

Hilbert Functions and Low-Degree Randomness Extractors

Alexander Golovnev* Zeyu Guo† Pooya Hatami‡ Satyajeet Nagargoje§
Chao Yan¶

Abstract

For $S \subseteq \mathbb{F}^n$, consider the linear space of restrictions of degree- d polynomials to S . The Hilbert function of S , denoted $h_S(d, \mathbb{F})$, is the dimension of this space. We obtain a tight lower bound on the smallest value of the Hilbert function of subsets S of arbitrary finite grids in \mathbb{F}^n with a fixed size $|S|$. We achieve this by proving that this value coincides with a combinatorial quantity, namely the smallest number of low Hamming weight points in a down-closed set of size $|S|$.

Understanding the smallest values of Hilbert functions is closely related to the study of degree- d closure of sets, a notion introduced by Nie and Wang (Journal of Combinatorial Theory, Series A, 2015). We use bounds on the Hilbert function to obtain a tight bound on the size of degree- d closures of subsets of \mathbb{F}_q^n , which answers a question posed by Doron, Ta-Shma, and Tell (Computational Complexity, 2022).

We use the bounds on the Hilbert function and degree- d closure of sets to prove that a random low-degree polynomial is an extractor for samplable randomness sources. Most notably, we prove the existence of low-degree extractors and dispersers for sources generated by constant-degree polynomials and polynomial-size circuits. Until recently, even the existence of arbitrary deterministic extractors for such sources was not known.

*Georgetown University. Email: alexgolovnev@gmail.com. Supported by the NSF CAREER award (grant CCF-2338730).

†The Ohio State University. Email: zguotcs@gmail.com.

‡The Ohio State University. Email: pooyahat@gmail.com. Supported by NSF grant CCF-1947546.

§Georgetown University. Email: satyajeetn2012@gmail.com. Supported by the NSF CAREER award (grant CCF-2338730).

¶Georgetown University. Email: cy399@georgetown.edu.

Contents

1	Introduction	1
1.1	Our Results on Hilbert Functions	3
1.1.1	Degree- d Closure of Sets	4
1.2	Our Results on Randomness Extractors	5
1.2.1	Low-Degree Dispersers	6
1.2.2	Low-Degree Extractors	7
1.3	Remarks	9
2	Preliminaries	10
2.1	Probability Distributions	10
2.2	Randomness Sources, Dispersers, and Extractors	11
2.3	Hilbert Functions and Standard Monomials	13
3	Hilbert Functions of Sets in Finite Grids	14
4	Number of Points with Low Hamming Weight in Down-Closed Sets	16
4.1	The Boolean Case, $q = 2$	17
4.2	The General Case of Finite Grids	19
5	A Tight Bound on the Size of Degree-d Closures of Sets	21
6	Low-Degree Dispersers	23
6.1	Dispersers for Small Families of Sources	23
6.2	Dispersers for Polynomial and Circuit Sources	24
7	Random Low-Degree Polynomials Extract from Fixed Sources	27
8	Low-Degree Extractors	31
8.1	Extractors Outputting Multiple Bits	32

1 Introduction

Hilbert Functions. Low-degree polynomials are fundamental objects in theoretical computer science, and their properties are extensively studied due to their important role in areas such as error correcting codes and circuit lower bounds. Let $d \geq 0$ be an integer, \mathbb{F} be a field, and $S \subseteq \mathbb{F}^n$ be a set. Each degree- d n -variate polynomial p over \mathbb{F} can be naturally viewed as a map $p : \mathbb{F}^n \rightarrow \mathbb{F}$, and hence also defines a map $p|_S : S \rightarrow \mathbb{F}$. Considering the linear space of all such maps in \mathbb{F}^S , which is a subspace of the space of all maps from S to \mathbb{F} , allows one to tap into a wide array of algebraic techniques in order to prove useful facts about the set S . This approach was for example utilized in complexity theory famously in the work of Smolensky [Smo87], where proving bounds on the dimension of the aforementioned subspace was used to obtain lower-bounds for $\text{AC}^0[\oplus]$ circuits computing the indicator function of the set S , for various $S \subseteq \{0,1\}^n$. The dimension of the space consisting of $p|_S$ for all degree- d polynomials p is indeed a well-studied and classical concept in algebraic geometry known as the (affine) Hilbert function of S , denoted by $h_S(d, \mathbb{F})$. Hilbert functions encode important geometric and algebraic information, such as the dimension, degree, and regularity of varieties, in a more general context.

Hilbert functions have previously been studied in complexity theory due to their applications in circuit lower bounds, in particular for $\text{AC}^0[\oplus]$ circuits, that were established by Smolensky [Smo87] and Razborov [Raz87]. Such applications require finding sets $S \subseteq \{0,1\}^n$ where the Hilbert function takes a very large value. However, it is also interesting to prove general lower bounds or find lower-bounding methods for arbitrary sets S . An example of such a result is the work of Moran and Rashtchian [MR16], who showed upper and lower bounds on $h_S(d, \mathbb{F})$ for $S \subseteq \{0,1\}^n \subseteq \mathbb{F}^n$ via various concepts in VC theory. [MR16] treated the Hilbert function as a complexity measure of the set S and compared it to measures that arise naturally in learning theory, including “shattering” and “strong shattering” values.

Suppose $r > 0$ is an integer. It is natural to wonder, what the extreme values of $h_S(d, \mathbb{F})$ are, among all sets S of size $|S| = r$. It is not hard to show that the maximum value is equal to $\min(r, h_{\mathbb{F}^n}(d, \mathbb{F}))$ when $S \subseteq \mathbb{F}^n$. For example, the maximum value of the Hilbert function of a set $S \subseteq \mathbb{F}_2^n$ of size r is $\min(r, \binom{n}{\leq d})$.

On the other hand, finding the true smallest value of $h_S(d, \mathbb{F})$ is a natural and intriguing question even when S is restricted to subsets of some finite and structured set in \mathbb{F}^n .

Question 1. *Let $0 \leq d \leq n$ be integers, \mathbb{F} be a field, and $A = A_1 \times \dots \times A_n \subseteq \mathbb{F}^n$ where $A_i \subseteq \mathbb{F}$ are finite sets. For any $r \leq |A|$, what is the smallest value of $h_S(d, \mathbb{F})$ among all subsets $S \subseteq A$ of cardinality $|S| = r$?*

This question has been answered in the case of $\mathbb{F} = \mathbb{F}_2$ and $A = \mathbb{F}_2^n$ by Keevash and Sudakov [KS05] and Ben-Eliezer, Hod, and Lovett [BEHL12], and later generalized to $\mathbb{F} = \mathbb{F}_p$ and $A = \mathbb{F}_p^n$ by Beame, Oveis Gharan, and Yan [BGY20]. For simplicity, let $r = p^k$ for some $k \geq 0$. [BGY20] proved that the smallest value of $h_S(d, \mathbb{F}_p)$ with $|S| = r$ is equal to the number of degree- $\leq d$ monomials on k variables, for example when $p = 2$, this is equal to simply $\binom{k}{\leq d} = \binom{\log_2 r}{\leq d}$.

We prove a more general result that answers **Question 1** for arbitrary finite grids $A \subseteq \mathbb{F}^n$ in arbitrary fields \mathbb{F} . We show that the smallest values of Hilbert functions are exactly determined by an extremal combinatorial question about the number of low-Hamming-weight elements in down-closed sets, which we solve by building on the work of Beelen and Datta [BD18].

The prior works discussed above were motivated by applications in bounding the list-size of the Reed-Muller codes and obtaining certain extensions of Frankl–Ray-Chaudhuri–Wilson theorems

on cross-intersecting sets. In contrast, in this paper, we are interested in [Question 1](#) due to its applications in pseudorandomness, particularly in randomness extraction.

Understanding the smallest values of Hilbert functions is closely related to the study of *degree- d closure* of sets, a notion introduced by Nie and Wang [[NW15](#)].

Definition 1.1. The degree- d closure of a set $T \subseteq \mathbb{F}^n$ is defined as

$$\text{cl}_d(T) := \{a \in \mathbb{F}^n \mid \text{for every degree-}d \text{ polynomial } f, f|_T \equiv 0 \Rightarrow f(a) = 0\} .$$

Equivalently, $\text{cl}_d(T)$ is the set of all points $a \in \mathbb{F}^n$ such that the values $f(a)$ of a degree- d polynomial f are determined by $f|_T$.

The existence of a small set with a large degree- d closure has application to hitting-set generators for polynomials [[DTT22](#)]. As an application of our answer to [Question 1](#), we obtain an upper bound on the size of $\text{cl}_d(T)$ in terms of $|T|$. Our bound in fact yields an optimal way of creating a small set with a large degree- d closure.

Futhermore, [Question 1](#) has direct implications to the theory of randomness extractors, which we discuss next.

Randomness Extractors. The theory of randomness extractors is an active research area that was initiated in [[SV86](#), [Blu86](#)] with the motivation of simulating randomized algorithms with access to “weak” randomness sources. The main objective of this theory is to design *extractors* that are capable of purifying imperfect randomness sources into high-quality random bits or bit sequences. Extractors and related objects such as *dispersers*, *samplers*, and *condensers* have since found numerous applications in constructing other pseudorandom objects such as pseudorandom generators [[NZ96](#)] and expander graphs [[WZ99](#)], as well as applications in other areas of theoretical computer science and mathematics including cryptography [[DOPS04](#)], combinatorics [[Li23](#)], hardness of approximation [[Zuc07](#)], error correcting codes [[TSZ01](#)], and metric embeddings [[Ind07](#)].

A deterministic extractor for a family \mathcal{X} of distributions over $\{0, 1\}^n$ is a map $f : \{0, 1\}^n \rightarrow \{0, 1\}^m$ such that for any $\mathbf{X} \in \mathcal{X}$, $f(\mathbf{X})$ is close to the uniform distribution in statistical distance. It is common to measure the amount of randomness in a random variable \mathbf{X} by its min-entropy, defined $H_\infty(\mathbf{X}) := -\log_2 \max_{x \in \{0, 1\}^n} \Pr[\mathbf{X} = x]$. It is easy to show that no deterministic extractor can extract from general n -bit randomness sources of min-entropy as high as $n - 1$ [[CG88](#)]. As a result, researchers in the area have explored two directions. Much of the focus in the area has been given to the more powerful *seeded extractors* that have access to an additional short purely random seed. This article contributes to another line of work that has extensively investigated the extra assumptions on the randomness sources that allow for explicit deterministic extractors and dispersers to exist. A widely studied class of sources in this latter direction, introduced in [[TV00](#)] is “samplable sources”, where the sources of randomness are distributions sampled by applying a low-complexity map (e.g., a decision forest, local map, NC^0 circuit, AC^0 circuit, an affine or a low-degree map) to the uniform distribution. Unfortunately, constructing explicit extractors even for sources samplable by really low-complexity maps has been quite challenging, and for example all the known constructions of extractors for local sources require quite high min-entropy of $\Omega(\sqrt{n})$ [[Vio14](#), [DW11](#)]. Due to the difficulty of constructing good explicit extractors and motivated by applications in complexity theory such as circuit lower bounds [[LY22](#)] and lower bounds for distribution-sampling [[Vio12](#)], researchers have considered the seemingly easier task of proving the existence of low-complexity extractors [[Vio05](#), [GVW15](#), [CL18](#), [Vad04](#), [BG13](#), [DY21](#), [HIV22](#), [CT15](#), [ACG+22](#)].

The state of affairs is much worse when it comes to randomness extraction from sources sampled by more powerful maps such as $\text{AC}^0[\oplus]$ or low-degree \mathbb{F}_2 -polynomial maps. In this case obtaining nontrivial explicit constructions and even non-explicit low-complexity extractors remains open. In fact, the same problems are open even in the case of dispersers. Here a map $f : \{0, 1\}^n \rightarrow \{0, 1\}$ is a *disperser* for \mathcal{X} if for every source $\mathbf{X} \in \mathcal{X}$, the support of $f(\mathbf{X})$ is $\{0, 1\}$. On the positive side, Chattopadhyay, Goodman, and Gurumukhani [CGG24], recently proved the existence of deterministic (not necessarily low-complexity) extractors for low-degree \mathbb{F}_2 -polynomial sources with logarithmic min-entropy.

1.1 Our Results on Hilbert Functions

We obtain our answer to [Question 1](#) by first reducing it to a purely combinatorial problem. In particular, via an algebraic geometric argument, we prove the following theorem which states that the minimum value of Hilbert functions over subsets of a grid is exactly captured by a combinatorial quantity related to down-closed sets. (A set $T \subseteq \mathbb{N}^n$ is said to be down-closed if T is closed under decreasing any coordinates of its elements.)

Theorem 1.2 (See [Corollary 3.3](#)). *Let \mathbb{F} be a field, and $A_1, \dots, A_n \subseteq \mathbb{F}$ be finite sets of size $|A_i| = r_i$. Define $A = A_1 \times \dots \times A_n$. For every $k \leq |A|$,*

$$\min_{S \subseteq A: |S|=k} h_S(d, \mathbb{F}) = \min_{\text{down-closed } T \subseteq F: |T|=k} |T_{\leq d}| ,$$

where $F = \prod_i \{0, \dots, r_i - 1\}$ and $T_{\leq d} = \{x \in T : \sum_i x_i \leq d\}$.

Let I be the ideal of $\mathbb{F}[X_1, \dots, X_n]$ associated with a set $S \subseteq A$, that is, the set of all polynomials vanishing on S . Classical results in algebraic geometry (such as Hilbert's Nullstellensatz) establish close connections between the structure of S and the structure of I , which allows us to focus on studying I .

