



# Low-Degree Polynomials Are Good Extractors

Omar Alrabiah   

University of California, Berkeley, CA, USA

Jesse Goodman   

The University of Texas at Austin, TX, USA

Jonathan Mosheiff   

Ben-Gurion University of the Negev, Be'er-Sheva, Israel

João Ribeiro   

Instituto de Telecomunicações and Departamento de Matemática, Instituto Superior Técnico, Universidade de Lisboa, Portugal

---

## 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  $\varepsilon$ -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).

**2012 ACM Subject Classification** Theory of computation  $\rightarrow$  Pseudorandomness and derandomization

**Keywords and phrases** randomness extractors, low-degree polynomials, local sources, polynomial sources, variety sources, sumset sources, sumset extractors, Reed-Muller codes, lower bounds

**Digital Object Identifier** 10.4230/LIPIcs.APPROX/RANDOM.2025.38

**Category** RANDOM

**Related Version** *Full Version:* <https://eccc.weizmann.ac.il/report/2024/093/>

**Funding** *Omar Alrabiah:* Supported by a Saudi Arabian Cultural Mission (SACM) Scholarship, NSF Award CCF-2210823, and a Simons Investigator Award (Venkatesan Guruswami).

*Jesse Goodman:* Supported by a Simons Investigator Award (#409864, David Zuckerman).

*Jonathan Mosheiff:* Supported by Israel Science Foundation grant 3450/24, an Alon Fellowship, and DOE grant #DE-SC0024124.

*João Ribeiro:* Supported by NOVA LINC (ref. UIDB/04516/2020), FCT/MECI through national funds and when applicable co-funded EU funds under UID/50008: Instituto de Telecomunicações, and DOE grant #DE-SC0024124.



© Omar Alrabiah, Jesse Goodman, Jonathan Mosheiff, and João Ribeiro;  
licensed under Creative Commons License CC-BY 4.0

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2025).

Editors: Alina Ene and Eshan Chattopadhyay; Article No. 38; pp. 38:1–38:25



Leibniz International Proceedings in Informatics

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

**Acknowledgements** We thank Alexander Golovnev, Zeyu Guo, Pooya Hatami, Satyaajeet 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 authors were visiting NTT Research in Sunnyvale, CA. The fourth author was based at NOVA LINCS and NOVA School of Science and Technology for most of the duration of this work.

## 1 Introduction

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

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

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  $\alpha_S$  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  $\mathbb{F}_2^n$ ,

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

For these definitions of *random* and *bias*, Ben-Eliezer, Hod, and Lovett [7] – building on a pair of earlier papers [50, 3] – provided a complete answer to the above question. In particular, they showed sharp concentration bounds on  $|\text{bias}(f)|$ , 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 [7, 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  $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$ ,*

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

They showed these bounds were tight, and a later work extended these results to all prime fields [6].

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

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

Given this definition, a simple observation is that  $\text{bias}_{\mathbf{U}_n}(f) = \text{bias}(f)$ , and thus establishing concentration bounds for  $|\text{bias}_{\mathbf{X}}(f)|$  is a strictly more general problem than doing so for  $|\text{bias}(f)|$ . But how should we use this generality? And for what distributions  $\mathbf{X}$  is it actually interesting to understand  $|\text{bias}_{\mathbf{X}}(f)|$ ? Recall that we would like to understand  $|\text{bias}_{\mathbf{X}}(f)|$  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 0** (Randomness extractor). *Let  $\mathcal{X}$  be a family of sources  $\mathbf{X} \sim \{0,1\}^n$ . A deterministic function  $\text{Ext} : \{0,1\}^n \rightarrow \{0,1\}^m$  is an extractor for  $\mathcal{X}$  with error  $\varepsilon$  if for any  $\mathbf{X} \in \mathcal{X}$ ,*

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

where  $\Delta(\cdot, \cdot)$  denotes statistical distance. For short, we also call  $\text{Ext}$  an  $\varepsilon$ -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 [61, 39].

In this paper, we will focus on extractors that output one bit.<sup>1</sup> In this case, the requirement in Definition 0 reduces to a requirement that for any  $\mathbf{X} \in \mathcal{X}$ , it holds that  $|\Pr[\text{Ext}(\mathbf{X}) = 1] - 1/2| \leq \varepsilon$ . Or in other words,

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

Thus, returning to our original discussion, we see that getting good bounds on  $|\text{bias}_{\mathbf{X}}(f)|$  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  $|\text{bias}_{\mathbf{X}}(f)|$ , 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 source  $\mathbf{X} \sim \mathbb{F}_2^n$  (equivalently, in order to upper bound  $|\text{bias}_{\mathbf{X}}(f)|$ ), 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.

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

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

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 [23].<sup>2</sup> 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 [61, 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  $\mathbf{X}$ ?*

If we can show that  $|\text{bias}_{\mathbf{X}}(f)|$  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 [7], affine sources [25], and local sources [4].<sup>3</sup> 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* [18], 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  $\mathbf{A}, \mathbf{B} \sim \mathbb{F}_2^n$  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 [18, 14, 19], including: independent sources [23], interleaved sources [59], and small-space sources [48] (and affine sources [8], 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

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

<sup>3</sup> 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  $\mathbf{X} = f(\mathbf{U}_m)$ , where  $f$  is some function where each output bit depends on a bounded number of input bits.

blocks in the construction of key cryptographic [56, 62] and pseudorandom primitives [53, 54].<sup>4</sup> Their study has also led to important structural results for well-studied families of distributions [25, 4]. Because of this, low-complexity extractors have received a lot of attention in the literature [33, 56, 62, 63, 26, 53, 11, 35, 25, 21, 4, 47, 22, 54], with the works of Cohen and Tal [25] and Alrabiah, Chattopadhyay, Goodman, Li, and Ribeiro [4] 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. For example, the work of Ben-Eliezer, Hod, and Lovett [7] (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 [25] (on *low-degree extractors for affine sources*) showed that every low-degree polynomial must have a big subspace in its solution set. And the work of Alrabiah, Chattopadhyay, Goodman, Li, and Ribeiro [4] (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., [45]). In more detail, extractors are known to exhibit strong lower bounds against a variety of well-studied complexity classes, including low-depth circuits [46, 38], general circuits [28, 37, 34, 38, 52], various flavors of branching programs [46, 17, 42, 20], and more [24, 38, 40, 15]. 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 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. 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 [61, Proposition 6.12], and (2) a generalization of the result that a random low-degree polynomial has low bias [7, Lemma 1.2] (Theorem 0).

<sup>4</sup> In fact, several well-known explicit extractors are *themselves* low-degree polynomials [23, 5, 8, 13, 31, 51]! The most canonical example is the *inner-product extractor* [23], which works for independent sources and, more generally, sumset sources (see Section A).

