Li and Yan Zhong. Explicit directional affine extractors and improved hardness for linear branching programs. In *39th Computational Complexity Conference (CCC 2024)*, pages 10:1–10:14. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024.

[Mra16] Rudi Mrazović. Extractors in Paley graphs: A random model. *European Journal of Combinatorics*, 54:154–162, 2016.

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

[PR04] Pavel Pudlák and Vojtěch Rödl. Pseudorandom sets and explicit constructions of Ramsey graphs. *Complexity of Computations and Proofs. Quad. Mat. 13 Dept. Math. Seconda Univ. Napoli Caserta*, pages 327–346, 2004.

[Rem16] Zachary Remscrim. The Hilbert function, algebraic extractors, and recursive Fourier sampling. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 197–208, 2016.

[RY11] Ran Raz and Amir Yehudayoff. Multilinear formulas, maximal-partition discrepancy and mixed-sources extractors. *Journal of Computer and System Sciences*, 77(1):167–190, 2011. Celebrating Karp's Kyoto Prize. Preliminary version in FOCS 2008.

[Sha11] Ronen Shaltiel. An introduction to randomness extractors. In *Automata, Languages and Programming: 38th International Colloquium, ICALP 2011*, pages 21–41. Springer, 2011.

[ST23] Benny Sudakov and István Tomon. Evasive sets, covering by subspaces, and point-hyperplane incidences. *Discrete & Computational Geometry*, pages 1–15, 2023.

[Tre01] Luca Trevisan. Extractors and pseudorandom generators. *Journal of the ACM*, 48(4):860–879, 2001.

[TV06] Terence Tao and Van H Vu. *Additive combinatorics*, volume 105. Cambridge University Press, 2006.

[Vad04] Salil P. Vadhan. Constructing locally computable extractors and cryptosystems in the bounded-storage model. *Journal of Cryptology*, 17:43–77, 2004.

[Vad12] Salil Vadhan. Pseudorandomness. *Foundations and Trends in Theoretical Computer Science*, 7(1–3):1–336, 2012.

[Vio05] Emanuele Viola. The complexity of constructing pseudorandom generators from hard functions. *computational complexity*, 13(3-4):147–188, 2005.

[Vio14] Emanuele Viola. Extractors for circuit sources. *SIAM Journal on Computing*, 43(2):655–672, 2014.

A Existential results for sumset extractors via a uniformly random function

As a special case of our second main result (Theorem 2), we know that a random degree $d = O(\log(n/\varepsilon))$ polynomial is an ε-extractor for (n,k)-sumset sources with min-entropy $k = O(\log(n/\varepsilon))$. A natural question is whether there is a simpler proof that a uniformly random function can achieve this min-entropy requirement. In this section, we show that the answer is yes, and give a simple proof of the following.

Theorem 13 (Existential result for sumset extractors, originally proved in [Mra16])**.** *For every constant $\gamma > 0$ there exists a constant $C > 0$ such that for any $n \geq k \in \mathbb{N}$ and $\varepsilon > 0$ such that $k \geq (2+\gamma)\log n + (4+2\gamma)\log(1/\varepsilon) + C$, the following holds. A uniformly random function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is a (k,ε)-sumset extractor with probability at least 0.99.*

Theorem 13 was posed as an open problem by Chattopadhyay and Liao at STOC 2022 [CL22], and the previous version of our work claimed to resolve it for the first time.[21] However, it turns out that this result was already known in the Cayley sum graph literature since 2015. Namely, Mrazović [Mra16] showed that a uniformly random function is, with high probability, a (k,ε)-sumset extractor for $k = 2\log n + 6\log(1/\varepsilon) + O(1)$. Later work by Konyagin and Shkredov [KS18] improved the constant in front of the $\log n$ term at the expense of a worse dependence on ε. More recently, Conlon, Fox, Pham, and Yepremyan [CFPY24] extended this to more groups (among other results).

Thus, the existential result presented in Theorem 13 is not novel, and should be attributed to Mrazović. However, we opted to keep this section in the current version of our work, since our proof seems somewhat novel,[22] and because we believe the presentation here may nevertheless be helpful to other people in the field. In what follows, we give an overview of our proof of Theorem 13 vs. the proof of Mrazović, and conclude with a formal presentation of our proof.

Our approach vs. the original proof of Mrazović Unlike many pseudorandom objects, it is not easy to show that a uniformly random function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is a good extractor for (n,k)-sumset sources. This is because a naive counting argument shows that there are $\approx \binom{2^n}{2^k}\binom{2^n}{2^k} \geq 2^{2(n-k)2^k}$ such sources, but some of these sources only have 2^k elements in their support (for example, if the sumset source is actually an affine (n,k)-source). Thus, even if one had access to a perfect concentration bound, a naive application of the probabilistic method is destined to fail (since $2^{-2^k} \cdot 2^{2(n-k)2^k} \geq 1$).