The proof of [Theorem 1.2](#) is based on the idea that the ideal I can be reasonably approximated by another ideal, *the ideal of leading terms of I* . This approximation preserves important information about I , and consequently, about S as well. In particular, when the ideal of leading terms of I is defined with respect to a specific total order of monomials compatible with the total degree, it can be shown that such an approximation preserves the value of the Hilbert function. One advantage of working with the ideal of leading terms is that it is a *monomial ideal*, that is, an ideal generated by monomials, whose relatively simple structure can be analyzed using combinatorial tools.

We remark that the concept of transforming a general ideal into a monomial ideal is closely related to the theory of Gröbner bases, which serves as a basis of computational algebraic geometry. For a detailed discussion, see, e.g., [BM93]. This concept is also used in Smolensky's algebraic method for proving circuit lower bounds [Smo93].

[Theorem 1.2](#) allows us to reduce the problem of determining the smallest value of Hilbert function of a set of size k to understanding the smallest number of low-Hamming-weight points in down-closed sets of the same size. We then solve this combinatorial problem by proving that the minimum is obtained by the down-closed set $M_F(k)$ which is defined as the set of k lexicographically first elements of F .

Theorem 1.3 (See [Theorem 4.1](#)). *Let $1 \leq r_1 \leq \dots \leq r_n$ be integers and let $F = \prod_{i=1}^n \{0, \dots, r_i - 1\}$. Then*

$$\min_{\text{down-closed } T \subseteq F} |T_{\leq d}| = |M_F(k)_{\leq d}| .$$

In the case of $r_1 = \dots = r_n = 2$, we prove the above theorem via an elementary combinatorial argument, that via a series of operations turns any set of k elements into $M_F(k)$ without increasing the number of elements of Hamming weight $\leq d$. We prove the general case by building on a recent result of Beelen and Datta [BD18]. This result generalizes the work of Wei [Wei91] and Heijnen–Pellikaan [HP98, Hei99] in studying the generalized Hamming weights of certain linear codes.

We record the following corollary of our results specialized to finite fields which generalizes the bounds due to [KS05, BEHL12, BGY20].

Corollary 1.4 (See [Corollary 4.2](#)). *For every prime power q , and $n, k, d \in \mathbb{N}$ where $k \leq q^n$, we have*

$$\min_{S \subseteq \mathbb{F}_q^n: |S|=k} h_S(d, \mathbb{F}_q) = |M_q^n(k)_{\leq d}|.$$

In particular, setting $q = 2$, for every $n, k, d \in \mathbb{N}$ where $k \leq 2^n$, and every $S \subseteq \mathbb{F}_2^n$ of size $|S| = k$,

$$h_S(d, \mathbb{F}_2) \geq \binom{\lfloor \log(k) \rfloor}{\leq d}.$$

1.1.1 Degree- d Closure of Sets

Motivated by its applications to combinatorial geometry, the notion of degree- d closures of subsets of \mathbb{F}_q^n was introduced in [NW15]. This concept has since found further applications and connections to complexity theory [KS18, OSS19, Sri23] and pseudorandomness [DTT22].

Recall that the degree- d closure $\text{cl}_d(T)$ of a set $T \subseteq \mathbb{F}^n$ over a finite field \mathbb{F} is the set of all points $a \in \mathbb{F}^n$ such that any degree- d polynomial vanishing on T also vanishes at a . Nie and Wang [NW15] proved the following result.

Theorem 1.5 ([NW15, Theorem 5.6]). *Let $n, d \in \mathbb{N}$ and $T \subseteq \mathbb{F}_q^n$. Then*

$$|\text{cl}_d(T)| \leq \frac{q^n}{h_{\mathbb{F}_q^n}(d, \mathbb{F}_q)} \cdot |T|.$$

Building on our results on Hilbert functions, we obtain an improvement of [Theorem 1.5](#) by obtaining a tight upper bound on the size of degree- d closures of sets.

Theorem 1.6 (See [Theorem 5.2](#) and [Theorem 5.3](#)). *Let $n, d, m \in \mathbb{N}$. Let $T \subseteq \mathbb{F}_q^n$ be a set of size m . Then*

$$|\text{cl}_d(T)| \leq \max_{0 \leq k \leq q^n: |M_q^n(k)_{\leq d}| \leq m} k = \begin{cases} \max_{0 \leq k \leq q^n: |M_q^n(k)_{\leq d}| = m} k & \text{if } m \leq h_{\mathbb{F}_q^n}(d, \mathbb{F}_q), \\ q^n & \text{otherwise.} \end{cases} \quad (1)$$

Moreover, this bound is tight in the sense that for any $0 \leq m \leq q^n$, there exists $T \subseteq \mathbb{F}_q^n$ of size m for which (1) holds with equality.

In fact, the set T of size m that attains the bound in the above theorem can be constructed explicitly; see [Theorem 5.3](#) for details.

For convenience, we state the following corollary which is used later in the paper. For $n, d, \delta \in \mathbb{N}$, denote by $N(n, d, \delta)$ the number of monomials $X_1^{e_1} \dots X_n^{e_n}$ with $e_1, \dots, e_n \leq \delta$ and $e_1 + \dots + e_n \leq d$.

Corollary 1.7. *Let $n, d, \ell \in \mathbb{N}$. If $T \subseteq \mathbb{F}_q^n$ is a set of size less than $N(\ell, d, q-1)$, then $|\text{cl}_d(T)| < q^\ell$. In particular, if $q = 2$ and $T \subseteq \mathbb{F}_2^n$ is a set of size less than $\binom{\ell}{\leq d}$, then $|\text{cl}_d(T)| < 2^\ell$.*

Proof. Observe that $|\mathbb{M}_q^n(q^\ell)_{\leq d}| = N(\ell, d, q-1)$. Then apply [Theorem 1.6](#). □

Let us compare our bound with the bound of Nie and Wang in some specific settings.

Example 1.8. *Suppose $\ell \leq n$. Let $T \subseteq \mathbb{F}_2^n$ be a set of size $\binom{\ell}{\leq d} - 1$. Then by [Corollary 1.7](#), we have the bound $|\text{cl}_d(T)| \leq 2^\ell - 1$. On the other hand, the bound of Nie and Wang ([Theorem 1.5](#)) gives*

$$|\text{cl}_d(T)| \leq \frac{2^n}{\binom{n}{\leq d}} \cdot |T|,$$

which is exponential in n , rather than in ℓ , at least when $d \leq (\frac{1}{2} - c)n$ for some constant $c > 0$.

Example 1.9. *Suppose $\ell \leq n$ and $d < q$. Let $T \subseteq \mathbb{F}_q^n$ be a set of size $N(\ell, d, q-1) - 1 = \binom{\ell+d}{d} - 1$. Then by [Corollary 1.7](#), we have the bound $|\text{cl}_d(T)| \leq q^\ell - 1$, which is exponential in $\ell \log q$. On the other hand, the bound of Nie and Wang ([Theorem 1.5](#)) gives*

$$|\text{cl}_d(T)| \leq \frac{q^n}{\binom{n+d}{d}} \cdot |T|,$$

which is exponential in $n \log q$, rather than in $\ell \log q$, at least when $n + d \leq q^{1-c}$ for some constant $c > 0$.

In [\[DTT22\]](#), Doron, Ta-Shma, and Tell explicitly asked if there exists a small set $T \subseteq \mathbb{F}_q^n$ whose degree- d closure is very large. Our [Theorem 1.6](#) gives an upper bound on the size of the degree- d closure of T in terms of the size of T , which is tight in the sense that there exist sets T that exactly meet this bound for every cardinality of T . Moreover, such sets T can be constructed explicitly (see [Theorem 5.3](#)). Thus, we completely resolve the question posed by Doron, Ta-Shma, and Tell.

1.2 Our Results on Randomness Extractors

Continuing the line of work on low-complexity extractors, in this paper we investigate the power of low-degree polynomials in randomness extraction.

Question 2. *For which families \mathcal{X} of sources does there exist a low-degree disperser? Similarly, for which families \mathcal{X} of sources does there exist a low-degree extractor?*

Let us first discuss the easier task of obtaining low-degree dispersers before moving on to our main application of low-degree extractors. For simplicity, we will focus on the most important case of extracting randomness over \mathbb{F}_2 , but all our results easily generalize to \mathbb{F}_q . Non-explicit constructions of low-degree dispersers can be obtained via understanding the probability that a random low-degree polynomial is a disperser for a family \mathcal{X} of distributions over $\{0, 1\}^n$ which we identify with \mathbb{F}_2^n in the natural way. Our starting point is the observation that the notion of Hilbert functions can be used to exactly describe the probability that a random degree- d polynomial $f : \{0, 1\}^n \rightarrow \{0, 1\}$ is a disperser for a fixed source $\mathbf{X} \in \mathcal{X}$. Indeed, this probability is exactly equal to $1 - 2^{1-h_S(d, \mathbb{F}_2)}$, where $S = \text{support}(\mathbf{X})$. Thus, in particular, [Corollary 1.4](#) can be used to bound the probability that a random degree- d polynomial is not a disperser for a fixed source.

1.2.1 Low-Degree Dispersers

Let \mathcal{X} be a family of sources of min-entropy $\geq k$. Observing that the support of any distribution $\mathbf{X} \in \mathcal{X}$ is of size $\geq 2^k$, one gets as an immediate corollary of [Corollary 1.4](#), the existence of low-degree dispersers \mathcal{X} as long as $|\mathcal{X}|$ is small.

Theorem 1.10 (Informal, see [Corollary 6.2](#)). *Let $n, d, k \geq 1$. Let \mathcal{X} be a family of distributions of min-entropy $\geq k$. Then a random degree- d polynomial over \mathbb{F}_2 is a disperser for \mathcal{X} with probability at least*

$$1 - |\mathcal{X}| \cdot 2^{1 - \binom{k}{\leq d}}.$$

This theorem itself implies the existence of low-degree dispersers for several interesting families of samplable sources such as sources sampled by local maps, bounded-depth decision forests, and polynomial-sized bounded-fan-in circuits, to name a few.

A map $f : \{0, 1\}^m \rightarrow \{0, 1\}^n$ is called ℓ -local if each of its output bits depends on at most ℓ input bits. A depth- ℓ decision forest is a map f where each output bit can be computed as a depth- ℓ decision tree. It is easy to obtain an upper bound exponential in $\text{poly}(n)$ on the number of local or decision forest sources. Hence we get the following as a corollary of [Theorem 1.10](#).

Corollary 1.11 (Informal, see [Corollary 6.3](#)). *Let $1 \leq \ell \leq d \leq n$ be integers. There exists a degree- d disperser*

- *for the family of ℓ -local sources on $\{0, 1\}^n$ with min-entropy $k > d(2^\ell n + 2\ell n \log n)^{1/d}$.*
- *for the family of depth- ℓ decision forest sources on $\{0, 1\}^n$ with min-entropy $k > d((\ell + \log n)2^{\ell+1}n)^{1/d}$.*

As mentioned above, since in addition to the min-entropy requirement, the only requirement in [Theorem 1.10](#) about the family \mathcal{X} is a bound on $|\mathcal{X}|$, it can be used to immediately obtain low-degree dispersers for various other families of sources as well. For example, since for any c , the number of Boolean circuits with $\leq n^c$ bounded fan-in gates is at most $2^{O(n^{c+1})}$, one can also use [Theorem 1.10](#) to obtain a degree- $O(c)$ disperser for such families of circuits. However, we will not do an exhaustive search for all such applications, and instead our main disperser applications will focus on two powerful families of sources, namely sources sampled by low-degree polynomials over \mathbb{F}_2 and $\text{AC}[\oplus]$ circuits which we define as the family of unbounded-depth polynomial-size Boolean circuits with AND, OR, XOR, NOT gates of unbounded fan-in, where the input gates are not counted towards the size.

Note that low-degree polynomial maps $f : \{0, 1\}^m \rightarrow \{0, 1\}^n$, even affine ones, can depend on the entire input for any $m \gg n$ and thus one cannot simply bound $|\mathcal{X}|$ when \mathcal{X} is the family of sources sampled by low-degree polynomials. This property holds for $\text{AC}[\oplus]$ circuits, as we allow them to non-trivially depend on an arbitrary number of input gates (since the circuit gates have unbounded fan-in). Nevertheless, utilizing an “input-reduction” trick of [\[CGG24\]](#) which applies to both the foregoing families of sources, it can be shown that for our disperser purposes we may assume the input of both families of sources to be of length $O(n)$. This allows us to apply [Theorem 1.10](#) to obtain low-degree dispersers for both of these families.

Theorem 1.12 (Informal, see [Theorems 6.5](#) and [6.6](#)). *For every $1 \leq \ell < d \leq n$, there exists a degree- d disperser*

- for the family of degree- ℓ sources on $\{0, 1\}^n$ with min-entropy $k \geq (12^\ell \cdot d^d \cdot n)^{\frac{1}{d-\ell}} + 1$.
- for the family of n^ℓ -size $\text{AC}[\oplus]$ circuit sources on $\{0, 1\}^n$ with min-entropy $k \geq (30^2 \cdot d^d \cdot n^{2\ell})^{\frac{1}{d-2}} + 1$.

In particular, for every $\ell \in \mathbb{N}$, there is a degree- $(\ell + 2)$ disperser for degree- ℓ sources on $\{0, 1\}^n$ with min-entropy $\Omega(\sqrt{n})$.

We note that both of these source families are very powerful, and to the best of our knowledge, no nontrivial low-complexity dispersers for either of these families of sources was known prior to this work (except in the easier case of degree-1 sources which corresponds to affine sources for which explicit extractors for logarithmic entropy was recently proved [Li23]). Let us also point out that the two foregoing classes have incomparable power, and that it is straightforward to use our proof technique to conclude the same result for a class of sources that generalizes both $\text{AC}[\oplus]$ and constant-degree polynomials. Indeed, the input-reduction and counting idea works for the “hybrid” class of polynomial-size circuits which extends $\text{AC}[\oplus]$ by allowing additional unbounded fan-in gates computing arbitrary polynomials of fixed constant degree. However, for ease of exposition, we have chosen to present only the results for $\text{AC}[\oplus]$ and low-degree sources separately.