► **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  $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$ ,*

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

We highlight some key aspects of this result. First, it has a simple proof, which follows by combining [7, Lemma 1.2] (Theorem 0) with the leftover hash lemma [44]. Second, it is easy to verify that it is tight.<sup>5</sup> 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.<sup>6</sup>

We note that by a standard application of the XOR lemma [30, Lemma 3.8], it is easy to extend Theorem 1 to show that a *sequence* of independent, uniformly random degree  $d$  polynomials  $f_1, f_2, \dots, f_m : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$  can be concatenated to create a multi-bit extractor for  $\mathbf{X}$ .<sup>7</sup> In fact, this can further be extended to show that the sequence  $f_1, f_2, \dots, f_m$  not only extracts  $m$  uniform bits from  $\mathbf{X}$ , but has low correlation with any (short) fixed function  $g$  applied to  $\mathbf{X}$ .<sup>8</sup> 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. [49, 1, 2]), 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” [16, 25].

► **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  $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$  is a  $2^{-ck/d}$ -extractor for  $\mathcal{X}$ , except with probability at most  $2^{-c \binom{k}{\leq d}}$ . In particular, we get the following for a sufficiently large constant  $C > 0$  depending only on  $\delta$ .*

■ **Local sources:** *There exists a degree  $\leq d$  polynomial  $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$  that is a  $2^{-ck/d}$ -extractor for  $r$ -local sources with min-entropy*

$$k \geq Cd(2^r n + rn \log n)^{1/d}.$$

<sup>5</sup> 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.

<sup>6</sup> Even though arbitrary  $(n, k)$ -sources are known to be convex combinations of flat  $(n, k)$ -sources [61, Lemma 6.10], this convex combination may end up bringing the source  $\mathbf{X}$  out of the “small family”  $\mathcal{X}$ .

<sup>7</sup> 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  $\mathbf{X}$ .

<sup>8</sup> 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.

- **Polynomial sources:** *There exists a degree  $\leq d$  polynomial  $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$  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  $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$  that is a  $2^{-ck/d}$ -extractor for degree  $r$  variety sources with min-entropy*

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

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 [4, Theorem 1.1], which required min-entropy  $k \geq C 2^r r^2 d (2^r n \log n)^{1/d}$  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 [32, 16], 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 that our result on variety sources may be useful for establishing fine-grained complexity separations between low-degree polynomials and other models of computation, given known hardness results for variety extractors [46, 38]. Finally, we note that in the full version, we show that Corollary 1 (in fact, Theorem 0) can be used to prove low-degree polynomials are good *linear seeded extractors*.<sup>9</sup>

## Concurrent work

In a concurrent and independent work, Golovnev, Guo, Hatami, Nagargoje, and Yan [36] 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 [50, Theorem 1.5]. The differences are as follows: our result achieves better error (exponentially small vs. polynomially small),<sup>10</sup> and we also establish results for sumset sources (discussed in Section 1.1.2 below).<sup>11</sup> On the other hand, the work [36] proves a significant generalization of [50, 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* – 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  $k \geq C d (n/\varepsilon^2)^{1/\lfloor d/2 \rfloor}$ , a random degree  $d$  polynomial  $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$  is an  $\varepsilon$ -extractor for  $(n, k)$ -sumset sources, with probability at least  $1 - 2^{-\varepsilon^2 \binom{k/C}{2 \lfloor d/2 \rfloor}} \geq 1 - 2^{-n^2/\varepsilon^2}$ .*

<sup>9</sup> 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.

<sup>10</sup> Our exponentially small error is crucial to our strict improvement over the main positive result in [4].

<sup>11</sup> We find the sumset setting to be much more challenging, and view these results as the main *technical* contribution of this work.

We highlight a few key specializations of Theorem 2 – focusing on even  $d$ , for simplicity. First, we note that in the *dispenser* regime,<sup>12</sup> it 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 [25] 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 the full version, we show that 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*  $\varepsilon = n^{-\gamma}$ , it shows that there exist degree  $\leq d$  polynomials that are  $\varepsilon$ -extractors for sumset sources with min-entropy  $k = O(dn^{2(1+2\gamma)/d})$ . Third, in the *arbitrary error regime*  $\varepsilon > 0$ , it shows that a random degree  $d = O(\log(n/\varepsilon))$  polynomial  $f$  is an  $\varepsilon$ -extractor for sumset sources with min-entropy  $k = O(\log(n/\varepsilon))$ . This strengthens the existential result of Mrazović [57], who obtained such a min-entropy requirement for a uniformly random function  $f$ .

Finally, 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 computational model [42], 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).<sup>13</sup> We observe that by leveraging standard results on finite fields [55, Lemma 6.21] (see also [12, Lemma 17]), linear ROBPs of constant width  $w = O(1)$  can compute any polynomial of degree 2.<sup>14</sup> On the other hand, Theorem 2 (combined with [20, 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.<sup>15</sup>

### 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, in the full version, we use this to build new “evasive sets,” which we then use to improve the extractor impossibility results of Chattopadhyay, Goodman, and Gurumukhani [16], 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 [41].

<sup>12</sup> A *dispenser* is an extractor  $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$  with nontrivial error  $\varepsilon < 1/2$ .

<sup>13</sup> The formal definition is slightly more technical - see [20, Definition 2].

<sup>14</sup> In more detail, [12, Lemma 17] asserts that for any quadratic  $q : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$ , there is some  $B \in \mathbb{F}_2^{m \times n}$  with full row rank and some affine  $L : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$  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  $\leq 2$  bits of storage. Then, using  $\leq 1$  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.

<sup>15</sup> Indeed, one might expect the width  $w$  to somehow grow proportionately with the degree  $d$  of the polynomial that must be computed. However, this shows that the width jumps from constant to *exponential*, simply by moving from degree 2 to degree 4.

## A novel reduction between extractors

The second key ingredient in the proof of Theorem 2 is a new type of reduction between extractors. While most reductions 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 using a careful (*correlated*) *randomized process* can make this task much easier. In the full version, 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 [57].

## 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  $|\text{bias}_{\mathbf{X}}(f)| = |\mathbb{E}_{x \sim \mathbf{X}}[(-1)^{f(x)}]|$  with  $\mathbf{X}$  an arbitrary  $(n, k)$ -source. Our simple argument combines the original result of [7] 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  $\text{eval}_d(x)$  the tuple of evaluations of all monomials of degree at most  $d$  on  $x$ , i.e.,

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

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

In order to obtain the desired concentration, it suffices to appropriately upper bound the high-order moments  $\mathbb{E}[\text{bias}_{\mathbf{X}}(f)^t]$  for a large  $t$ , which is also the approach followed in [7]. And, also analogously to [7], it is not hard to show that

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

Intuitively, we would like to reduce the task of bounding  $\mathbb{E}_f[\text{bias}_{\mathbf{X}}(f)^t]$  to the task of bounding  $\mathbb{E}_g[\text{bias}_{L(\mathbf{X})}(g)^t]$ , which can be handled via the concentration bounds from [7]. We establish such a reduction via the leftover hash lemma, which guarantees the existence of a *linear* map  $L : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^m$  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

$$\mathbb{E}_f[\text{bias}_{\mathbf{X}}(f)^t] \leq \mathbb{E}_g[\text{bias}_{L(\mathbf{X})}(g)^t] \quad (2)$$