In order to get around this issue, the key observation is as follows. While it seems difficult to dramatically improve the naive count of (n,k)-sumset sources, it does seem like *most* (n,k)-sumset sources should have $\gg 2^k$ elements in their support. And if we restrict our attention to (n,k)-sumset sources of this type, then there is some hope that the naive probabilistic argument (outlined above) will go through.

As it turns out, this is exactly the case. In particular, if we only consider the family $\mathcal{Y}$ of (n,k)-sumset sources $\mathbf{W} = \mathbf{X} + \mathbf{Y}$ with low *additive energy*,[23] then it is easy to show that a uniformly random function

[21] In fact, in the previous version of our work, Theorem 2 only worked for polynomially small error, and thus did not show that degree $d = O(\log(n/\varepsilon))$ polynomials can extract from sumset sources with min-entropy $k = O(\log(n/\varepsilon))$ (as it does now). Thus, in the previous version of our work, Theorem 13 was presented as a completely separate result that established the existence of sumset extractors for min-entropy $k = O(\log(n/\varepsilon))$, instead of as a "simpler version" of Theorem 2.

[22] In particular, our proof follows immediately from a seemingly novel structural result for sumset sources (Lemma 20). This structural result can be thought of as a simpler version of our key structural result used in the low-degree setting (Lemma 15), and the proof of Lemma 20 can be viewed as a simpler variant of the randomized convex combination approach used in the proof of Lemma 15. Moreover, as we will soon discuss, the structural result in Lemma 20 implies the ("one-sided" version of the) main structural result used in the work of Mrazović [Mra16].

[23] Having low additive energy can be thought of as a stronger (statistical) version of having support size $\gg 2^k$ - see Definition 19.

is an excellent extractor for $\mathcal{Y}$. This idea first appeared in the work of Mrazović [Mra16], and subsequently in the work of Chattopadhyay and Liao [CL22].

The final (and main) step is to show (via a structural result) that an extractor for sumset sources with low additive energy is automatically an extractor for *all* sumset sources. This is the step that is missing in [CL22], and provided by both Mrazović [Mra16, Proposition 4] and, subsequently, our work (Lemma 20, below). We outline and compare these structural results, below.

For his structural result, Mrazović shows that for any fixed function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ and sets $X, Y \subseteq \mathbb{F}_2^n$, there exist slightly smaller sets $X^\star, Y^\star \subseteq \mathbb{F}_2^n$ that (1) have low additive energy, and (2) have the property that $\mathsf{bias}(f(X+Y)) \leq \mathsf{bias}(f(X^\star + Y^\star)) + \gamma$ for some small γ.[24] As a result, it follows that any extractor for sumset sources with low additive energy automatically works for all sumset sources.[25] Mrazović shows the existence of such sets $X^\star, Y^\star$ by taking them to be uniformly random subsets of X, Y (respectively) of an appropriate size.

For our structural result, we show that every sumset source $\mathbf{X} + \mathbf{Y}$ is a convex combination of sumset sources $\mathbf{X}^\star + \mathbf{Y}^\star$ with low additive energy. This also shows that any extractor for sumset sources with low additive energy automatically works for all sumset sources. Furthermore, it implies the (one-sided) structural result of Mrazović (even with $\gamma = 0$), by an averaging argument over the participants in the convex combination. In order to obtain our structural result, we assume (without loss of generality) that $\mathbf{X}, \mathbf{Y}$ are uniform over some sets X, Y, and randomly partition these sets into subsets $\{X_i^\star\}, \{Y_i^\star\}$ that each have the same size. The participants in our convex combination are then taken to be all random variables of the form $\mathbf{X}_i^\star + \mathbf{Y}_j^\star$, where $\mathbf{X}_i^\star, \mathbf{Y}_j^\star$ are uniform over $X_i^\star, Y_j^\star$.

A formal presentation of our approach

We now present our formal proof of Theorem 13. For this, we need the notion of *additive energy* between two sets X, Y (which is closely related to the collision probability, or Rényi entropy H_2, of their sum).

Definition 19 ([TV06, Definition 2.8])**.** *The* additive energy *between two sets $X, Y \subseteq \mathbb{F}_2^n$ is defined as*

$$E(X,Y) := \sum_{w \in \mathbb{F}_2^n} |\{(x,y) \in X \times Y : x+y = w\}|^2 = |\{(x,y,x',y') \in X \times Y \times X \times Y : x+y = x'+y'\}|.$$

Theorem 13 is an easy consequence of the following structural lemma, which states that every pair of independent flat sources can be exactly partitioned into pairs of independent flat sources with low additive energy with slightly lower min-entropy. We prove this result by analyzing random partitions of the supports of the two independent sources.