1.2.2 Low-Degree Extractors

Next, we move on to another application concerning the existence of low-degree extractors for samplable sources. Can we prove the existence of low-degree extractors for all the families for which we proved the existence of low-degree dispersers? We prove this by showing an analogue of [Theorem 1.10](#) for extractors.

Theorem 1.13 (Informal, see [Theorem 8.1](#) for the more general statement). *Let \mathcal{X} be a family of distributions of min-entropy $k \geq 5 \log n$ over $\{0, 1\}^n$ for large enough n . Then for every $d \geq 6$, a uniformly random degree- d polynomial is an ε -extractor for \mathcal{X} with probability at least*

$$1 - |\mathcal{X}| \cdot e^{3n - O(k^{d/2})/n^2}$$

for $\varepsilon = (2d)^d \cdot k^{-d/4}$.

A similar statement (see [Theorem 8.1](#)) holds for families of sources that are close to convex combinations of another small family of sources. Combined with the input-reduction trick, we obtain as a corollary, the existence of low-degree extractors for various families of sources, notably, lower-degree sources and $\text{AC}[\oplus]$ circuits.

Theorem 1.14 (Informal, see [Theorem 8.3](#)). *For all $\ell, d \geq 1$, and all large enough n , and $k \geq 5 \log n$. There exists a degree- d \mathbb{F}_2 -polynomial that is an ε -extractor for the following families of sources over $\{0, 1\}^n$ for $\varepsilon = (2d)^d \cdot k^{-d/4}$:*

- ℓ -local sources for $k \geq (2^\ell n^3 \log n)^{2/d}$.
- depth- ℓ decision forest sources for $k \geq (2^\ell n^3 (\log n + \ell))^{2/d}$.
- degree- ℓ sources for $k \geq (3^\ell n)^{\frac{6}{d-2\ell}}$.
- n^ℓ -size $\text{AC}[\oplus]$ circuit sources (with unbounded number of input gates) for $k \geq 3n^{\frac{4(\ell+1)}{d-4}}$.

In [Theorem 8.4](#), we further extend our low-degree extractors to multi-output extractors that output $\Theta(k)$ bits. This is done by independently picking random degree- d polynomials p_1, \dots, p_t for some $t = \Theta(k)$, and analyzing the probability that each p_i is an extractor for the family of sources obtained by \mathcal{X} conditioned on the values of p_1, \dots, p_{i-1} .

Let us now discuss our proof technique for [Theorem 1.13](#). Recall that [Theorem 1.10](#) was a corollary to [Corollary 1.4](#) which showed that a random polynomial is with high probability non-constant on the support of any fixed high min-entropy distribution. A priori it is not clear how to use this bound on the Hilbert function to prove [Theorem 1.13](#).

Indeed, let us consider the simpler case of a fixed k -flat source \mathbf{X} over $\{0, 1\}^n$, which is uniformly distributed over a set $S \subseteq \{0, 1\}^n$ with $|S| = 2^k$. Note that a map $p : \{0, 1\}^n \rightarrow \{0, 1\}$ is an extractor for \mathbf{X} if it has small bias on S . Thus, for example, to prove the special case of [Theorem 1.13](#) for small families of k -flat sources, we would need to prove that a random degree- d polynomial is small-biased on S with high probability. However, [Corollary 1.4](#) only tells us that $h_S(d, \mathbb{F}_2) \geq \binom{k}{\leq d}$, which is not enough to prove concentration bounds for the bias of a random degree- d polynomial on an arbitrary set S . We note that when S is highly structured, that is when it is an affine subspace, this problem is equivalent to questions about list-decoding size of Reed-Muller codes, and known results such as one by Kaufman, Lovett, and Porat [[KLP12](#)] that show that the number of distinct ε -biased degree- d polynomials on a k -dimensional subspace S is at most $(1/\varepsilon)^{k^{d-1}}$ could be utilized. However, for our application we have to deal with arbitrary sets S .

Uniform covering by sets of full Hilbert dimension. We say that a set $T \subseteq \{0, 1\}^n$ has “full Hilbert dimension” if $h_T(d, \mathbb{F}_2) = |T|$. Note that when T has full Hilbert dimension, then the restriction of a random degree- d polynomial to T is uniformly distributed over $\{0, 1\}^T$. In particular, if T is a sufficiently large set of full Hilbert dimension, then a random degree- d polynomial is small-biased on T with high probability. We use this observation to design our technique for bounding the probability that a random degree- d polynomial is small-biased on any fixed source \mathbf{X} of large min-entropy. For simplicity we describe the idea for flat sources. In this case, \mathbf{X} is uniformly distributed over a set S with $|S| \geq 2^k$. It is sufficient to prove the existence of an almost-uniform covering of S by large sets T_1, \dots, T_t of the same size with full Hilbert dimensions, where we call a covering almost-uniform if each element $x \in S$ belongs to roughly $tm/|S|$ many sets, where we assume $|T_i| = m$.

We obtain such a covering by analyzing the probability that a uniformly picked subset $T_i \subseteq S$ has full Hilbert dimension. Using our bound on the Hilbert function, [Corollary 1.4](#), which allows us to bound the size of the “degree- d closure” of small sets, we prove that a random set T_i of size m , for some $m = \binom{\Theta(k)}{\leq d}$, has full Hilbert dimension with high probability. Similarly, we prove using the Bayes rule, that we may pick these good sets T_i ’s of full Hilbert dimension in a way that leads to an almost uniform covering. Since T_i ’s are of sufficiently large size $\binom{\Theta(k)}{\leq d}$ and of full Hilbert dimension, we can use the Hoeffding inequality to bound the probability that a random degree- d polynomial is biased on a T_i to be exponentially small in $\Theta(k)^d$, which is good enough for our applications to existence of low-degree extractors. We obtain the following result which can be used to prove [Theorem 1.13](#).

Theorem 1.15 (Informal, see [Theorem 7.3](#)). *Let $n, d, k \geq 1$, and $\varepsilon > 0$ be a real. Then for every distribution \mathbf{X} over $\{0, 1\}^n$ with $H_\infty(\mathbf{X}) \geq k$, a uniformly random degree- d polynomial f is an ε -extractor for \mathbf{X} , with probability at least $1 - e^{3n - \varepsilon^2 \binom{\ell}{\leq d} / (Cn^2)}$ where $\ell = k/2 - \log(32n/\varepsilon)$ and $C = 7 \cdot (32)^2$.*

We find our technique of obtaining almost uniform coverings with sets of full Hilbert dimension to be powerful, and hope that it will find other applications beyond the ones explored here.

1.3 Remarks

Correlation bounds over arbitrary subsets. We note that our proof of [Theorem 1.15](#) ([Theorem 8.1](#)) can be modified to the following correlation bounds with any fixed function.

Theorem 1.16. *Let $n, d, k \geq 1$, $\varepsilon > 0$ be a real, and $g : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$ be a fixed function. Then for every distribution \mathbf{X} over $\{0, 1\}^n$ with $H_\infty(\mathbf{X}) \geq k$, for a uniformly random degree- d polynomial f we have*

$$\Pr[f(\mathbf{X}) = g(\mathbf{X})] = \frac{1}{2} \pm \varepsilon,$$

with probability at least $1 - e^{-3n - \varepsilon^2 \binom{\ell}{\leq d} / (Cn^2)}$ where $\ell = k/2 - \log(32n/\varepsilon)$ and $C = 7 \cdot (32)^2$.

This generalization is quite straightforward, as once we obtain a uniform covering by sets of maximum Hilbert dimension, then Hoeffding bounds can be used to bound the correlation of a random polynomial with the fixed function restricted to the sets belonging to the cover. This can then be used to bound the over-all correlation with the fixed function in a similar way to the proof of [Theorem 8.1](#).

Punctured Reed-Muller codes. The special case of [Theorem 1.16](#) when \mathbf{X} is a flat source can be interpreted as a bound on the list-decoding size of Reed-Muller codes when “punctured” on a large set $S \subseteq \mathbb{F}_2^n$. Recall that the binary Reed-Muller code $\text{RM}[d, n]$ consists of codewords in \mathbb{F}_2^n that correspond to the evaluation vectors of degree $\leq d$ polynomials over \mathbb{F}_2 . Given a set $S \subseteq \mathbb{F}_2^n$, the resulting punctured code consists of the evaluation of degree $\leq d$ polynomials on S . In this context, [Theorem 1.16](#) can be used to bound the list-size of any puncturing of the Reed-Muller code, showing that for any word w from \mathbb{F}_2^S , only a small fraction of codewords are within radius $\frac{1}{2} - \varepsilon$ of w . Another interpretation of [Theorem 1.16](#) is that any puncturing of the Reed-Muller codes over a set S can be turned into a “small-biased” code without much loss in the rate of the code.

Sampling lower bound for polynomial sources. Our low-degree extractor for lower-degree sources ([Theorem 1.14](#)) has a direct application in distributions that are hard to sample by low-degree polynomials. Indeed, an argument similar to the proof of [[Vio16](#), Lemma 3], [Theorem 1.14](#) implies the existence of a degree- $O(d)$ polynomial p for which the distribution $(\mathbf{U}, p(\mathbf{U}))$ cannot be sampled by any degree- d source, where $\mathbf{U} \sim \mathbf{U}_n$.

Suppose that p is a degree- $O(d)$ polynomial that is an ε -extractor for the family of degree $\leq 2d$ sources over $\{0, 1\}^n$ of min-entropy $\geq \frac{n}{2}$, where $\varepsilon = o(1)$. The existence of such a polynomial p is guaranteed by [Theorem 1.14](#). Now suppose that $(G(\mathbf{U}'), g(\mathbf{U}'))$, where $\mathbf{U}' \sim \mathbf{U}_m$ for some $m \geq 1$, is a degree $\leq d$ source sampling $(\mathbf{U}, p(\mathbf{U}))$. In particular, G is an n -bit degree $\leq d$ source and g is a degree $\leq d$ polynomial. Consider the n -bit random variable $\mathbf{R} = G(\mathbf{U}') \cdot g(\mathbf{U}') + \mathbf{U}_n \cdot (1 - g(\mathbf{U}'))$. Since \mathbf{R} is sampled by a degree $\leq 2d$ source of min-entropy $n - O(1)$, $\Pr[p(\mathbf{R}) = 1] = \frac{1}{2} + o(1)$. On the other hand, by the definition \mathbf{R} , we have $\Pr[p(\mathbf{R}) = 1] \geq \frac{1}{2} + \Omega(1)$, which is a contradiction.

Related Work. An independent and concurrent paper by Alrabiah, Goodman, Mosheiff, and Ribeiro [AGMR24] proves the existence of low-degree extractors for similar families of sources that are considered in our work, as well as sumset sources. While the proofs are quite different, they both rely on bounds on the dimension of punctured Reed-Muller codes (equivalently the Hilbert function).

Acknowledgments. We thank Omar Alrabiah, Jesse Goodman, Jonathan Mosheiff, and João Ribeiro for sharing with us an early draft of their work. We would also like to thank Jesse Goodman and S. Venkitesh for helpful discussions and pointers. We are very grateful to the anonymous reviewers for their comments and pointers to related work. Part of this work was conducted while the second author was visiting the Simons Institute for the Theory of Computing at UC Berkeley; he extends his thanks to the institute for its support and hospitality.

2 Preliminaries

All logarithms in this paper are base 2. By \mathbb{N} we denote the set of non-negative integers. For a positive integer n , by $[n]$ we denote the set $\{1, \dots, n\}$. For a prime power q , denote by \mathbb{F}_q the finite field q elements.

For simplicity, throughout this paper, we refer to a polynomial as a degree- d polynomial if its total degree is *at most* d . When q is a prime power, by $\mathcal{P}_q(n, d)$ we denote the set of all degree- d polynomials from $\mathbb{F}[X_1, \dots, X_n]$ with individual degrees at most $q - 1$. Note that each element of $\mathcal{P}_q(n, d)$ corresponds to a unique map $\mathbb{F}_q^n \rightarrow \mathbb{F}_q$.

Let $r_1, \dots, r_n \geq 1$ be integers and $F = \prod_{i=1}^n \{0, \dots, r_i - 1\}$. For $x \in F$ and $i \in [n]$, x_i denotes the i th coordinate of x . For $x \in F$, we define its *generalized Hamming weight* as $|x| := \sum_i x_i$, where the summation is over the integers. For an integer $d \geq 0$, and a set $T \subseteq F$, we denote the set of its elements of generalized Hamming weight $\leq d$ by

$$T_{\leq d} := \{x \in T : |x| \leq d\}.$$

For $a, b \in F$, we write $a \leq_P b$ if $a_i \leq b_i$ for all $i \in [n]$. We say a subset $T \subseteq F$ is *down-closed* if for all $a, b \in F$ such that $a \leq_P b$, if b is in T , then so is a . Similarly, we say a subset $T \subseteq F$ is *up-closed* if for all $a, b \in F$ such that $a \leq_P b$, if a is in T , then so is b .

The *lexicographic order* \prec on F is defined as follows. For distinct $x, y \in F$, x precedes y , denoted $x \prec y$, in lexicographic order if $x_i < y_i$, where i is the smallest index such that $x_i \neq y_i$.

We will be studying the following quantity.

Definition 2.1. For $F = \prod_{i=1}^n \{0, \dots, r_i - 1\}$ and $k \leq |F|$, let

$$\mathcal{H}_F(d, k) := \min_T |T_{\leq d}|,$$

where the minimum is taken over all down-closed sets $T \subseteq F$ with $|T| = k$. Moreover, denote $\mathcal{H}_F(d, k)$ by $\mathcal{H}_q^n(d, k)$ in the special case where $r_1 = \dots = r_n = q$ for some $q \geq 1$.