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

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

where the inequality uses the fact that  $q \circ L : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$  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  $\mathbb{E}_g[\text{bias}_{L(\mathbf{X})}(g)^t] \approx \mathbb{E}_g[\text{bias}(g)^t]$ . Moreover, we can upper bound  $\mathbb{E}_g[\text{bias}(g)^t]$  via the known concentration bound from [7]. Combining this with Equation (2) yields the desired upper bound on  $\mathbb{E}_f[\text{bias}_{\mathbf{X}}(f)^t]$ , which we translate into a concentration bound on  $|\text{bias}_{\mathbf{X}}(f)|$  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  $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$  will satisfy  $f(\text{supp}(\mathbf{W})) = \mathbb{F}_2$  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

$$\Pr_f[f(X + Y) \equiv 0] \leq 2^{-\text{rank}(\text{eval}_d(X + Y))}. \quad (3)$$

This is because we may write  $f(x) = \langle v, \text{eval}_d(x) \rangle$  for a uniformly random vector  $v \in \mathbb{F}_2^{\binom{n}{\leq d}}$ , 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  $\text{eval}_d(x)$  and  $\text{eval}_d(y)$  are linearly independent.

Given Equation (3), it is clear that we must understand  $\text{rank}(\text{eval}_d(X + Y))$ . 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  $\text{rank}(\text{eval}_d(X + Y))$  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  $\text{rank}(\text{eval}_d(X' + Y'))$  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 2** (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  $\text{rank}(\text{eval}_d(A' + B'))$  is roughly  $\binom{k}{\leq d}$ .*

We sketch how Claim 2 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  $\text{eval}_d(X' + Y')$  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 2, it is instructive to consider existing results of a similar flavor. Keevash and Sudakov [50] (and later Ben-Eliezer, Hod, and Lovett [7, 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  $\text{rank}(\text{eval}_d(S')) = |S'|$ . This lemma is one of the most important steps in the proof of the main result of [7], 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  $g : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^k$  that is surjective on  $S$  and satisfies

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

Applying this same rank-preserving surjection in the setting of Claim 2 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 [7] with a *linear* map, guaranteed by the leftover hash lemma. In fact, by the leftover hash lemma, there exists a linear map  $L : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^{k'}$ , 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

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

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

## 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  $\text{rank}(\text{eval}_d(X + Y)) = |X| \cdot |Y|$ , 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 uniform over  $\mathbb{F}_2$ , with high probability over  $f$ . This happens because, under the conditions above,  $\mathbf{W}$  is a flat source and all the vectors in  $\text{eval}_d(X + Y)$  are linearly independent, which means that the bits  $(f(w) = \langle v, \text{eval}_d(w) \rangle)_{w \in X + Y}$  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,  $\text{rank}(\text{eval}_d(X + Y)) = |X| \cdot |Y|$ , 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  $\varepsilon$ -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  $B_1 = \{u_1, \dots, u_t\}$  be the set of weight- $d/2$  vectors supported on  $\{1, \dots, k/3\}$ , and let  $B_2 = \{v_1, \dots, v_t\}$  be the set of weight- $d/2$  vectors supported on  $\{2k/3 + 1, \dots, k\}$ . 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

$$\text{rank eval}_d(B_1 + B_2) = |B_1| \cdot |B_2| = \left(\frac{k/3}{d/2}\right)^2.$$

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

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

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  $\mathbf{X}, \mathbf{Y}$  independent and uniformly distributed over  $\mathbb{F}_2^k$ . If  $\mathbf{L} \in \mathbb{F}_2^{k \times k}$  were a uniformly random matrix, 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

We begin with some 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  $\text{supp}(\mathbf{X})$  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  $\text{wt}(x)$  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  $\binom{n}{\leq r} := \sum_{i=0}^r \binom{n}{i}$ .

### 3.1 Probability

We now 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  $\varepsilon$ -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 3** (Data-processing inequality). *For any random variables  $\mathbf{X}, \mathbf{Y} \sim V$  and function  $f : V \rightarrow W$ , it holds that*

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

Next, we define the notion of min-entropy, and various types of sources.

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

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

► **Definition 3** ( $(n, k)$ -source). *We say that  $\mathbf{X} \sim \mathbb{F}_2^n$  is an  $(n, k)$ -source if  $H_\infty(\mathbf{X}) \geq k$ .*

► **Definition 3** ( $(n, k)$ -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'll also need the following “dependency reversal” lemma from [15].

► **Lemma 3** (Dependency reversal [15]). *For any random variable  $\mathbf{X} \sim X$  and deterministic function  $f : X \rightarrow Y$ , there exists an independent random variable  $\mathbf{A} \sim A$  and deterministic function  $g : Y \times A \rightarrow X$  such that*

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

and such that  $g(\cdot, a)$  is a pseudoinverse<sup>16</sup> of  $f$ , for all  $a \in \text{supp}(\mathbf{A})$ .

### 3.2 Extractors

Finally, we collect some basic definitions of randomness extractors and useful auxiliary results.

<sup>16</sup>We say that  $g' : Y \rightarrow X$  is a *pseudoinverse* of  $f : X \rightarrow Y$  if  $f(g'(f(x))) = f(x)$  for all  $x \in X$ .

► **Definition 3** (Strong seeded extractor). A function  $\text{Ext} : \mathbb{F}_2^n \times \mathbb{F}_2^s \rightarrow \mathbb{F}_2^m$  is called a  $(k, \varepsilon)$ -strong seeded extractor if for every  $(n, k)$ -source  $\mathbf{X}$  it holds that

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

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

The leftover hash lemma, stated below, is a crucial ingredient in our proofs.

► **Lemma 3** (Leftover Hash Lemma [44]). For every  $0 < k < n$ ,  $\varepsilon > 0$ , and  $m \leq k - 2 \log(1/\varepsilon)$ , there exists an explicit linear  $(k, \varepsilon)$ -strong seeded extractor  $\text{Ext} : \mathbb{F}_2^n \times \mathbb{F}_2^n \rightarrow \mathbb{F}_2^m$ .

We'll also need the following (special case of) the leftover hash lemma with  $\ell_\infty$  guarantees.

► **Lemma 3** ([29, 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  $L : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^m$  with output length  $m = \lfloor k - C \log k \rfloor$  are surjective when restricted to  $S$ .

We now define various types of extractors and dispersers that will appear throughout.

► **Definition 3** (Two-source disperser). A function  $\text{Disp} : \mathbb{F}_2^n \times \mathbb{F}_2^n \rightarrow \mathbb{F}_2$  is called a  $k$ -two-source disperser if for any two independent  $(n, k)$ -sources  $\mathbf{X}, \mathbf{Y}$  it holds that  $\text{supp}(\text{Disp}(\mathbf{X}, \mathbf{Y})) = \mathbb{F}_2$ .