Lemma 20. *Let $X, Y \subseteq \mathbb{F}_2^n$ be sets of size at least 2^k each. Then, we can partition X and Y into $T = 2^t$ subsets $(X_i)_{i \in [T]}$ and $(Y_j)_{j \in [T]}$, respectively, such that $|X_i|, |Y_j| = 2^{k-t}$ and $E(X_i, Y_j) \leq \ell^2 \cdot 2^{2(k-t)}$ for all pairs $(i,j) \in [T]^2$, provided that $\ell \geq 4$ and $t \geq \frac{k}{\ell - 1}(1 + \ell/2)$.*

In particular, this means that we can write $\mathbf{X}$ and $\mathbf{Y}$ uniformly distributed over X and Y, respectively, as convex combinations $\mathbf{X} = \sum_i p_i \mathbf{X}_i$ and $\mathbf{Y} = \sum_j q_j \mathbf{Y}_j$ for random variables $\mathbf{X}_i$ and $\mathbf{Y}_j$ uniformly distributed over sets X_i and Y_j, respectively, satisfying the properties detailed above.

[24] Mrazović actually shows the slightly stronger result $|\mathsf{bias}(f(X+Y)) - \mathsf{bias}(f(X^\star + Y^\star))| \leq \gamma$, but only the "one-sided" version, stated above, is actually needed/used.

[25] This can be seen by taking f in property (2), above, to be the extractor.

Proof. Fix $\ell \geq 4$ and $t \geq \frac{k}{\ell-1}(1+\ell/2)$. Note that these constraints guarantee that $t \leq k$. Consider randomly partitioning each of X and Y (without replacement) into $T = 2^t$ subsets $X_1, \ldots, X_T$ and $Y_1, \ldots, Y_T$, respectively, of size exactly 2^{k-t}. Define $A_w = \{(x,y) \in X \times Y : x + y = w\}$ and $A_w^{i,j} = \{(x,y) \in X_i \times Y_j : x + y = w\}$. We show that we will sample with positive probability partitions (X_i) and (Y_j) such that $|A_w^{i,j}| \leq \ell$ for all $w \in \mathbb{F}_2^n$ and pairs (i,j). In turn this implies that

$$E(X_i, Y_j) = \sum_{w \in \mathbb{F}_2^n} |A_w^{i,j}|^2 \leq \ell^2 \cdot 2^{2(k-t)},$$

since the set $A_w^{i,j}$ is non-empty for at most $|X_i| \cdot |Y_j| = 2^{2(k-t)}$ choices of w, as desired.

Fix a vector $w \in \mathbb{F}_2^n$ and distinct pairs $(x^{(1)}, y^{(1)}), \ldots, (x^{(\ell)}, y^{(\ell)}) \in A_w$. Note that, since $x^{(i)} + y^{(i)} = x^{(j)} + y^{(j)}$, it follows that $x^{(1)}, \ldots, x^{(\ell)}$ are pairwise distinct, and the same holds for $y^{(1)}, \ldots, y^{(\ell)}$. Therefore, the probability that the pairs $(x^{(1)}, y^{(1)}), \ldots, (x^{(\ell)}, y^{(\ell)})$ all end up in the same product set $X_i \times Y_j$ for some (i,j) is at most[26]

$$\left(\frac{1}{T^2}\right)^{\ell-1} = 2^{-2t(\ell-1)}.$$

For each vector $w \in \mathbb{F}_2^n$, define $s_w = |A_w|$, and, for pairs $(x^{(a)}, y^{(a)})_{a \in [\ell]} \in A_w^\ell$, let $\mathbf{Z}_{(x^{(a)}, y^{(a)})_{a \in [\ell]}}$ be the indicator random variable of the event that $(x^{(a)}, y^{(a)}) \in X_i \times Y_j$ for all $a \in [\ell]$ and some indices $i, j \in [T]$. Then, taking the expectation over the sampling of the partitions $X_1, \ldots, X_T$ and $Y_1, \ldots, Y_T$, we have

$$\begin{aligned}
\mathbb{E}\left[\sum_{w \in \mathbb{F}_2^n} \sum_{(x^{(i)}, y^{(i)})_{i \in [\ell]} \in A_w^\ell} \mathbf{Z}_{(x^{(i)}, y^{(i)})_{i \in [\ell]}}\right] &\leq \sum_{w \in \mathbb{F}_2^n} \binom{s_w}{\ell} 2^{-2t(\ell-1)} \\
&\leq \sum_{w \in \mathbb{F}_2^n : s_w \neq 0} \binom{2^k}{\ell} 2^{-2t(\ell-1)} \\
&< 2^{2k} \cdot 2^{k\ell} \cdot 2^{-2t(\ell-1)} \\
&\leq 1.
\end{aligned}$$