2.1 Probability Distributions

We use lowercase letters such as x, y to denote vectors, uppercase bold letters such as \mathbf{X}, \mathbf{Y} to denote random variables, and \mathcal{X}, \mathcal{Y} to denote families of distributions. By \mathbf{U}_n we denote the uniform distribution over $\{0, 1\}^n$.

The statistical distance between two distributions \mathbf{A} and \mathbf{B} over a finite domain X is

$$\Delta(\mathbf{A}, \mathbf{B}) = \frac{1}{2} \left(\sum_{x \in X} |\Pr[x \in \mathbf{A}] - \Pr[x \in \mathbf{B}]| \right).$$

We say two distributions \mathbf{A} and \mathbf{B} are ε -close if $\Delta(\mathbf{A}, \mathbf{B}) \leq \varepsilon$. For a distribution $\mathbf{X} \sim \{0, 1\}^n$, the min-entropy of \mathbf{X} is

$$H_\infty(\mathbf{X}) = \min_{x \in \text{support}(\mathbf{X})} -\log(\Pr[\mathbf{X} = x]).$$

We will use following forms of Chernoff's and Hoeffding's bounds (see, e.g., [MU17, Hoe63]).

Theorem 2.2 (Chernoff bound). *Let $X_1, \dots, X_n \in \{0, 1\}$ be independent random variables. Let $X = \sum_{i=1}^n X_i$ and $\mu = \mathbb{E}(X)$. Then we have*

$$\Pr[|X - \mu| \geq \delta\mu] \leq 2e^{-\mu\delta^2/3}$$

for all $0 < \delta < 1$.

Theorem 2.3 (Hoeffding's inequality). *Let $X_1, \dots, X_n \in [0, 1]$ be independent random variables, $X = \sum_{i=1}^n X_i$ and $\mu = \mathbb{E}[X]$. Then,*

$$\Pr[|X - \mu| \geq R] \leq 2e^{-\frac{2R^2}{n}}.$$

2.2 Randomness Sources, Dispersers, and Extractors

Definition 2.4 (Sources and Their Convex Combinations). A distribution $\mathbf{X} \sim \{0, 1\}^n$ is a *source* from a class \mathcal{C} of functions, if $\mathbf{X} = f(\mathbf{U}_m)$ for some $f : \{0, 1\}^m \rightarrow \{0, 1\}^n \in \mathcal{C}$. A distribution \mathbf{Y} is a *convex combination* of sources \mathbf{X}_i if $\mathbf{Y} = \sum_i p_i \mathbf{X}_i$ for some non-negative p_i satisfying $\sum_i p_i = 1$, i.e., \mathbf{Y} samples from each \mathbf{X}_i with probability p_i .

One of the most powerful classes of sources that we consider in this work is the class of circuits of polynomial size.

Definition 2.5 (AC[\oplus] circuits). An AC[\oplus] *circuit* is an unbounded-depth Boolean circuit consisting of AND, OR, XOR, NOT gates of unbounded fan-in. The size of such a circuit is the number of non-input gates in it.

We focus on the class of AC[\oplus] circuit as it generalizes circuit classes previously studied in this context: unbounded-depth circuits of bounded fan-in from P/poly, and bounded-depth circuits of unbounded fan-in from, say, AC⁰. We remark that we define AC[\oplus] sources (see [Definition 2.6](#)) as sources where each output is computed by an AC[\oplus] circuit of polynomial size but with an arbitrary (possibly super-polynomial) number of inputs. This explains why in this context P/poly and AC⁰ circuits are incomparable, and why we work with AC[\oplus] circuits generalizing both of the aforementioned classes. In fact, our results hold even for a larger class of circuits where not only XOR but arbitrary constant-degree polynomials over \mathbb{F}_2 can be computed at gates (see the discussion at the end of [Section 6](#)).

Definition 2.6 (Structured Sources). Let $n, d, m \in \mathbb{N}$, $f : \{0, 1\}^m \rightarrow \{0, 1\}^n$, and \mathbf{X} be a distribution over $\{0, 1\}^n$ that is generated as $f(\mathbf{U}_m)$.

- \mathbf{X} is called a *d-local source* if every output bit of f depends only on at most d of its input bits.
- \mathbf{X} is called a *depth-d decision forest source* if every output bit of f is determined by a depth- d decision tree of its input variables.
- \mathbf{X} is called a *degree-d source* if every output bit of f is a degree- d polynomial over \mathbb{F}_2 .
- \mathbf{X} is called a *size- n^d circuit source* if there is an $\text{AC}[\oplus]$ circuit of size n^d that computes all output bits of f .

Note that every d -local source is a depth- d decision forest source, and a degree- d source. Also, every depth- d decision forest source is a degree- d source and a 2^d -local source.

We will use the following bounds on the numbers of d -local sources and depth- d decision forest sources.

Proposition 2.7. *Let $n, d \geq 1$.*

- *The number of d -local sources over $\{0, 1\}^n$ is bounded from above by $2^{2^d n + 2dn \log n}$.*
- *The number of depth- d decision forest sources is bounded from above by $2^{(d+\log n)2^{d+1}n}$.*

Proof. Every d -local source over $\{0, 1\}^n$ can be expressed as $f(\mathbf{U}_m)$ where $f : \{0, 1\}^m \rightarrow \{0, 1\}^n$ is a d -local function, and we may assume without loss of generality that $m \leq dn$, as the number of relevant variables is bounded from above by dn . Then the number of distinct d -local sources is at most

$$\left(\binom{dn}{d} \cdot 2^{2^d} \right)^n \leq 2^{2^d n + 2dn \log n} .$$

Similarly, for depth- d decision forest sources we may assume that $m \leq 2^d n$, as each depth- d decision tree depends on at most 2^d variables. The number of depth- d decision trees on m variables is at most $(2m)^{2^d}$. Thus, the number of distinct depth- d decision forest sources on $\{0, 1\}^n$ is at most

$$2^{(d+\log n)2^{d+1}n} . \quad \square$$

For polynomial and circuit sources where the number of input bits cannot be bounded by a small function of n (unlike the sources considered in [Proposition 2.7](#)), we will need the following bounds on the number of such sources for a *fixed* number of input bits m .

Proposition 2.8. *Let $n, d, m \geq 1$.*

- *The number of degree- d polynomials $f : \mathbb{F}_2^m \rightarrow \mathbb{F}_2^n$ is bounded from above by $2^{n \binom{m}{\leq d}}$.*
- *The number of $\text{AC}[\oplus]$ circuits $C : \{0, 1\}^m \rightarrow \{0, 1\}^n$ of size n^d is bounded from above by $2^{4n^d(n^d+m)}$.*

Proof. The bound on the number of degree- d polynomials follows from the observation that the number of multilinear monomials in m variables of degree at most d is $\binom{m}{\leq d}$.

For the bound on the number of n^d -size circuits, first notice that each gate of C can be described by $m + n^d$ bits specifying the set of gates feeding it, and 3 bits specifying the function computed at the gate. Additional $n \log(n^d + m)$ bits are sufficient to specify the n output bits of the circuit. Therefore, the total number of such circuits is bounded from above by

$$2^{n^d(m+n^d+3)+n \log(n^d+m)} \leq 2^{4n^d(n^d+m)} . \quad \square$$

Definition 2.9 (Disperser). A function $\text{Disp} : \{0, 1\}^n \rightarrow \{0, 1\}$ is a disperser for a family \mathcal{X} of sources over $\{0, 1\}^n$ with min-entropy k , if for every source $\mathbf{X} \in \mathcal{X}$ with $H_\infty(\mathbf{X}) \geq k$, the support of $\text{Disp}(\mathbf{X})$ is $\{0, 1\}$.

Definition 2.10 (Extractor). A function $\text{Ext} : \{0, 1\}^n \rightarrow \{0, 1\}^m$ is an ε -extractor for a family \mathcal{X} of sources over $\{0, 1\}^n$ with min-entropy k , if for every source $\mathbf{X} \in \mathcal{X}$ with $H_\infty(\mathbf{X}) \geq k$, $\Delta(\text{Ext}(\mathbf{X}(\mathbf{U}_t)), \mathbf{U}_m) \leq \varepsilon$.

For clarity of presentation, in this paper when working with sources that are guaranteed to have entropy $H_\infty(\mathbf{X}) \geq k$, we will always assume that k is an integer.

2.3 Hilbert Functions and Standard Monomials

In this section, we recall some necessary definitions (see, e.g., [CLO13]). Let \mathbb{F} be a field, X_1, \dots, X_n be indeterminates, and $\mathbb{F}[X_1, \dots, X_n]$ be the polynomial ring in n indeterminates over \mathbb{F} . For a polynomial $f \in \mathbb{F}[X_1, \dots, X_n]$ and $S \subseteq \mathbb{F}^n$, let $f|_S \in \mathbb{F}^S$ be the restriction of f to S . For $d \in \mathbb{N}$, by $\Gamma_S(d) \subseteq \mathbb{F}^S$ we denote the vector space spanned by $f|_S$ for all degree- d polynomials f :

$$\Gamma_S(d) := \{f|_S : f \in \mathbb{F}[X_1, \dots, X_n], \deg(f) \leq d\}.$$

Definition 2.11 (Hilbert function). For a set $S \subseteq \mathbb{F}^n$, the (affine) Hilbert function of S over \mathbb{F} , $h_S(\cdot, \mathbb{F}) : \mathbb{N} \rightarrow \mathbb{N}$, is defined as the dimension of $\Gamma_S(d)$ over \mathbb{F} , i.e.,

$$h_S(d, \mathbb{F}) := \dim_{\mathbb{F}}(\Gamma_S(d)).$$

Definition 2.12 (Monomial order). Let \preceq be a total order on the monomials in a polynomial ring $\mathbb{F}[X_1, \dots, X_n]$. The order \preceq is called a *monomial order* if 1 is the minimal element of \preceq , and for all monomials m_1, m_2, m satisfying $m_1 \preceq m_2$, we have that $m_1 m \preceq m_2 m$. The order \preceq is *degree-compatible* if for all monomials m_1, m_2 such that $\deg(m_1) < \deg(m_2)$, we have that $m_1 \preceq m_2$.

Examples of degree-compatible monomial orders include the graded lexicographic and graded reverse lexicographic orders.

Definition 2.13 (Graded orders). The *graded lexicographic order* \leq_{grlex} and the *graded reverse lexicographic order* \leq_{grevlex} are defined as follows. For a pair of monomials $m_1 = X_1^{\alpha_1} \cdots X_n^{\alpha_n}$ and $m_2 = X_1^{\beta_1} \cdots X_n^{\beta_n}$, let $\alpha = \sum_{i=1}^n \alpha_i$, $\beta = \sum_{i=1}^n \beta_i$, and $\gamma = (\beta_1 - \alpha_1, \dots, \beta_n - \alpha_n)$. We have that $m_1 \leq_{\text{grlex}} m_2$ if and only if either $\alpha < \beta$, or $\alpha = \beta$ and the leftmost non-zero entry of γ is positive. Similarly, $m_1 \leq_{\text{grevlex}} m_2$ if and only if either $\alpha < \beta$, or $\alpha = \beta$ and the rightmost non-zero entry of γ is negative.

Definition 2.14 (Leading monomial). For a nonzero polynomial $f \in \mathbb{F}[X_1, \dots, X_n]$, the *leading monomial* of f under a monomial order \preceq is the largest monomial of f under \preceq .

Let R be a commutative ring (such as the polynomial ring $\mathbb{F}[X_1, \dots, X_n]$). An *ideal* of R is a subset I of R such that for all $a, b \in I$ and $r \in R$, we have that $a + b \in I$ and $ra \in I$.

Definition 2.15 (Standard monomial). Let I be an ideal of $\mathbb{F}[X_1, \dots, X_n]$, and \preceq be a monomial order. A *standard monomial* m of I is a monomial in X_1, \dots, X_n that is not the leading monomial of any nonzero polynomial in I .

For an ideal I and $d \in \mathbb{N}$, $\text{SM}(I)$ denotes the set of all standard monomials of I , and $\text{SM}_{\leq d}(I)$ denotes the set of all standard monomials of I of degree at most d :

$$\text{SM}_{\leq d}(I) = \{m \in \text{SM}(I) : \deg(m) \leq d\}.$$

For a set $S \subseteq \mathbb{F}^n$, by $I(S)$ we denote the ideal of polynomials in $\mathbb{F}[X_1, \dots, X_n]$ vanishing on S ,

$$I(S) = \{f \in \mathbb{F}[X_1, \dots, X_n] : f|_S = 0^S\}.$$

For an ideal I of $\mathbb{F}[X_1, \dots, X_n]$, define the set $V(I) \subseteq \mathbb{F}^n$ by

$$V(I) = \{a \in \mathbb{F}^n : f(a) = 0 \text{ for all } f \in I\}.$$

By definition, for all $f \in I$ and $a \in V(I)$, we have $f(a) = 0$. So $I \subseteq I(V(I))$.

Finally, for a set $S \subseteq \mathbb{F}^n$, define

$$\text{SM}(S) = \text{SM}(I(S)) \quad \text{and} \quad \text{SM}_{\leq d}(S) = \text{SM}_{\leq d}(I(S)).$$

We say that a set T of monomials is *down-closed* if for all monomials m and m' such that $m \in T$ and m' divides m , it holds that $m' \in T$. It is easy to see that $\text{SM}(S)$ is down-closed. Indeed, if m' was the leading monomial of a polynomial $p \in I(S)$, then m would be the leading monomial of the polynomial $p \cdot (m/m') \in I(S)$.