► **Definition 3** (Two-source extractor). A function  $\text{Ext} : \mathbb{F}_2^n \times \mathbb{F}_2^n \rightarrow \mathbb{F}_2^m$  is called a  $(k, \varepsilon)$ -two-source extractor if for any two independent  $(n, k)$ -sources  $\mathbf{X}, \mathbf{Y}$  it holds that

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

► **Definition 3** (Sumset disperser). A function  $\text{Disp} : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$  is called a  $k$ -sumset disperser if for any  $(n, k)$ -sumset source  $\mathbf{W}$  it holds that  $\text{supp}(\text{Disp}(\mathbf{W})) = \mathbb{F}_2$ .

► **Definition 3** (Sumset extractor). A function  $\text{Ext} : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^m$  is called a  $(k, \varepsilon)$ -sumset extractor if for any  $(n, k)$ -sumset source  $\mathbf{W}$  it holds that

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

Note that every  $k$ -sumset disperser  $\text{Disp}$  automatically gives a  $k$ -two-source disperser  $\text{Disp}'$ , simply by setting  $\text{Disp}'(x, y) := \text{Disp}(x + y)$ . The same holds for 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  $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$ ,

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

We'll use the following characterization of the bias's moments, whose statement and proof are analogous to those of [7, Claim 2.1] (which only focuses on uniform input.)

► **Lemma 3** (Simple extension of [7, Claim 2.1]). *If  $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$  is a random degree  $d$  polynomial, it holds that*

$$\mathbb{E}_f [\text{bias}_{\mathbf{X}}(f)^t] = \Pr_{x^{(1)}, \dots, x^{(t)} \sim \mathbf{X}} [\forall p : \mathbb{F}_2^n \rightarrow \mathbb{F}_2, \deg(p) \leq d : p(x^{(1)}) + \dots + p(x^{(t)}) = 0]$$

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

With this lemma in hand, we are ready to prove Theorem 1.

**Proof of Theorem 1.** The proof follows along the same lines as the proof of [7, Lemma 1.2], combined with a linear hashing trick. For an integer  $t > 0$ , we focus on bounding the  $t$ -th moment of  $\text{bias}_{\mathbf{X}}(f)$ . By Lemma 3, we have that

$$\mathbb{E}_f [\text{bias}_{\mathbf{X}}(f)^t] = \Pr_{x^{(1)}, \dots, x^{(t)} \sim \mathbf{X}} [\forall p : \mathbb{F}_2^n \rightarrow \mathbb{F}_2, \deg(p) \leq d : p(x^{(1)}) + \dots + p(x^{(t)}) = 0].$$

Fix any linear map  $L : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^m$ . Observe that the above is

$$\leq \Pr_{x^{(1)}, \dots, x^{(t)} \sim \mathbf{X}} [\forall p : \mathbb{F}_2^m \rightarrow \mathbb{F}_2, \deg(p) \leq d : p(L(x^{(1)})) + \dots + p(L(x^{(t)})) = 0],$$

since  $\deg(p \circ L) \leq \deg(p) \leq d$ , where  $p \circ L$  denotes the composition of the polynomial  $p$  and the linear map  $L$ . Then, applying Lemma 3 again, the above is exactly

$$\begin{aligned} &= \Pr_{w^{(1)}, \dots, w^{(t)} \sim L(\mathbf{X})} [\forall p : \mathbb{F}_2^m \rightarrow \mathbb{F}_2, \deg(p) \leq d : p(w^{(1)}) + \dots + p(w^{(t)}) = 0] \\ &= \mathbb{E}_g [\text{bias}_{L(\mathbf{X})}(g)^t], \end{aligned}$$

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

$$\mathbb{E}_f [\text{bias}_{\mathbf{X}}(f)^t] \leq \mathbb{E}_g [\text{bias}_{L(\mathbf{X})}(g)^t] \quad (4)$$

for all linear maps  $L$ .

Now, 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  $L : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^m$  with  $m = k(1 - 2c/d)$  such that

$$L(\mathbf{X}) \approx_{2^{-ck/d}} \mathbf{U}_m. \quad (5)$$

Note that, by our choice of  $c$ , we have that  $m \geq (1 - \delta)k \geq d$ . Equation (5) implies that

$$\Pr_g [|\text{bias}_{L(\mathbf{X})}(g)| > 2^{-\frac{ck}{2d}+1}] \leq \Pr_g [|\text{bias}(g)| > 2^{-\frac{ck}{2d}+1} - 2^{-\frac{ck}{d}}] \leq 2^{-c\binom{m}{\leq d}}. \quad (6)$$

The last inequality follows from Theorem 0 applied to  $\mathbb{F}_2^m$  because  $2^{-\frac{ck}{2d}+1} - 2^{-\frac{ck}{d}} \geq 2^{-\frac{ck}{2d}} \geq 2^{-cm/d}$ , since  $m = k(1 - 2c/d) \geq k/2$  (as we enforced that  $c < 1/4$ ). Now, since  $\binom{m}{\leq d} = \binom{k(1-2c/d)}{\leq d} \geq \alpha \binom{k}{\leq d}$  for some constant  $\alpha = \alpha(c, \delta) > 0$ , we get from Equation (6) that

$$\Pr_g [|\text{bias}_{L(\mathbf{X})}(g)| > 2^{-c_1 k/d}] \leq 2^{-c_1 \binom{k}{\leq d}} \quad (7)$$

for some absolute constant  $c_1 > 0$ .

We now bound the  $t$ -th moment  $\mathbb{E}_f [\text{bias}_{\mathbf{X}}(f)^t]$  for an appropriate  $t$ . Set  $t = c_2 \cdot \frac{d}{k} \cdot \binom{k}{\leq d}$  for a sufficiently large constant  $c_2 = c_2(c_1) > 0$ . Using Equations (4) and (7), we have that

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

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

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

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

## 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 [27, 64, 4]. A source  $\mathbf{X} \sim \mathbb{F}_2^n$  is said to be  $r$ -local if  $\mathbf{X} = g(\mathbf{U}_m)$ , where  $g : \mathbb{F}_2^m \rightarrow \mathbb{F}_2^n$  is some function such that each output bit depends on at most  $r$  input bits and  $\mathbf{U}_m$  denotes the uniform distribution over  $\mathbb{F}_2^m$ . In the full version, we show that by combining Theorem 1 with a simple upper bound on the number of local sources, we immediately obtain the following.

► **Corollary 3** (Low-degree polynomials extract from local sources). *There exist constants  $C, c > 0$  such that for all  $n, k, d, r \in \mathbb{N}$  with  $k \geq Cdn^{1/d}(r \log n + 2^r)^{1/d}$ , the following holds. A random degree  $d$  polynomial  $f$  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 3 both improves on and simplifies the proof of [4, 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 was obtained via an intricate initial reduction to local “non-oblivious bit-fixing (NOBF)” sources. In particular, observe that when  $d > r$ , we now get that the min-entropy requirement is just  $O(d(n \log n)^{1/d})$ . Previously, this was only possible for  $d \gg 2^r$ .