The second inequality uses the fact that $s_w \leq 2^k$ for all $w \in \mathbb{F}_2^n$. The third inequality holds because $\binom{2^k}{\ell} < 2^{k\ell}$ when $\ell \geq 4$ and because $\sum_{w \in \mathbb{F}_2^n} s_w \leq 2^{2k}$, meaning that there are at most 2^{2k} vectors w such that $s_w \neq 0$. The last inequality follows from the choice $t \geq \frac{k}{\ell-1}(1 + \ell/2)$. This implies that there must exist a partition of X and Y into $X_1, \ldots, X_T$ and $Y_1, \ldots, Y_T$, respectively, such that the associated sum inside the expectation is 0, which means that $|A_w^{i,j}| \leq \ell$ for all $i, j \in [T]$, as desired. □

This lemma guarantees that to extract from any k-sumset source it suffices to extract from any k-sumset source $\mathbf{W} = \mathbf{X} + \mathbf{Y}$ such that $\mathbf{X}$ and $\mathbf{Y}$ are uniformly distributed over sets X and Y, respectively, of slightly smaller size and such that the additive energy $E(X,Y)$ is appropriately small. The proof of Theorem 13 is then a consequence of Lemma 20, the following lemma of Chattopadhyay and Liao [CL22] (which shows that random functions extract from sumset sources with low additive energy), and a union bound.

Lemma 21 ([CL22, Lemma B.2])**.** *For any independent flat k-sources $\mathbf{X}, \mathbf{Y} \sim \mathbb{F}_2^n$, a uniformly random function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ satisfies $f(\mathbf{X} + \mathbf{Y}) \approx_\varepsilon \mathbf{U}_1$ with probability at least $1 - 2 \cdot 2^{-\frac{2\varepsilon^2 2^{4k}}{E(X,Y)}}$.*

[26] This uses the fact that $\frac{a-1}{b-1} \leq \frac{a}{b}$ whenever $a \leq b$.

We are now ready to prove Theorem 13.

Proof of Theorem 13. Without loss of generality, fix a k-sumset $\mathbf{X}+\mathbf{Y} \sim \mathbb{F}_2^n$ with $\mathbf{X}$ and $\mathbf{Y}$ flat k-sources. Set $\ell = 1+k(1/2-\alpha)$ for an arbitrary constant $\alpha > 0$ and $t \geq \frac{k}{\ell-1}(1+\ell/2) = (1+\beta)\frac{k}{2}+O(1)$. According to Lemma 20, we can then write

$$\mathbf{X}+\mathbf{Y} = \sum_{i,j\in[T]} p_{i,j}(\mathbf{X}_i+\mathbf{Y}_j),$$

where $(\mathbf{X}_i, \mathbf{Y}_j)$ are independent flat k'-sources with $k' = k-t = (1-\beta)\frac{k}{2} - O(1)$ and supports X_i and Y_j satisfying $E(X_i, Y_j) \leq \ell^2 \cdot 2^{2(k-t)} = \ell^2 \cdot 2^{2k'}$, for $T = 2^t$ and some $(p_{i,j})_{i,j\in[T]}$. Therefore, it suffices to show that a uniformly random function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ satisfies $f(\mathbf{X}'+\mathbf{Y}') \approx_\varepsilon \mathbf{U}_1$ simultaneously for all pairs $(\mathbf{X}', \mathbf{Y}')$ of independent flat k'-sources with $E(\mathbf{X}', \mathbf{Y}') \leq \ell^2 2^{2k'}$ with the desired probability.

Lemma 21 guarantees that, for a fixed pair $(\mathbf{X}', \mathbf{Y}')$ as above, $f(\mathbf{X}'+\mathbf{Y}') \approx_\varepsilon \mathbf{U}_1$ holds for a uniformly random function f with probability at least

$$1 - 2 \cdot 2^{-\frac{2\varepsilon^2 2^{4k'}}{E(\mathbf{X}', \mathbf{Y}')}} \geq 1 - 2 \cdot 2^{-\frac{2\varepsilon^2 2^{2k'}}{\ell^2}}.$$

Since there are at most $\binom{2^n}{2^{k'}}^2 \leq 2^{2n2^{k'}}$ such pairs of sources, a union bound over all these pairs shows that a uniformly random function satisfies the desired property except with probability at most

$$2^{2n2^{k'}} \cdot 2 \cdot 2^{-\frac{2\varepsilon^2 2^{2k'}}{\ell^2}} = 2 \cdot 2^{2^{k'}\left(2n - \frac{2\varepsilon^2 2^{k'}}{\ell^2}\right)}. \tag{20}$$

Setting $k = (2+\gamma)(\log n + 2\log(1/\varepsilon)) + C$ for $\gamma > 0$ an arbitrarily small constant and $C > 0$ a sufficiently large constant gives that the term in Equation (20) is strictly smaller than 1. □

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il