We will use the following facts about $\text{SM}(S)$ and $\text{SM}_{\leq d}(S)$, which are proven, for example, in [Rem16, Lemma 1] and [Fel07, Corollary 2.1.21].

Lemma 2.16. *Let $S \subseteq \mathbb{F}^n$ be a finite set. Then*

(a) *for every monomial order \preceq ,*

$$|S| = |\text{SM}(S)|;$$

(b) *for every degree-compatible monomial order \preceq and every $d \in \mathbb{N}$,*

$$h_S(d, \mathbb{F}) = |\text{SM}_{\leq d}(S)|.$$

3 Hilbert Functions of Sets in Finite Grids

Let \mathbb{F} be a field. We consider Hilbert functions of subsets of a finite grid $A = \prod_{i=1}^n A_i$, where each A_i is a finite subset of the field \mathbb{F} . The main result of this section is that the minimum value $h_S(d, \mathbb{F})$ of a set $S \subseteq A$ of size k equals the quantity $\mathcal{H}_F(d, k)$ introduced in Definition 2.1, where $F = \prod_{i=1}^n \{0, 1, \dots, |A_i| - 1\}$.

Consider the following setting: Let r_1, \dots, r_n be integers such that $1 \leq r_i \leq |\mathbb{F}|$ for $i \in [n]$. For each $i \in [n]$, let A_i be a subset of \mathbb{F} consisting of r_i distinct elements $a_{i,1}, \dots, a_{i,r_i} \in \mathbb{F}$. Let A be the Cartesian product $\prod_{i=1}^n A_i$. Let \mathcal{M} be the set of monomials dividing $\prod_{i=1}^n X_i^{r_i-1}$. Let σ_A be the bijection from \mathcal{M} to A defined by

$$\sigma_A : \prod_{i=1}^n X_i^{e_i} \mapsto (a_{1,e_1+1}, \dots, a_{n,e_n+1}). \quad (2)$$

Finally, fix a degree-compatible monomial order \preceq .

The next lemma states that every down-closed subset $T \subseteq \mathcal{M}$ can be realized as the set of standard monomials of the set $\sigma_A(T) \subseteq A$.

Lemma 3.1. *Let T be a down-closed subset of \mathcal{M} . Then $\text{SM}(\sigma_A(T)) = T$.*

Proof. Let I be the ideal of $\mathbb{F}[X_1, \dots, X_n]$ generated by the set of polynomials

$$\left\{ \prod_{i=1}^n \prod_{j=1}^{\deg_{X_i}(m)} (X_i - a_{i,j}) : m \in \mathcal{M} \setminus T \right\} \cup \left\{ \prod_{j=1}^{r_1} (X_1 - a_{1,j}), \dots, \prod_{j=1}^{r_n} (X_n - a_{n,j}) \right\}.$$

Let $S = V(I)$, so that $I \subseteq I(S)$. As $\prod_{j=1}^{r_1} (X_1 - a_{1,j}), \dots, \prod_{j=1}^{r_n} (X_n - a_{n,j}) \in I$, we have

$$S \subseteq V \left(\prod_{j=1}^{r_1} (X_1 - a_{1,j}), \dots, \prod_{j=1}^{r_n} (X_n - a_{n,j}) \right) = A.$$

Next, we show that $\text{SM}(S) \subseteq T$. Consider $m \in \text{SM}(S)$. Suppose $\deg_{X_i}(m) \geq r_i$ for some $i \in [n]$. Then m is the leading monomial of $(m/X_i^{r_i}) \cdot \prod_{j=1}^{r_i} (X_i - a_{i,j})$, and the latter is in $I(A) \subseteq I(S)$. This contradicts the assumption that $m \in \text{SM}(S) = \text{SM}(I(S))$. Therefore, in the following we assume that for all $i \in [n]$, $\deg_{X_i}(m) < r_i$, or equivalently, $m \in \mathcal{M}$. Now suppose $m \notin T$. Then by the definition of I , the polynomial $\prod_{i=1}^n \prod_{j=1}^{\deg_{X_i}(m)} (X_i - a_{i,j})$ is in $I \subseteq I(S)$, and its leading monomial is m . This again contradicts the assumption that $m \in \text{SM}(S) = \text{SM}(I(S))$. This shows that every $m \in \text{SM}(S)$ also belongs to T . Therefore, $\text{SM}(S) \subseteq T$.

Consider arbitrary $m_0 = \prod_{i=1}^n X_i^{e_i} \in T$ and let $a = \sigma_A(m_0) = (a_{1,e_1+1}, \dots, a_{n,e_n+1})$. Consider arbitrary $m \in \mathcal{M} \setminus T$. As T is down-closed, there exists $k \in [n]$ such that $\deg_{X_k}(m) > \deg_{X_k}(m_0) = e_k$. The polynomial $\prod_{i=1}^n \prod_{j=1}^{\deg_{X_i}(m)} (X_i - a_{i,j})$ then contains the factor $X_k - a_{k,e_k+1}$, so it vanishes at a . As $m \in \mathcal{M} \setminus T$ is arbitrary and $\prod_{j=1}^{r_1} (X_1 - a_{1,j}), \dots, \prod_{j=1}^{r_n} (X_n - a_{n,j})$ also vanish at a , this shows $a \in V(I) = S$ by the definition of I . As $m_0 \in T$ is arbitrary, we see that $\sigma_A(T) \subseteq S$. It follows that

$$|T| = |\sigma_A(T)| \leq |S| = |\text{SM}(S)|$$

where the first equality follows from the injectivity of σ_A and the last equality holds by [Lemma 2.16](#) (a). As $\text{SM}(S) \subseteq T$, we must have $\text{SM}(S) = T$ and $|\sigma_A(T)| = |S|$. The latter equality together with $\sigma_A(T) \subseteq S$ yields $\sigma_A(T) = S$. So $\text{SM}(\sigma_A(T)) = \text{SM}(S) = T$. \square

Lemma 3.2. *Let $k, d \in \mathbb{N}$ such that $k \leq |A|$. Then*

$$\min_{S \subseteq A: |S|=k} h_S(d, \mathbb{F}) = \min_{\text{down-closed } T \subseteq \mathcal{M}: |T|=k} |\{m \in T : \deg(m) \leq d\}|.$$

Proof. Let $S \subseteq A$ be a set of size k such that $h_S(d, \mathbb{F})$ achieves the minimum. Let $T = \text{SM}(S)$, which is down-closed. Moreover, as $S \subseteq A$, we have $T = \text{SM}(S) \subseteq \text{SM}(A) = \mathcal{M}$. Note that $|T| = |\text{SM}(S)| = |S| = k$ by [Lemma 2.16](#) (a). And $|\{m \in T : \deg(m) \leq d\}| = |\text{SM}_{\leq d}(S)| = h_S(d, \mathbb{F})$ by [Lemma 2.16](#) (b). So LHS \geq RHS.

Conversely, let T be a down-closed subset of \mathcal{M} of size k such that $|\{m \in T : \deg(m) \leq d\}|$ achieves the minimum. Let $S = \sigma_A(T)$. Then $\text{SM}(S) = T$ by [Lemma 3.1](#). We have $|S| = |\text{SM}(S)| = |T| = k$ by [Lemma 2.16](#) (a) and $h_S(d, \mathbb{F}) = |\text{SM}_{\leq d}(S)| = |\{m \in T : \deg(m) \leq d\}|$ by [Lemma 2.16](#) (b). So LHS \leq RHS. This finishes the proof of the lemma. \square

Let $F = \prod_{i=1}^n \{0, 1, \dots, r_i - 1\}$. Let $\phi : \mathcal{M} \rightarrow F$ be the bijection

$$\phi : \prod_{i=1}^n X_i^{e_i} \mapsto (e_1, \dots, e_n). \quad (3)$$

Lemma 3.2 can now be reformulated as follows.

Corollary 3.3. *Let $k, d \in \mathbb{N}$ such that $k \leq |A|$. Then*

$$\min_{S \subseteq A: |S|=k} h_S(d, \mathbb{F}) = \mathcal{H}_F(d, k). \quad (4)$$

Proof. Recall that by **Definition 2.1**,

$$\mathcal{H}_F(d, k) = \min_{\text{down-closed } S \subseteq F: |S|=k} |S_{\leq d}|.$$

Identify \mathcal{M} with F via the bijection ϕ . The claim follows from **Lemma 3.2** by noting that $T \subseteq \mathcal{M}$ is down-closed iff $\phi(T) \subseteq F$ is down-closed, and that for $m \in \mathcal{M}$, $\deg(m) \leq d$ iff $\phi(m) \in F_{\leq d}$. \square

For the special case of a finite field $\mathbb{F} = \mathbb{F}_q$, $r_1 = \dots = r_n = q$, and $A_1 = \dots = A_n = \mathbb{F}_q$, we have $A = \mathbb{F}_q^n$, and the right-hand side of **Equation (4)** becomes $\mathcal{H}_q^n(d, k)$ from **Definition 2.1**. This leads us to the following corollary.

Corollary 3.4. *For every $n, k, d \in \mathbb{N}$ where $k \leq q^n$, a prime power q , and every set $S \subseteq \mathbb{F}_q^n$ of size $|S| = k$, we have that*

$$h_S(d, \mathbb{F}_q) \geq \mathcal{H}_q^n(d, k).$$

Finally, we state the following lemma, which will be used in **Section 5**. Its proof reuses ideas from the previous proofs in this section.

Lemma 3.5. *Let $n, d \in \mathbb{N}$. Let $\sigma_A : \mathcal{M} \rightarrow A$ and $\phi : \mathcal{M} \rightarrow F$ be the bijections (2) and (3) respectively. Let $S \subseteq A$ such that $T := \sigma_A^{-1}(S) \subseteq \mathcal{M}$ is down-closed. Let $T' = \phi(T) \subseteq F$. Then $h_S(d, \mathbb{F}) = |T'_{\leq d}|$.*

Proof. By definition, we have $\sigma_A(T) = S$. So $\text{SM}(S) = \text{SM}(\sigma_A(T)) = T$ by **Lemma 3.1**. Then we have

$$h_S(d, \mathbb{F}) = |\text{SM}_{\leq d}(S)| = |\{m \in \text{SM}(S) : \deg(m) \leq d\}| = |\{m \in T : \deg(m) \leq d\}| = |T'_{\leq d}|.$$

where the first inequality holds by **Lemma 2.16** (b), the third one holds by the fact that $\text{SM}(S) = T$, and the last one holds since ϕ maps $\{m \in T : \deg(m) \leq d\}$ bijectively to $T'_{\leq d}$. \square

4 Number of Points with Low Hamming Weight in Down-Closed Sets

In this section, we will find the exact values of all $\mathcal{H}_q^n(d, k)$ which, by **Corollary 3.4**, will give us tight lower bounds on the Hilbert function of sets of size k .

For every n, k, q where $k \leq q^n$, we define $M_q^n(k)$ as the set of the first k elements of $\{0, \dots, q-1\}^n$ in lexicographic order.

The main result of this section is the following theorem.

Theorem 4.1. For every $n, k, d, q \in \mathbb{N}$ where $k \leq q^n$,

$$\mathcal{H}_q^n(d, k) = |\mathbb{M}_q^n(k)_{\leq d}|.$$

Combining [Corollary 3.4](#) and [Theorem 4.1](#), we obtain the following bounds on the Hilbert function.

Corollary 4.2. For every prime power q , and $n, k, d \in \mathbb{N}$ where $k \leq q^n$, we have

$$\min_{S \subseteq \mathbb{F}_q^n, |S|=k} h_S(d, \mathbb{F}_q) = |\mathbb{M}_q^n(k)_{\leq d}|.$$

In particular, setting $q = 2$, for every $n, k, d \in \mathbb{N}$ where $k \leq 2^n$, and every $S \subseteq \mathbb{F}_2^n$ of size $|S| = k$,

$$h_S(d, \mathbb{F}_2) \geq \binom{\lfloor \log(k) \rfloor}{\leq d}.$$

We will use the following notation: For $t \in \{0, 1, \dots, n\}$, define $\mathcal{D}_q^n(t)$ to be the set of $x \in \{0, \dots, q-1\}^n$ whose first $n-t$ coordinates are zero.

Note that for every q, k , and n ,

$$\mathcal{D}_q^n(\lfloor \log_q k \rfloor) \subseteq \mathbb{M}_q^n(k) \subseteq \mathcal{D}_q^n(\lceil \log_q k \rceil).$$

When n and q are clear from the context, we omit the superscript n and the subscript q from $\mathbb{M}_q^n(k)$, $\mathcal{D}_q^n(t)$, and $\mathcal{H}_q^n(d, k)$.

4.1 The Boolean Case, $q = 2$

For a set $S \subseteq \{0, 1\}^n$, let $\min(S)$ and $\max(S)$ be respectively the smallest and the largest strings in S in lexicographic order. We say a set $S \subseteq \{0, 1\}^n$ is a *contiguous k -set* if $|S| = k$ and S consists of all x such that $\min(S) \preceq x \preceq \max(S)$.

We first show that $\mathbb{M}(k)$ has the largest number of low Hamming weight strings among all contiguous k -sets.

Lemma 4.3. Let $n, k, d \in \mathbb{N}$ be integers such that $k \leq 2^n$. Let $S^k \subseteq \{0, 1\}^n$ be a contiguous k -set. Then $|\mathbb{M}(k)_{\leq d}| \geq |S^k_{\leq d}|$.

Proof. We prove this lemma by induction on k and $\max(S^k)$. The base cases of $k \leq 1$ hold trivially, as in this cases $|\mathbb{M}(k)_{\leq d}| = k$. For each k , the base case of minimal $\max(S^k)$ holds trivially as in this case $S^k = \mathbb{M}(k)$ and $|S^k_{\leq d}| = |\mathbb{M}(k)_{\leq d}|$.