It is also instructive to compare our improved result with the lower bound from [4, 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 3 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  $p_1, \dots, p_n : \mathbb{F}_2^m \rightarrow \mathbb{F}_2$  of degree at most  $r$  for some positive integer  $m$  (the *input length*) such that  $\mathbf{X} = P(\mathbf{U}_m)$ , where  $P = (p_1, \dots, p_n)$  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 [32, 11, 43, 16]. In the full version, we show that by combining Theorem 1 with a recently-established *input reduction lemma* for polynomial sources [16], we immediately obtain the following.

► **Corollary 3** (Low-degree polynomials extract from polynomial sources). *There exist constants  $C, c > 0$  such that for all  $n, k, d, r \in \mathbb{N}$  with  $d > r$  and  $k \geq C \left( \frac{C^r d^d n}{r^r} \right)^{\frac{1}{d-r}}$ , the following holds. A random degree  $d$  polynomial  $f$  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  $1 - 2^{-c \binom{k}{\leq d}}$  over  $f$ . In particular, if we take  $d = 2r$ , then the above holds for any  $k \geq Cdn^{2/d}$ .*

## 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  $p_1, \dots, p_t : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$  of degree at most  $r$  such that  $\mathbf{X}$  is uniformly distributed over the variety

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

This family of sources has received significant interest over the last decade, both over  $\mathbb{F}_2$  and larger fields [31, 60, 51, 43]. Sufficiently strong explicit extractors for variety sources are known to imply breakthrough circuit lower bounds [38]. In the full version, we show that by combining Theorem 1 with a known *input reduction lemma* for variety sources [25], we immediately obtain the following.

► **Corollary 3** (Low-degree polynomials extract from variety sources). *There exist constants  $C, c > 0$  such that for all  $n, k, d, r \in \mathbb{N}$  with  $d > r$  and  $k \geq Cdn^{\frac{r+1}{d}}$ , the following holds. A random degree  $d$  polynomial  $f$  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  $f$ .*

The best explicit extractors for variety sources over  $\mathbb{F}_2$  either work for constant degree  $r$  and min-entropy  $k = (1 - c_r)n$  [51], or large degree  $r = n^\alpha$  and very high min-entropy  $n - n^\beta$  with  $\alpha + \beta < 1/2$  [60]. By [38], an explicit version of Corollary 3 would be more than enough to imply significantly improved circuit lower bounds.

## 5 Low-degree polynomials extract from sumset sources

### 5.1 Low-degree polynomials disperse from sumset sources

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

► **Proposition 3.** *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  $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$  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 2) states that for any two subsets  $A$  and  $B$ , we can find small subsets  $A' \subseteq A$  and  $B' \subseteq B$  such that  $\text{rank}(\text{eval}_d(A' + B'))$  is large.

► **Lemma 3.** *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  $\text{rank}(\text{eval}_d(A' + B')) \geq \binom{k - C \log n}{\leq d}$ .*

In order to prove this result, the following simple claim will come in handy. It says that applying a linear map to a set can only decrease its “eval<sub>d</sub>-rank.”

► **Claim 3.** *If  $L : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^m$  is linear, then  $\text{rank}(\text{eval}_d(S)) \geq \text{rank}(\text{eval}_d(L(S))) \forall S \subseteq \mathbb{F}_2^n$ .*

This claim’s proof is straightforward, and we include it in the full version. Next, by combining this claim with the leftover hash lemma, we can prove Lemma 3.

**Proof of Lemma 3.** Fix arbitrary sets  $A, B \subseteq \mathbb{F}_2^n$  of size  $2^k$ . By Lemma 3 and a union bound, there exists a linear map  $L : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^t$  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

### 38:18 Low-Degree Polynomials Are Good Extractors

$L(A'), L(B') = \mathcal{B}_{d/2}^t(0)$ . Since  $L$  is linear, we have that  $L(A' + B') = L(A') + L(B') = \mathcal{B}_d^t(0)$ , and so

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

where the first inequality uses Claim 3, and the last equality uses the fact that  $\mathcal{B}_d^t(0)$  is an interpolating set for degree  $\leq d$  polynomials (see, e.g., [58, Proposition 6.21]). ◀

Finally, we can now use Lemma 3 to prove Proposition 3.

**Proof of Proposition 3.** 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 3, there exist subsets  $A' \subseteq A$  and  $B' \subseteq B$  of size  $\binom{t}{\leq d/2}$  such that

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

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

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

By taking a union bound over all of the at most  $\binom{2^n}{\leq d/2}^2 \leq 2^{2n \binom{t}{\leq d/2}}$  choices of  $A'$  and  $B'$ , we conclude that the probability that  $f$  is constant on some set  $A + B$  is at most

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

By combining the guaranteed bounds on  $k, d$  (from the hypothesis) with standard estimates on binomial coefficients, it is straightforward to show that this is at most

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

which completes the proof. (For detailed calculations, see the full version.) ◀

## 5.2 Low-degree polynomials extract from sumset sources

We now prove Theorem 2 in full generality. For convenience, we restate it, below.

► **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  $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$  is an  $\varepsilon$ -extractor for  $(n, k)$ -sumset sources, with probability at least  $1 - 2^{-\varepsilon^2 \binom{k/C}{2 \lfloor d/2 \rfloor}} \geq 1 - 2^{-n^2/\varepsilon^2}$ .*

In order to prove Theorem 2, we start by defining a special type of sumset sources.

► **Definition 3.** *We say that a sumset source  $\mathbf{W} = \mathbf{X} + \mathbf{Y}$  has full  $\text{eval}_d$ -rank if the set  $\text{eval}_d(\text{supp}(\mathbf{X}) + \text{supp}(\mathbf{Y}))$  is a collection of  $|\text{supp}(\mathbf{X})| \cdot |\text{supp}(\mathbf{Y})|$  linearly independent vectors.*

As it turns out, random low-degree polynomials can easily be shown to extract from sumset sources with full  $\text{eval}_d$ -rank. Indeed, a straightforward application of the Chernoff (and union) bounds yields the following (for a proof, see the full version).

► **Lemma 3.** *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  $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$ . For any sets  $X, Y \subseteq \mathbb{F}_2^n$  of size  $t$  satisfying  $\text{rank}(\text{eval}_d(X + Y)) = t^2$ , we have*

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

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

Next, we show that every sumset source is close to a convex combination of sumset sources with full  $\text{eval}_d$ -rank. This is the main ingredient in our proof.