Assume for the induction hypothesis that the above statement is true for smaller contiguous sets (i.e. $|\mathbb{M}(t)_{\leq d}| \geq |S^t_{\leq d}|$ for all contiguous sets S^t of size $t \leq k-1$) as well as for contiguous sets A^k of size k having $\max(A^k) \prec \max(S^k)$.

Let m be the smallest integer such that $S^k \subseteq \mathcal{D}(m)$. If $d \geq m$, then $|S^k_{\leq d}| = k$. Also, since S^k is a contiguous set such that $S^k \subseteq \mathcal{D}(m)$, we have $\mathbb{M}(k) \subseteq \mathcal{D}(m)$, implying $|\mathbb{M}(k)_{\leq d}| = |\mathbb{M}(k)| = k$. Thus for the remainder of the proof we assume $d < m$.

If $T = S^k \cap \mathbb{M}(k) \neq \emptyset$, then the claim follows by applying the induction hypothesis to the contiguous $(k - |T|)$ -set $S^k - T$. Thus, we may assume from now that $S^k \cap \mathbb{M}(k) = \emptyset$.

Case 1: Assume $x_{n-m+1} = 1$ for all $x \in S^k$. Define the contiguous k -set

$$B^k = \{x - e_{n-m+1} : x \in S^k\},$$

where e_i denotes the element in $\{0, 1\}^n$ whose i th coordinate is one, and all other coordinates are zeros. Since $\max(B^k) \prec \max(S^k)$, by the induction hypothesis, $|\mathbb{M}(k)_{\leq d}| \geq |B^k_{\leq d}|$. Also, $|B^k_{\leq d}| \geq |S^k_{\leq d}|$ by definition, thereby implying $|\mathbb{M}(k)_{\leq d}| \geq |S^k_{\leq d}|$.

Case 2: Let $k^* > 0$ be the number of elements x in S^k with $x_{n-m+1} = 0$. Define

$$S^{k^*} = \{x \in S^k : x_{n-m+1} = 0\} \quad \text{and} \quad S^{k-k^*} = S^k \setminus S^{k^*}.$$

Note that in this case, we are guaranteed that S^{k-k^*} and S^{k^*} are both non-empty contiguous sets. Since $k - k^* < k$ and S^{k-k^*} is a contiguous $(k - k^*)$ -set, the induction hypothesis gives us

$$|\mathbb{M}(k - k^*)_{\leq d}| \geq |S^{k-k^*}_{\leq d}|.$$

It suffices to show that $|T_{\leq d}| \geq |S^{k^*}_{\leq d}|$ where $T = \mathbb{M}(k) \setminus \mathbb{M}(k - k^*)$. For $U \subseteq \{0, 1\}^n$ and $r \leq n$, define the set $\gamma(U, r) = \{0^{n-r}1^r - x : x \in U\}$, where $0^{n-r}1^r - x$ denotes coordinate-wise subtraction. Since $\max(S^{k^*}) = 0^{n-m+1}1^{m-1}$, S^{k^*} consists of the k^* lexicographically largest elements $\preceq 0^{n-m+1}1^{m-1}$. Consequently, $\gamma(S^{k^*}, m-1) = \mathbb{M}(k^*)$. Note that $\gamma(T, m-1)$ is a contiguous k^* -set, hence by the induction hypothesis, we get

$$|\gamma(S^{k^*}, m-1)_{\leq m-1-d}| \geq |\gamma(T, m-1)_{\leq m-1-d}|,$$

where we used the fact that $d \leq m-1$. This implies $|T_{\leq d}| \geq |S^{k^*}_{\leq d}|$, as $|S^{k^*}_{\leq d}| = k^* - |\gamma(S^{k^*}, m-1)_{\leq m-1-d}|$ and $|T_{\leq d}| = k^* - |\gamma(T, m-1)_{\leq m-1-d}|$, where we used the fact that $\max(T) \preceq 0^{n-m+1}1^{m-1}$ as $\mathbb{M}(k) \cap S^k = \emptyset$. \square

We now use [Lemma 4.3](#) to prove that if a contiguous k -set S^k that does *not* contain any of the first k strings in lexicographic order, then the result of [Lemma 4.3](#) $|S^k_{\leq d}| \leq |\mathbb{M}(k)_{\leq d}|$ can be strengthened to $|S^k_{\leq d}| \leq |\mathbb{M}(k)_{\leq d-1}|$.

Lemma 4.4. *Let $n, k, d \in \mathbb{N}$ be integers such that $k \leq 2^n$. Let $S^k \subseteq \{0, 1\}^n$ be a contiguous k -set. If $S^k \cap \mathbb{M}(k) = \emptyset$, then $|\mathbb{M}(k)_{\leq d-1}| \geq |S^k_{\leq d}|$.*

Proof. We prove this lemma by induction on k . The base cases of $k \leq 1$ hold trivially. Suppose the statement is true for contiguous $(\leq k-1)$ -sets. Let S^k be a contiguous k -set and let m be the smallest integer such that $S^k \subseteq \mathcal{D}(m)$.

Case 1: If $x_{n-m+1} = 1$ for all $x \in S^k$, consider $B^k = \{x - e_{n-m+1} : x \in S^k\}$. [Lemma 4.3](#) shows $|\mathbb{M}(k)_{\leq d-1}| \geq |B^k_{\leq d-1}| = |S^k_{\leq d}|$.

Case 2: Define the contiguous k^* -set $S^{k^*} = \{x \in S^k : x_{n-m+1} = 0\}$. We denote $S^{k-k^*} = S^k \setminus S^{k^*}$ and $T = \mathbb{M}(k) \setminus \mathbb{M}(k - k^*)$.

1. If $d \geq m-1$, the claim is trivial since $|\mathbb{M}(k)_{\leq d-1}| = |\mathbb{M}(k)| = k$, as $\mathbb{M}(k) \subseteq \mathcal{D}(m-1)$ due to our assumption of $S^k \cap \mathbb{M}(k) = \emptyset$.

2. If $d \leq m - 2$, using the analysis in Case 1, we get $|\mathsf{M}(k - k^*)_{\leq d-1}| \geq |S_{\leq d}^{k-k^*}|$. It now suffices to show that $|T_{\leq d-1}| \geq |S_{\leq d}^{k^*}|$, where $T = \mathsf{M}(k) \setminus \mathsf{M}(k - k^*)$. Analogous to the argument made in Case 2 in the proof of [Lemma 4.3](#), one can show that $\mathsf{M}(k^*) = \gamma(S^{k^*}, m - 1)$. Since $S^{k^*} \cap (\mathsf{M}(k) \setminus \mathsf{M}(k^*)) = \emptyset$, we also have $\gamma(S^{k^*}, m - 1) \cap \gamma(\mathsf{M}(k) \setminus \mathsf{M}(k^*), m - 1) = \emptyset$ and by the induction hypothesis, we get $|\mathsf{M}(k^*)_{\leq m-d-1}| \geq |\gamma(T, m - 1)_{\leq m-d}|$.

Now, note that $|S_{\leq d}^{k^*}| = k^* - |\mathsf{M}(k^*)_{\leq m-d-1}|$, and since $T \subseteq \mathcal{D}(m - 1)$, we have $|T_{\leq d-1}| = k^* - |\gamma(T, m - 1)_{\leq m-d}|$. Combining these, we get $|T_{\leq d-1}| \geq |S_{\leq d}^{k^*}|$, as desired. \square

We are finally ready to prove the Boolean case of [Theorem 4.1](#).

Proof of [Theorem 4.1](#), the $q = 2$ case. Let $S \subseteq \{0, 1\}^n$ be a down-closed set of size k . We prove this theorem by a simultaneous induction on $k, d \geq 0$.

For the base cases, we consider pairs (k, d) such that $d = 0$ or $k \leq 2^d$. The case of $d = 0$ is trivial. For the case where $k \leq 2^d$, a down-closed set S of size k cannot have strings of Hamming weight $> d$, thereby showing $|S_{\leq d}| = k$. Also, by construction, $\mathsf{M}(k)$ is a down-closed set of size k , implying $\mathcal{H}(d, k) = |\mathsf{M}(k)_{\leq d}| = k$ in this case.

Given $d \geq 1$ and $k > 2^d$, assume that the theorem is true for all (k', d') such that either $k' < k$, or $k' = k$ and $d' < d$. Suppose S is a down-closed set of size k and let m be the smallest integer such that $S \subseteq \mathcal{D}(m)$. Define

$$\begin{aligned} S^0 &:= \{x \in S : x_{n-m+1} = 0\}, \\ S^1 &:= \{x - e_{n-m+1} : x \in S \text{ and } x_{n-m+1} = 1\}. \end{aligned}$$

Since S is down-closed, we have $S^1 \subseteq S^0$. Moreover,

$$|S_{\leq d}| = |S_{\leq d}^0| + |S_{\leq d-1}^1|.$$

Applying the induction hypothesis for $k' = |S^0| < k$ and d , we get $|\mathsf{M}(|S^0|)_{\leq d}| \leq |S_{\leq d}^0|$. Let $T = \mathsf{M}(k) \setminus \mathsf{M}(|S^0|)$. Since $|S^1| \leq |S^0|$, we have $\mathsf{M}(|S^1|) \cap T = \emptyset$, and we may apply [Lemma 4.4](#) to get $|T_{\leq d}| \leq |\mathsf{M}(|S^1|)_{\leq d-1}|$. Now applying the induction hypothesis for $k' = |S^1|$ and $d' = d - 1$, we get $|\mathsf{M}(|S^1|)_{\leq d-1}| \leq |S_{\leq d-1}^1|$. Combining these observations, we get

$$\begin{aligned} |\mathsf{M}(k)_{\leq d}| &= |\mathsf{M}(|S^0|)_{\leq d}| + |T_{\leq d}| \\ &\leq |S_{\leq d}^0| + |\mathsf{M}(|S^1|)_{\leq d-1}| \\ &\leq |S_{\leq d}^0| + |S_{\leq d-1}^1| \\ &= |S_{\leq d}|. \end{aligned}$$

This concludes the induction, and shows that for every $k, d \geq 0$, and down-closed set S of size k , $|\mathsf{M}(k)_{\leq d}| \leq |S_{\leq d}|$. \square

4.2 The General Case of Finite Grids

We prove [Theorem 4.1](#) in this subsection. In fact, we prove the theorem in a more general setting, described as follows.

Let $F = \prod_{i=1}^n \{0, 1, \dots, r_i - 1\}$ where $r_1 \leq r_2 \leq \dots \leq r_n$. Let $d \in \mathbb{N}$. We introduce the following notations:

For $S \subseteq F$, define $\nabla(S) := \{a \in F : b \leq_P a \text{ for some } b \in S\}$, i.e., $\nabla(S)$ is the up-closure of S . For $k \in \{0, \dots, |F|\}$, denote by $M(k)$ the set of the smallest k elements of F in lexicographic order. And for $r \in \{0, \dots, |F_{\leq d}|\}$, denote by $L_{\leq d}(r)$ the set of the largest r elements of $F_{\leq d}$ in lexicographic order.

The main result of this subsection is the following generalization of [Theorem 4.1](#).

Theorem 4.5. *For every $k \in \mathbb{N}$ such that $k \leq |F|$,*

$$\mathcal{H}_F(d, k) = |M(k)_{\leq d}|.$$

We derive [Theorem 4.5](#) from a combinatorial result of Beelen and Datta [[BD18](#)], which generalizes the earlier work of Wei [[Wei91](#)] and Heijnen–Pellikaan [[HP98](#), [Hei99](#)].

Theorem 4.6 ([[BD18](#), Theorem 3.8]). *Let $S \subseteq F_{\leq d}$ and $r = |S|$. Then $|\nabla(L_{\leq d}(r))| \leq |\nabla(S)|$.¹*

Define $\Delta(S) := F \setminus \nabla(S)$ for $S \subseteq F$. The next lemma gives a characterization of $\Delta(S)$.

Lemma 4.7. *Let $T \subseteq F_{\leq d}$ be down-closed and $S = F_{\leq d} \setminus T$. Then $\Delta(S)$ is the unique maximal set with respect to inclusion among all down-closed subsets U of F satisfying $U_{\leq d} = T$.*

Proof. Suppose $\nabla(S) \cap T$ contains an element a . Then $b \leq_P a$ for some $b \in S$. As T is down-closed and $a \in T$, we have $b \in T$ and hence $b \in S \cap T$, contradicting the fact that $S = F_{\leq d} \setminus T$. So $\nabla(S) \cap T = \emptyset$. Therefore, $T \subseteq F_{\leq d} \setminus \nabla(S)$. On the other hand, we have $F_{\leq d} \setminus \nabla(S) \subseteq F_{\leq d} \setminus S \subseteq T$. So $T = F_{\leq d} \setminus \nabla(S)$.

By definition, $\nabla(S)$ is up-closed. So $\Delta(S) = F \setminus \nabla(S)$ is down-closed. And $\Delta(S)_{\leq d} = F_{\leq d} \setminus \nabla(S) = T$ by definition.

Finally, consider an arbitrary down-closed set $U \subseteq F$ satisfying $U_{\leq d} = T$. Then $S \cap U = \emptyset$. Applying the first part of the proof to U in place of T then shows that $\nabla(S) \cap U = \emptyset$. So $U \subseteq F \setminus \nabla(S) = \Delta(S)$. This proves the unique maximality of $\Delta(S)$ with respect to inclusion. \square

Lemma 4.8. *Let $r \in \{0, \dots, |F_{\leq d}|\}$ and $k = |\Delta(L_{\leq d}(r))|$. Then $\Delta(L_{\leq d}(r)) = M(k)$.*

Proof. If $r = 0$, then $\Delta(L_{\leq d}(r)) = \Delta(\emptyset) = F \setminus \nabla(\emptyset) = F$, and the lemma trivially holds. So assume $r > 0$. Let $x = (x_1, \dots, x_n)$ be the smallest element in $L_{\leq d}(r)$.

Let U be the subset of elements in F smaller than x in lexicographic order. Then $U \cap \nabla(L_{\leq d}(r)) = \emptyset$ and hence $U \subseteq \Delta(L_{\leq d}(r))$. It suffices to show that $U = \Delta(L_{\leq d}(r))$. Consider arbitrary $y = (y_1, \dots, y_n) \in F \setminus U$. Then y is at least x in lexicographic order.

We claim that $y \in \nabla(L_{\leq d}(r))$. If $y = x$, then it is in $L_{\leq d}(r) \subseteq \nabla(L_{\leq d}(r))$, so the claim holds in this case. Now suppose y is greater than x in lexicographic order. Let $i \in [n]$ be the smallest integer such that $y_i \neq x_i$. We have $y_i > x_i$. Let $x' = (x_1, \dots, x_{i-1}, x_i + 1, 0, \dots, 0)$.

Suppose x_{i+1}, \dots, x_n are not all zero. Then the generalized Hamming weight of x' is bounded by that of x . So $x' \in L_{\leq d}(r)$. As $x' \leq_P y$, we have $y \in \nabla(L_{\leq d}(r))$.

Now suppose $x_{i+1} = \dots = x_n = 0$. Then $x \leq_P y$. Again, this implies $y \in \nabla(L_{\leq d}(r))$.

So the claim $y \in \nabla(L_{\leq d}(r))$ holds in all cases. Therefore, $y \notin \Delta(L_{\leq d}(r))$. As $y \in F \setminus U$ is arbitrary, this shows $U \supseteq \Delta(L_{\leq d}(r))$ and hence $U = \Delta(L_{\leq d}(r))$. \square

Now we are ready to prove [Theorem 4.5](#).

¹In [[BD18](#)], $L_{\leq d}(r)$ is denoted by $M(r)$, while we use $M(r)$ to denote the set of the smallest r elements of F in lexicographic order.

Proof of Theorem 4.5. For $k \in \{0, 1, \dots, |F|\}$, define $f(k) = |\mathbb{M}(k)_{\leq d}|$. Then

$$f(0) = 0, \quad f(|F|) = |F_{\leq d}|, \quad \text{and} \quad f(k) \leq f(k+1) \leq f(k) + 1 \quad \text{for } k \in \{0, \dots, |F| - 1\}. \quad (5)$$

For $s \in \{0, 1, \dots, |F_{\leq d}|\}$, let $g(s)$ be the largest integer in $\{0, 1, \dots, |F|\}$ such that $f(g(s)) = s$. Note that by (5), $g(\cdot)$ is a well-defined, strictly monotonically increasing function and $g(|F_{\leq d}|) = |F|$.

Fix $k \leq |F|$. By the definition of $\mathcal{H}_F(d, k)$, we want to show that the smallest possible value of $|T_{\leq d}|$ over all down-closed sets $T \subseteq F$ of size k is $|\mathbb{M}(k)_{\leq d}|$. Obviously, this value is attained by choosing $T = \mathbb{M}(k)$.

Now assume to the contrary that there exists a down-closed set $T \subseteq F$ of size k such that $|T_{\leq d}| < |\mathbb{M}(k)_{\leq d}|$, i.e., $|T_{\leq d}| < f(k)$. Let s be the smallest integer in $\{0, 1, \dots, |F_{\leq d}|\}$ such that $g(s) \geq k$. Such s exists since $g(|F_{\leq d}|) = |F|$. Then

$$|T_{\leq d}| < f(k) \leq f(g(s)) = s$$

where the second step uses the monotonicity of $f(\cdot)$. We must have $f(k) > s - 1$, since otherwise we would have $s \geq 1$ and $f(k) = s - 1$, which implies $g(s - 1) \geq k$ by the definition of $g(\cdot)$. But this contradicts the choice of s . Therefore, $f(k) = s$.

Let $S = F_{\leq d} \setminus T$. Let $s' = |T_{\leq d}| < s$ and $r = |S| = |F_{\leq d}| - s'$. Note that $T_{\leq d}$ is down-closed since T and $F_{\leq d}$ are. Applying Lemma 4.7 to $T_{\leq d}$ shows that $\Delta(S)$ is the largest down-closed subset U of F satisfying $U_{\leq d} = T_{\leq d}$. As T is another set satisfying this property, we have

$$|\Delta(S)| \geq |T| = k. \quad (6)$$

By Theorem 4.6, $|\nabla(\mathbb{L}_{\leq d}(r))| \leq |\nabla(S)|$. So $|\Delta(\mathbb{L}_{\leq d}(r))| \geq |\Delta(S)|$. Combining this with (6) gives

$$|\Delta(\mathbb{L}_{\leq d}(r))| \geq k.$$

Let $k' = |\Delta(\mathbb{L}_{\leq d}(r))| \geq k$. By Lemma 4.8, $\Delta(\mathbb{L}_{\leq d}(r)) = \mathbb{M}(k')$. Then $f(k') = |\mathbb{M}(k')_{\leq d}| = |\Delta(\mathbb{L}_{\leq d}(r))_{\leq d}|$. And $\Delta(\mathbb{L}_{\leq d}(r))_{\leq d} = F_{\leq d} \setminus \mathbb{L}_{\leq d}(r)$ by Lemma 4.7. So $f(k') = |F_{\leq d} \setminus \mathbb{L}_{\leq d}(r)| = |F_{\leq d}| - r = s'$.

In summary, we have $s' < s$, $k' \geq k$, $f(k) = s$, and $f(k') = s'$, contradicting the monotonicity of $f(\cdot)$ as stated by (5). \square

5 A Tight Bound on the Size of Degree- d Closures of Sets

For $n, d, \delta \in \mathbb{N}$, denote by $N(n, d, \delta)$ the number of monomials $X_1^{e_1} \cdots X_n^{e_n}$ with $e_1, \dots, e_n \leq \delta$ and $e_1 + \cdots + e_n \leq d$. For example, $N(n, d, 1) = \binom{n}{\leq d}$ and $N(n, d, \delta) = \binom{n+d}{d}$ for $d \leq \delta$.

Lemma 5.1. $\mathfrak{h}_{\mathbb{F}_q^n}(d, \mathbb{F}_q) = N(n, d, q - 1)$.

Proof. By multivariate Lagrange interpolation, every function on \mathbb{F}_q^n can be uniquely written as a linear combination of $X_1^{e_1} \cdots X_n^{e_n} |_{\mathbb{F}_q^n}$ with $e_1, \dots, e_n \leq q - 1$ over \mathbb{F}_q . So the set $\{X_1^{e_1} \cdots X_n^{e_n} |_{\mathbb{F}_q^n} : e_1, \dots, e_n \leq q - 1, e_1 + \cdots + e_n \leq d\}$ is a basis of $\Gamma_{\mathbb{F}_q^n}(d)$ of size $N(n, d, q - 1)$. \square

In particular, Theorem 1.5, which was proved by Nie and Wang [NW15], can be restated as

$$|\text{cl}_d(T)| \leq \frac{q^n}{\mathfrak{h}_{\mathbb{F}_q^n}(d, \mathbb{F}_q)} \cdot |T| = \frac{q^n}{N(n, d, q - 1)} \cdot |T|. \quad (7)$$

We now give the following tight bound on the size of the degree- d closure of a set $T \subseteq \mathbb{F}_q^n$, improving (7).

Theorem 5.2. *Let $n, d, m \in \mathbb{N}$. Let $T \subseteq \mathbb{F}_q^n$ be a set of size m . Then*

$$|\text{cl}_d(T)| \leq \max_{0 \leq k \leq q^n: |\mathbb{M}_q^n(k)_{\leq d}| \leq m} k = \begin{cases} \max_{0 \leq k \leq q^n: |\mathbb{M}_q^n(k)_{\leq d}| = m} k & \text{if } m \leq N(n, d, q-1), \\ q^n & \text{otherwise.} \end{cases} \quad (8)$$

Proof. Note that $|\mathbb{M}_q^n(q^n)_{\leq d}| = N(n, d, q-1)$ and that $|\mathbb{M}_q^n(k-1)_{\leq d}| \leq |\mathbb{M}_q^n(k)_{\leq d}| \leq |\mathbb{M}_q^n(k-1)_{\leq d}| + 1$ for $1 \leq k \leq q^n$. It follows that

$$\max_{0 \leq k \leq q^n: |\mathbb{M}_q^n(k)_{\leq d}| \leq m} k = \begin{cases} \max_{0 \leq k \leq q^n: |\mathbb{M}_q^n(k)_{\leq d}| = m} k & \text{if } m \leq N(n, d, q-1), \\ q^n & \text{otherwise.} \end{cases} \quad (9)$$

This is because if $|\mathbb{M}_q^n(k)_{\leq d}| < m$ and $k < q^n$, we may always increase k by one, which increases $|\mathbb{M}_q^n(k)_{\leq d}|$ by at most one.

Let $T \subseteq \mathbb{F}_q^n$ be a set of size m . Let $U = \text{cl}_d(T)$ and $k_U = |U|$. We want to prove $k_U \leq \max_{0 \leq k \leq q^n: |\mathbb{M}_q^n(k)_{\leq d}| \leq m} k$. Assume to the contrary that this is not true. Then $|\mathbb{M}_q^n(k_U)_{\leq d}| > m$. So we have

$$h_U(d) \geq |\mathbb{M}_q^n(k_U)_{\leq d}| > m = |T| \geq h_T(d), \quad (10)$$

where the first inequality holds by [Corollary 4.2](#). On the other hand, the fact that $U = \text{cl}_d(T)$ implies that $h_U(d) = h_T(d)$, contradicting (10). So $|\text{cl}_d(T)| = k_U \leq \max_{0 \leq k \leq q^n: |\mathbb{M}_q^n(k)_{\leq d}| \leq m} k$. \square

The next theorem states that the bound in [Theorem 5.2](#) is tight and explicitly constructs sets that meet this bound.

Theorem 5.3. *Let $\sigma_A : \mathcal{M} \rightarrow A$ and $\phi : \mathcal{M} \rightarrow F$ be the bijections (2) and (3) respectively, where $A = \mathbb{F}_q^n$, $F = \{0, 1, \dots, q-1\}^n$, and $\mathcal{M} = \{\prod_{i=1}^n X_i^{e_i} : 0 \leq e_1, \dots, e_n \leq q-1\}$. Let m be any integer such that $0 \leq m \leq q^n$. Choose the maximum $k \leq q^n$ such that $|\mathbb{M}_q^n(k)_{\leq d}| \leq m$. Let $T_0 = (\sigma_A \circ \phi^{-1})(\mathbb{M}_q^n(k)_{\leq d}) \subseteq A = \mathbb{F}_q^n$. If $|T_0| \geq m$, let $T = T_0$. Otherwise, let T be an arbitrary set obtained by adding $m - |T_0|$ elements from $\mathbb{F}_q^n \setminus T_0$ to T_0 . Then T is a set of size m that attains the equality in (8).*

Proof. Let $S = (\sigma_A \circ \phi^{-1})(\mathbb{M}_q^n(k)_{\leq d}) \supseteq T_0$. Then $\phi(\sigma_A^{-1}(S)) = \mathbb{M}_q^n(k)_{\leq d}$. By the definitions of $\mathbb{M}_q^n(k)$ and ϕ , the set $\sigma_A^{-1}(S) = \phi^{-1}(\mathbb{M}_q^n(k)_{\leq d}) \subseteq \mathcal{M}$ is down-closed. So by [Lemma 3.5](#), we have

$$h_S(d) = |\phi(\sigma_A^{-1}(S))_{\leq d}| = |\mathbb{M}_q^n(k)_{\leq d}|.$$

Similarly, by the definitions of $\mathbb{M}_q^n(k)_{\leq d}$ and ϕ , the set $\sigma_A^{-1}(T_0) = \phi^{-1}(\mathbb{M}_q^n(k)_{\leq d}) \subseteq \mathcal{M}$ is down-closed. So by [Lemma 3.5](#), we have

$$h_{T_0}(d) = |\phi(\sigma_A^{-1}(T_0))_{\leq d}| = |(\mathbb{M}_q^n(k)_{\leq d})_{\leq d}| = |\mathbb{M}_q^n(k)_{\leq d}|.$$

It follows that $h_S(d) = h_{T_0}(d)$. So we have $S \subseteq \text{cl}_d(T_0)$. Therefore,

$$|\text{cl}_d(T)| \geq |\text{cl}_d(T_0)| \geq |S| = |\mathbb{M}_q^n(k)_{\leq d}| = k = \max_{0 \leq k' \leq q^n: |\mathbb{M}_q^n(k')_{\leq d}| \leq m} k', \quad (11)$$

where the last equality holds by the choice of k .

Suppose $m > N(n, d, q-1)$. Then $k = q^n$. So $|T_0| = |\mathbb{M}_q^n(k)_{\leq d}| = N(n, d, q-1) < m$. Then $|T| = m$ by definition. As $|\text{cl}_d(T)| \geq k = q^n$, we must have $|\text{cl}_d(T)| = q^n$. So T attains the equality in (8).

Now suppose $m \leq N(n, d, q-1)$. Then $|T_0| = |\mathbb{M}_q^n(k)_{\leq d}| = m$ by (9) and the maximality of k . So $T = T_0$ and $|T| = m$. Combining (8) and (11) shows that (8) holds with equality. \square

6 Low-Degree Dispersers

In this section, we will show how to use [Theorem 4.1](#) to conclude the existence of low-degree dispersers for various families of sources. In [Section 6.1](#), we will use [Corollary 4.2](#) to show that for every family of at most $2^{O(k^d)}$ sources of min-entropy k , there exists a degree- d disperser. In particular, this will imply dispersers for local sources and bounded-depth decision forest sources. In [Section 6.2](#), we will extend this result to large families of sources, including polynomial and circuit sources.