► **Lemma 3.** *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  $\mathbf{W}^* = \mathbf{X}^* + \mathbf{Y}^*$  with full  $\text{eval}_d$ -rank and such that  $|\text{supp}(\mathbf{X}^*)| = |\text{supp}(\mathbf{Y}^*)| = \binom{k/6}{\lfloor d/2 \rfloor}$ .*

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

$$\begin{aligned} \text{Ext}(\mathbf{X}) &\approx_{2\varepsilon} \mathbf{U}_m, \\ \text{Ext}(\mathbf{Y}) &\approx_{2\varepsilon} \mathbf{U}_m. \end{aligned} \tag{8}$$

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 3), there exist functions  $\text{Ext}_X^{-1}$  and  $\text{Ext}_Y^{-1}$  and random variables  $\mathbf{A} = (\mathbf{A}_0, \dots, \mathbf{A}_t)$  and  $\mathbf{B} = (\mathbf{B}_0, \dots, \mathbf{B}_t)$  such that

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

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

Now, by Equation (8) 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  $\text{Ext}(\mathbf{X}), \text{Ext}(\mathbf{Y})$  with them. Recalling that  $\mathbf{X}, \mathbf{Y}$  are independent, we get from an application of the data-processing inequality and Equation (8) that both

$$\begin{aligned} &\text{Pick}(\text{Ext}_X^{-1}(\text{Ext}(\mathbf{X}), \mathbf{A}_0), \dots, \text{Ext}_X^{-1}(\text{Ext}(\mathbf{X}), \mathbf{A}_t); \mathbf{U}) \\ &\approx_{2\varepsilon} \text{Pick}(\text{Ext}_X^{-1}(\mathbf{R}, \mathbf{A}_0), \dots, \text{Ext}_X^{-1}(\mathbf{R}, \mathbf{A}_t); \mathbf{U}), \text{ and} \\ &\text{Pick}(\text{Ext}_Y^{-1}(\text{Ext}(\mathbf{Y}), \mathbf{B}_0), \dots, \text{Ext}_Y^{-1}(\text{Ext}(\mathbf{Y}), \mathbf{B}_t); \mathbf{U}') \\ &\approx_{2\varepsilon} \text{Pick}(\text{Ext}_Y^{-1}(\mathbf{R}', \mathbf{B}_0), \dots, \text{Ext}_Y^{-1}(\mathbf{R}', \mathbf{B}_t); \mathbf{U}'). \end{aligned} \tag{9}$$

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

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

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  $\text{rank}(\text{eval}_d(\mathcal{B}_0 + \mathcal{B}_1)) = |\mathcal{B}_0| \cdot |\mathcal{B}_1|$ .

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

$$\begin{aligned} & \left( \text{Pick}(\text{Ext}_X^{-1}(\mathbf{R}, \mathbf{A}_0), \dots, \text{Ext}_X^{-1}(\mathbf{R}, \mathbf{A}_t); \mathbf{U}), \right. \\ & \quad \left. \text{Pick}(\text{Ext}_Y^{-1}(\mathbf{R}', \mathbf{B}_0), \dots, \text{Ext}_Y^{-1}(\mathbf{R}', \mathbf{B}_t); \mathbf{U}') \right) \\ & \approx_{m2^{-m/3}} \left( \text{Pick}(\text{Ext}_X^{-1}(\mathbf{L}u_0, \mathbf{A}_0), \dots, \text{Ext}_X^{-1}(\mathbf{L}u_t, \mathbf{A}_t); \mathbf{U}), \right. \\ & \quad \left. \text{Pick}(\text{Ext}_Y^{-1}(\mathbf{L}v_0, \mathbf{B}_0), \dots, \text{Ext}_Y^{-1}(\mathbf{L}v_t, \mathbf{B}_t); \mathbf{U}') \right) := (\mathbf{X}^*, \mathbf{Y}^*). \end{aligned} \quad (10)$$

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  $\text{supp}(u_i) \cup \text{supp}(v_j)$ , 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  $\text{eval}_d$ -rank of  $\mathbf{X}^* + \mathbf{Y}^*$ . Consider any fixing of the random variables  $\mathbf{L}$  and  $\mathbf{A}_0, \dots, \mathbf{A}_t, \mathbf{B}_0, \dots, \mathbf{B}_t$ . Upon such a fixing,  $(\mathbf{X}^*, \mathbf{Y}^*)$  becomes of the form  $(\mathbf{X}^*, \mathbf{Y}^*)$ , where  $\mathbf{X}^*, \mathbf{Y}^*$  are independent and uniform over the sets

$$\begin{aligned} \mathbf{X}^* &:= \{\text{Ext}_X^{-1}(\mathbf{L}u_0, a_0), \dots, \text{Ext}_X^{-1}(\mathbf{L}u_t, a_t)\}, \\ \mathbf{Y}^* &:= \{\text{Ext}_Y^{-1}(\mathbf{L}v_0, b_0), \dots, \text{Ext}_Y^{-1}(\mathbf{L}v_t, b_t)\}, \end{aligned}$$

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

$$S^* := \{\text{Ext}_X^{-1}(\mathbf{L}u_i, a_i) + \text{Ext}_Y^{-1}(\mathbf{L}v_j, b_j)\}_{i,j \in [t]}.$$

To analyze the  $\text{eval}_d$ -rank of  $S^*$ , recall from Claim 3 that applying linear transformations can only decrease the  $\text{eval}_d$ -rank. Furthermore, note that for any  $i, j \in [t]$  it holds that

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

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

$$\begin{aligned} \text{rank}(\text{eval}_d(S^*)) &\geq \text{rank}(\text{eval}_d(L^{-1}(\text{Ext}(S^*)))) \\ &= \text{rank}(\text{eval}_d(\{u_i + v_j\}_{i,j})) = \text{rank}(\text{eval}_d(\mathcal{B}_0 + \mathcal{B}_1)). \end{aligned}$$

Since  $|S^*| \leq |\mathcal{B}_0| \cdot |\mathcal{B}_1|$ , we get that  $\text{rank}(\text{eval}_d(S^*)) = |\mathcal{B}_0| \cdot |\mathcal{B}_1|$ , and so  $\mathbf{X}^* + \mathbf{Y}^*$  has full  $\text{eval}_d$ -rank. Recalling Equations (9) and (10) and the parameter settings  $m = k/2$  and  $\varepsilon = 2^{-k/4}$ , this means that the sumset source  $\mathbf{X} + \mathbf{Y}$  is  $\varepsilon^*$ -close to a convex combination of flat sumset sources  $\mathbf{X}^* + \mathbf{Y}^*$  with full  $\text{eval}_d$ -rank and support sizes  $|\text{supp}(\mathbf{X}^*)|, |\text{supp}(\mathbf{Y}^*)| = \binom{m/3}{\lfloor d/2 \rfloor} = \binom{k/6}{\lfloor d/2 \rfloor}$ , where  $\varepsilon^* = 4\varepsilon + m2^{-m/3} \leq 4 \cdot 2^{-k/4} + (k/2) \cdot 2^{-k/6} \leq 5k \cdot 2^{-k/6}$ . 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  $k \leq C$ , this completes the proof.  $\blacktriangleleft$

As we have seen, random low-degree polynomials easily extract from sumset sources with full  $\text{eval}_d$ -rank (Lemma 3), and every sumset source is close to a convex combination of sumset sources with full  $\text{eval}_d$ -rank (Lemma 3). As a result, we immediately obtain Theorem 2. For the detailed calculations, we refer the reader to the full version.

## 6 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.<sup>17</sup> We think that even showing this for small linear min-entropy would already be quite interesting.
- In the full version, we show that there exist (non-explicit) degree  $\leq 4$  low-error two-source extractors for any linear min-entropy via approximate duality [10]. This approach, however, provably cannot go below min-entropy  $\sqrt{n}$  [9]. Can we show the existence of low-degree low-error two-source extractors for min-entropy below  $\sqrt{n}$ ?

---

## References

- 1 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. doi:10.1109/TIT.2015.2462817.
- 2 Emmanuel Abbe, Amir Shpilka, and Min Ye. Reed–Muller codes: Theory and algorithms. *IEEE Transactions on Information Theory*, 67(6):3251–3277, 2020. doi:10.1109/TIT.2020.3004749.
- 3 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. doi:10.1007/978-3-540-85363-3\_22.
- 4 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. doi:10.4230/LIPIcs.ICALP.2022.10.
- 5 László Babai, Noam Nisan, and Mária Szegedy. Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *Journal of Computer and System Sciences*, 45(2):204–232, 1992. doi:10.1016/0022-0000(92)90047-M.
- 6 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. doi:10.1137/18M1215104.
- 7 Ido Ben-Eliezer, Rani Hod, and Shachar Lovett. Random low-degree polynomials are hard to approximate. *Comput. Complex.*, 21(1):63–81, 2012. doi:10.1007/s00037-011-0020-6.
- 8 Eli Ben-Sasson, Shlomo Hoory, Eyal Rozenman, Salil Vadhan, and Avi Wigderson. Extractors for affine sources. Unpublished manuscript, 2001.

---

<sup>17</sup>Note that Theorem 2 can handle min-entropy  $k = O(dn^{1/\lfloor d/2 \rfloor})$  if  $d \geq O(\log(1/\varepsilon))$ .

- 9 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. doi:10.1109/FOCS.2012.39.
- 10 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. doi:10.1137/12089003X.
- 11 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. doi:10.1145/2422436.2422496.
- 12 Andrej Bogdanov and Emanuele Viola. Pseudorandom bits for polynomials. *SIAM Journal on Computing*, 39(6):2464–2486, 2010. doi:10.1137/070712109.
- 13 Jean Bourgain. More on the sum-product phenomenon in prime fields and its applications. *International Journal of Number Theory*, 01(01):1–32, 2005. doi:10.1142/S1793042105000108.
- 14 Eshan Chattopadhyay and Jesse Goodman. Improved extractors for small-space sources. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS 2021)*, pages 610–621. IEEE, 2021. doi:10.1109/FOCS52979.2021.00066.
- 15 Eshan Chattopadhyay and Jesse Goodman. Leakage-resilient extractors against number-on-forehead protocols. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing (STOC 2025)*, pages 604–614, 2025. doi:10.1145/3717823.3718272.
- 16 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. doi:10.4230/LIPIcs.ITCS.2024.28.
- 17 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. doi:10.4230/LIPIcs.ITCS.2022.40.
- 18 Eshan Chattopadhyay and Xin Li. Extractors for sumset sources. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing (STOC 2016)*, pages 299–311, 2016. doi:10.1145/2897518.2897643.
- 19 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. doi:10.1145/3519935.3519963.
- 20 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. doi:10.4230/LIPIcs.CCC.2023.9.
- 21 Kuan Cheng and Xin Li. Randomness extraction in  $AC^0$  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. doi:10.4230/LIPIcs.APPROX-RANDOM.2018.37.
- 22 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. doi:10.4230/LIPIcs.APPROX-RANDOM.2024.69.
- 23 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. doi:10.1137/0217015.
- 24 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. doi:10.1145/2840728.2840734.

- 25 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. doi:10.4230/LIPIcs.APPROX-RANDOM.2015.680.
- 26 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. doi:10.1007/978-3-642-03685-9\_35.
- 27 Anindya De and Thomas Watson. Extractors and lower bounds for locally samplable sources. *ACM Trans. Comput. Theory*, 4(1), March 2012. doi:10.1145/2141938.2141941.
- 28 Evgeny Dement'kov 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. doi:10.1007/978-3-642-22993-0\_25.
- 29 Manik Dhar and Zeev Dvir. Linear hashing with  $\ell_\infty$  guarantees and two-sided Kakeya bounds. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS 2022)*, pages 419–428, 2022. doi:10.1109/FOCS54457.2022.00047.
- 30 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. doi:10.1137/120868414.
- 31 Zeev Dvir. Extractors for varieties. In *2009 24th Annual IEEE Conference on Computational Complexity (CCC 2009)*, pages 102–113, 2009. doi:10.1109/CCC.2009.7.
- 32 Zeev Dvir, Ariel Gabizon, and Avi Wigderson. Extractors and rank extractors for polynomial sources. *Comput. Complex.*, 18(1):1–58, 2009. doi:10.1007/S00037-009-0258-4.
- 33 Stefan Dziembowski and Ueli Maurer. Tight security proofs for the bounded-storage model. In *Thirty-Fourth Annual ACM Symposium on Theory of Computing (STOC 2002)*, pages 341–350. Association for Computing Machinery, 2002. doi:10.1145/509907.509960.
- 34 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. doi:10.1109/FOCS.2016.19.
- 35 Oded Goldreich, Emanuele Viola, and Avi Wigderson. On randomness extraction in  $AC^0$ . In *Proceedings of the 30th Conference on Computational Complexity (CCC 2015)*, pages 601–668. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2015.
- 36 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. doi:10.4230/LIPIcs.APPROX/RANDOM.2024.41.
- 37 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. doi:10.1145/2840728.2840755.
- 38 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. doi:10.4230/LIPIcs.ITCS.2021.24.
- 39 Jesse Goodman. *Seedless Extractors*. PhD thesis, Cornell University, 2023.
- 40 Jesse Goodman and Vipul Goyal. Two-source extractors don't shrink. Unpublished manuscript, 2025.
- 41 William Timothy Gowers, Ben Green, Freddie Manners, and Terence Tao. On a conjecture of Marton. *Annals of Mathematics*, 201(2):515–549, 2025.

- 42 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. doi:10.4230/LIPIcs.CCC.2022.4.
- 43 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*, pages 46–59, New York, NY, USA, 2023. Association for Computing Machinery. doi:10.1145/3564246.3585109.
- 44 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. doi:10.1137/S0097539793244708.
- 45 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. doi:10.1145/3564246.3585216.
- 46 Pavel Hrubes and Anup Rao. Circuits with medium fan-in. In *30th Conference on Computational Complexity (CCC 2015)*. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2015.
- 47 Xuanguo 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. doi:10.4230/LIPIcs.APPROX/RANDOM.2022.9.
- 48 Jesse Kamp, Anup Rao, Salil Vadhan, and David Zuckerman. Deterministic extractors for small-space sources. *J. Comput. Syst. Sci.*, 77(1):191–220, 2011. doi:10.1016/j.jcss.2010.06.014.
- 49 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. doi:10.1109/TIT.2012.2184841.
- 50 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. doi:10.1137/S0895480103434634.
- 51 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. doi:10.4230/LIPIcs.APPROX-RANDOM.2019.72.
- 52 Jiayu 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. doi:10.1145/3519935.3519976.
- 53 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. doi:10.1109/CCC.2011.27.
- 54 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. doi:10.4230/LIPIcs.CCC.2024.10.
- 55 Rudolf Lidl and Harald Niederreiter. *Finite Fields*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2 edition, 1997.
- 56 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. doi:10.1007/3-540-45708-9\_17.
- 57 Rudi Mrazović. Extractors in Paley graphs: A random model. *European Journal of Combinatorics*, 54:154–162, 2016. doi:10.1016/j.ejc.2015.12.009.
- 58 Ryan O’Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.

- 59 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. doi:10.1016/j.jcss.2010.06.013.
- 60 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. doi:10.1109/FOCS.2016.29.
- 61 Salil Vadhan. Pseudorandomness. *Foundations and Trends in Theoretical Computer Science*, 7(1–3):1–336, 2012. doi:10.1561/04000000010.
- 62 Salil P. Vadhan. Constructing locally computable extractors and cryptosystems in the bounded-storage model. *Journal of Cryptology*, 17:43–77, 2004. doi:10.1007/S00145-003-0237-X.
- 63 Emanuele Viola. The complexity of constructing pseudorandom generators from hard functions. *computational complexity*, 13(3-4):147–188, 2005. doi:10.1007/S00037-004-0187-1.
- 64 Emanuele Viola. Extractors for circuit sources. *SIAM Journal on Computing*, 43(2):655–672, 2014. doi:10.1137/11085983X.

## A Inner product as a sumset extractor

We show that the inner product function (a degree 2 polynomial) is a sumset extractor for min-entropy  $k > n/2$ .

► **Theorem 4.** *For any even  $n$  and  $n/2 \leq k \leq n$ , the function  $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$  given by  $f(x) = \langle (x_1, \dots, x_{n/2}), (x_{n/2+1}, \dots, x_n) \rangle$  is a  $(k, \varepsilon)$ -sumset extractor for  $\varepsilon \leq 4 \cdot 2^{\frac{1}{2}(n-2k)}$ .*

To prove Theorem 4, we only need the fact that inner product is a two-source extractor.

► **Lemma 4** ([23]). *Suppose that  $\mathbf{X}$  is an  $(n, k_1)$ -source and  $\mathbf{Y}$  is an  $(n, k_2)$ -source, and that  $\mathbf{X}$  and  $\mathbf{Y}$  are independent. Then,  $\langle \mathbf{X}, \mathbf{Y} \rangle \approx_\varepsilon \mathbf{U}_1$  for  $\varepsilon \leq 2^{\frac{1}{2}(n-k_1-k_2)}$ .*

**Proof of Theorem 4.** Without loss of generality, fix any two independent flat  $(n, k)$ -sources  $\mathbf{X}, \mathbf{Y}$ . For a vector  $w \in \mathbb{F}_2^n$ , define  $w^{(1)} := (w_1, \dots, w_{n/2})$  and  $w^{(2)} := (w_{n/2+1}, \dots, w_n)$ .

We first carry out the analysis assuming  $\langle \mathbf{X}^{(1)}, \mathbf{X}^{(2)} \rangle$  and  $\langle \mathbf{Y}^{(1)}, \mathbf{Y}^{(2)} \rangle$  are both constant. We will then remove this constraint by increasing the final error. For any  $b, b' \in \mathbb{F}_2$ , let  $\mathbf{X}_b := (\mathbf{X} \mid \langle \mathbf{X}^{(1)}, \mathbf{X}^{(2)} \rangle = b)$  and  $\mathbf{Y}_{b'} := (\mathbf{Y} \mid \langle \mathbf{Y}^{(1)}, \mathbf{Y}^{(2)} \rangle = b')$ , and denote  $k_{\mathbf{X},b} := H_\infty(\mathbf{X}_b)$  and  $k_{\mathbf{Y},b'} := H_\infty(\mathbf{Y}_{b'})$ . Then, defining  $\text{rev}(y) := (y^{(2)}, y^{(1)})$  for any  $y$ , we have that

$$\begin{aligned} f(x+y) &= \langle x^{(1)} + y^{(1)}, x^{(2)} + y^{(2)} \rangle = \langle x^{(1)}, x^{(2)} + y^{(2)} \rangle + \langle y^{(1)}, x^{(2)} + y^{(2)} \rangle \\ &= \langle x^{(1)}, y^{(2)} \rangle + \langle y^{(1)}, x^{(2)} \rangle + \langle x^{(1)}, x^{(2)} \rangle + \langle y^{(1)}, y^{(2)} \rangle = \langle x, \text{rev}(y) \rangle + b + b' \end{aligned}$$

for all  $x \in \text{supp}(\mathbf{X}_b), y \in \text{supp}(\mathbf{Y}_{b'})$ . Note that  $\mathbf{X}_b$  is a flat  $(n, k_{\mathbf{X},b})$ -source and  $\text{rev}(\mathbf{Y}_{b'})$  is a flat  $(n, k_{\mathbf{Y},b'})$ -source, and they are independent. So, by Lemma 4, we know that  $f(\mathbf{X}_b, \mathbf{Y}_{b'})$  is  $\varepsilon_{b,b'}$ -close to uniform with  $\varepsilon_{b,b'} \leq 2^{\frac{1}{2}(n-k_{\mathbf{X},b}-k_{\mathbf{Y},b'})}$ . Thus, by convexity, we have that

$$\begin{aligned} \Delta(f(\mathbf{X} + \mathbf{Y}), \mathbf{U}_1) &\leq \sum_{b,b' \in \mathbb{F}_2} \Pr[\langle \mathbf{X}^{(1)}, \mathbf{X}^{(2)} \rangle = b] \cdot \Pr[\langle \mathbf{Y}^{(1)}, \mathbf{Y}^{(2)} \rangle = b'] \cdot \varepsilon_{b,b'} \\ &= \sum_{b,b' \in \mathbb{F}_2} \frac{2^{k_{\mathbf{X},b}}}{2^k} \cdot \frac{2^{k_{\mathbf{Y},b'}}}{2^k} \cdot \varepsilon_{b,b'} \leq \sum_{b,b' \in \mathbb{F}_2} 2^{\frac{1}{2}(n+k_{\mathbf{X},b}+k_{\mathbf{Y},b'}-4k)} \\ &\leq \sum_{b,b' \in \mathbb{F}_2} 2^{\frac{1}{2}(n-2k)} = 4 \cdot 2^{\frac{1}{2}(n-2k)}. \end{aligned} \quad \blacktriangleleft$$