6.1 Dispersers for Small Families of Sources

In [Theorem 6.1](#), we use the bound of [Corollary 4.2](#) on the values of Hilbert functions to bound the probability that a random polynomial takes a fixed value on an arbitrary subset of \mathbb{F}_2^n .

Theorem 6.1. *Let $n, d \geq 1$, $S \subseteq \mathbb{F}_2^n$ be an arbitrary nonempty set, and $f : S \rightarrow \mathbb{F}_2$ be a function. Then,*

$$\Pr_{p \in_u \mathcal{P}_2(n, d)} [p|_S \equiv f] \leq 2^{-h_S(d, \mathbb{F}_2)} \leq 2^{-\binom{\lfloor \log_2 |S| \rfloor}{\leq d}}. \quad (12)$$

Proof. Given a function f , let V_f be the subset of $\mathcal{P}_2(n, d)$ consisting of polynomials p such that $p|_S \equiv f$. Note that V_0 is a subspace of $\mathcal{P}_2(n, d)$, and for every $f : S \rightarrow \mathbb{F}_2$, V_f is either empty or is a coset of the subspace V_0 . Thus by definition, for every f ,

$$\Pr_{p \in_u \mathcal{P}_2(n, d)} [p|_S \equiv f] \leq \frac{|V_0|}{|\mathcal{P}_2(n, d)|} = 2^{-h_S(d, \mathbb{F}_2)}.$$

The second inequality in [Equation \(12\)](#) follows from [Corollary 4.2](#). □

We will now use [Theorem 6.1](#) to prove the existence of low-degree dispersers for every small family of sources.

Corollary 6.2. *Let $n, d, k \geq 1$, and \mathcal{X} be a family of distributions of min-entropy $\geq k$ over $\{0, 1\}^n$. Then a uniformly random polynomial $p \in \mathcal{P}_2(n, d)$ is a disperser for \mathcal{X} with probability at least*

$$1 - |\mathcal{X}| \cdot 2^{1 - \binom{k}{\leq d}}.$$

Proof. Let \mathbf{X} be a distribution from \mathcal{X} . Since $H_\infty(\mathbf{X}) \geq k$, we have that $|\text{support}(\mathbf{X})| \geq 2^k$. By [Theorem 6.1](#),

$$\Pr_{p \in_u \mathcal{P}_2(n, d)} [p|_{\text{support}(\mathbf{X})} \text{ is constant}] \leq 2^{1 - \binom{k}{\leq d}}.$$

The corollary follows by applying the union bound over all $|\mathcal{X}|$ sources in \mathcal{X} . □

We will demonstrate two immediate applications of [Corollary 6.2](#) for the families of local and decision forests sources.

Corollary 6.3 (Low-degree dispersers for local sources). *Let $1 \leq \ell \leq d \leq n$ be integers. There exists $p \in \mathcal{P}_2(n, d)$ that is a disperser*

- for the family of ℓ -local sources on $\{0, 1\}^n$ with min-entropy $k > d(2^\ell n + 2\ell n \log n)^{1/d}$.

- for the family of depth- ℓ decision forest sources on $\{0, 1\}^n$ with min-entropy $k > d((\ell + \log n)2^{\ell+1}n)^{1/d}$.

Proof. Let \mathcal{X}_ℓ denote the family of ℓ -local sources on $\{0, 1\}^n$. By [Proposition 2.7](#),

$$|\mathcal{X}_\ell| \leq 2^{2^\ell n + 2\ell n \log n}.$$

Thus, by [Corollary 6.2](#), as long as $\binom{k}{\leq d} > 2^\ell n + 2\ell n \log n + 1$, there exists a degree- d disperser for ℓ -local sources of min-entropy k . For $k > d(2^\ell n + 2\ell n \log n)^{1/d}$, we have that

$$\binom{k}{\leq d} \geq \binom{k}{d} + 1 \geq (k/d)^d + 1 > 2^\ell n + 2\ell n \log n + 1.$$

For depth- ℓ decision forest sources, by [Proposition 2.7](#), the number of such sources on $\{0, 1\}^n$ is at most

$$2^{(\ell + \log n)2^{\ell+1}n}.$$

By [Corollary 6.2](#), as long as $\binom{k}{\leq d} > (\ell + \log n)2^{\ell+1}n + 1$, there exists a degree- d disperser for depth- ℓ decision forest sources of min-entropy k . For $k > d((\ell + \log n)2^{\ell+1}n)^{1/d}$, we have that

$$\binom{k}{\leq d} \geq \binom{k}{d} + 1 \geq (k/d)^d + 1 > (\ell + \log n)2^{\ell+1}n + 1. \quad \square$$

The recent result of [\[ACG⁺22\]](#) uses further properties of local sources to prove the existence of low-degree dispersers for local sources with min-entropy $k \geq c\ell^3 d \cdot (n \log n)^{1/d}$ for a constant $c > 0$. Noting that every depth- ℓ decision forest source is also a $(2^\ell - 1)$ -local source, the disperser of [\[ACG⁺22\]](#) for local sources implies a result similar to the above.

6.2 Dispersers for Polynomial and Circuit Sources

In this section, we will extend the results of the previous section to prove the existence of low-degree dispersers for powerful families of sources including polynomial-size circuits and low-degree polynomial sources. Unlike the previous examples such as local sources, the sources considered here may non-trivially depend on an arbitrary number of inputs. For example, even a degree-1 (i.e. affine) source defined by an affine map $f : \mathbb{F}_2^m \rightarrow \mathbb{F}_2^n$ can depend on an arbitrary number $m \gg n$ of input bits. We get around this by restricting the map $f : \{0, 1\}^m \rightarrow \{0, 1\}^n$ defining the source to a low-dimensional affine subspace. Specifically, we will use the input-reduction procedure from [\[CGG24\]](#), where it was used to prove that random (not necessarily bounded degree) maps extract from low-degree sources.

Lemma 6.4 ([\[CGG24, Lemma 4.5\]](#)). *Let $m, n, k \in \mathbb{N}$, $k > 1$, and $f : \mathbb{F}_2^m \rightarrow \mathbb{F}_2^n$ be a function. If $H_\infty(f(\mathbf{U}_m)) \geq k$, then there exists an affine map $L : \mathbb{F}_2^{11k} \rightarrow \mathbb{F}_2^m$ such that*

$$H_\infty(f(L(\mathbf{U}_{11k}))) \geq k - 1.$$

Equipped with [Lemma 6.4](#), we are ready to construct dispersers for low-degree sources.

Theorem 6.5 (Low-degree disperser for lower-degree polynomial sources). *Let $1 \leq \ell < d \leq n$ be integers. There exists $p \in \mathcal{P}_2(n, d)$ that is a disperser for the family of degree- ℓ sources on $\{0, 1\}^n$ with min-entropy $k \geq (12^\ell \cdot d^d \cdot n)^{\frac{1}{d-\ell}} + 1$.*

In particular, for every $\ell \in \mathbb{N}$, there is a degree- $(\ell + 2)$ disperser for degree- ℓ sources on $\{0, 1\}^n$ with min-entropy $\Omega(\sqrt{n})$.

Proof. Suppose $\mathbf{X} = f(\mathbf{U}_m)$ is a source of min-entropy $H_\infty(\mathbf{X}) \geq k$ defined by a degree- ℓ map $f : \mathbb{F}_2^m \rightarrow \mathbb{F}_2^n$.

We first reduce the number of inputs of the polynomial f . Specifically, we construct another degree- ℓ polynomial g with $r = 11k$ inputs such that every disperser for g is also a disperser for f . We will then conclude the proof by showing that there exists a degree- d disperser for the class of sources defined by polynomials with r inputs.

Let $L : \mathbb{F}_2^r \rightarrow \mathbb{F}_2^m$ be the affine map from [Lemma 6.4](#). Consider the new source $\mathbf{Y} = f(L(\mathbf{U}_r))$ over $\{0, 1\}^n$. By [Lemma 6.4](#), $H_\infty(\mathbf{Y}) \geq k - 1$, and the map $f \circ L$ defining \mathbf{Y} is also a degree- ℓ polynomial with only r input variables. Since $\text{Support}(\mathbf{Y}) \subseteq \text{Support}(\mathbf{X})$, a disperser for \mathbf{Y} is also a disperser for \mathbf{X} .

Therefore, it is now sufficient to prove that there exists a degree- d polynomial p that is a disperser for all degree- ℓ sources \mathbf{Y} with $H_\infty(\mathbf{Y}) \geq k - 1$ defined by polynomials with r inputs. By [Proposition 2.8](#), the number of such sources is bounded from above by $2^{n \cdot \binom{r}{\leq \ell}}$. Now, [Corollary 6.2](#) guarantees the existence of a degree- d disperser as long as

$$\binom{k-1}{\leq d} - 1 > n \cdot \binom{r}{\leq \ell} = n \cdot \binom{11k}{\leq \ell}.$$

For $k \geq (12^\ell \cdot d^d \cdot n)^{\frac{1}{d-\ell}} + 1$, we have that

$$\begin{aligned} \binom{k-1}{\leq d} - 1 &> \binom{k-1}{d} \\ &\geq ((k-1)/d)^d \\ &\geq (k-1)^\ell \cdot (k-1)^{d-\ell}/d^d \\ &\geq (k-1)^\ell \cdot (12^\ell \cdot d^d \cdot n)/d^d \\ &= n(12(k-1))^\ell \\ &\geq n \left((11k)^\ell + 1 \right) \\ &\geq n \cdot \binom{11k}{\leq \ell}, \end{aligned}$$

which finishes the proof of the theorem. \square

Theorem 6.6 (Low-degree disperser for circuit sources). *Let $\ell \geq 1$ and $n \geq d \geq 2\ell + 2$ be integers. There exists $p \in \mathcal{P}_2(n, d)$ that is a disperser for the family of n^ℓ -size circuit sources on $\{0, 1\}^n$ with min-entropy $k \geq (30^2 \cdot d^d \cdot n^{2\ell})^{\frac{1}{d-2}} + 1$.*

Proof. Let $\mathbf{X} = C(\mathbf{U}_m)$ be a source of min-entropy $H_\infty(\mathbf{X}) \geq k$ defined by an $\text{AC}[\oplus]$ circuit, $C : \mathbb{F}_2^m \rightarrow \mathbb{F}_2^n$ of size n^ℓ .

Similarly to the proof of [Theorem 6.5](#), we first reduce the number of inputs of the circuit C . Let $L : \mathbb{F}_2^r \rightarrow \mathbb{F}_2^m$ be the affine map from [Lemma 6.4](#) for $r = 11k$. Consider the circuit source $\mathbf{Y} = C(L(\mathbf{U}_r))$.

By [Lemma 6.4](#), $H_\infty(\mathbf{Y}) \geq k - 1$, and since $\text{Support}(\mathbf{Y}) \subseteq \text{Support}(\mathbf{X})$, a disperser for \mathbf{Y} is also a disperser for \mathbf{X} .

We will now show that there is an $\text{AC}[\oplus]$ circuit $C' : \mathbb{F}_2^r \rightarrow \mathbb{F}_2^n$ of size $|C'| \leq rn^\ell$ computing $C'(y) = C(L(y))$ for every $y = (y_1, \dots, y_r) \in \mathbb{F}_2^r$. The main difference between the circuits C and

C' is that we need to replace each input x_i for $i \in [m]$ of the circuit C by an affine form $L_i(y_1, \dots, y_r)$ of the inputs of C' given by L . The naive way of implementing this modification would require $m \gg n^\ell$ additional gates which we cannot afford. Instead, we will simulate this modification by updating the input wires of each gate in C' accordingly (and introducing at most r additional gates for each gate in C).

Every XOR gate of C fed by inputs x_{i_1}, \dots, x_{i_t} will now be replaced by an XOR gate in C' computing $L_{i_1}(y_1, \dots, y_r) \oplus \dots \oplus L_{i_t}(y_1, \dots, y_r)$. Note that this operation only modifies the wires in the circuit and does not introduce new gates. Every AND gate of C fed by inputs x_{i_1}, \dots, x_{i_t} will now be replaced by a gate computing AND of at most r linearly independent forms from L_{i_1}, \dots, L_{i_t} . This operation introduces at most r additional XOR gates for each gate of the original circuit C . We handle OR gates in a similar fashion. Finally, the wires between non-input gates of the circuit C stay unchanged in the circuit C' . We constructed a circuit of size rn^ℓ with r inputs that defines the source \mathbf{Y} .

It remains to show the existence of a degree- d polynomial p that is a disperser for all sources \mathbf{Y} with $H_\infty(\mathbf{Y}) \geq k - 1$ defined by rn^ℓ -size circuits with r inputs. By [Proposition 2.8](#), the number of such sources is bounded from above by

$$2^{4rn^\ell(rn^\ell+r)} = 2^{4r^2n^\ell(n^\ell+1)} = 2^{4(11k)^2n^\ell(n^\ell+1)} \leq 2^{(27k)^2n^{2\ell}}.$$

Now, [Corollary 6.2](#) guarantees the existence of a degree- d disperser as long as

$$\binom{k-1}{\leq d} - 1 > (27k)^2n^{2\ell}.$$

For $k \geq (30^2 \cdot d^d \cdot n^{2\ell})^{\frac{1}{d-2}} + 1$, we have that

$$\begin{aligned} \binom{k-1}{\leq d} - 1 &> \binom{k-1}{d} \\ &\geq ((k-1)/d)^d \\ &\geq (k-1)^2 \cdot (k-1)^{d-2}/d^d \\ &\geq (k-1)^2 \cdot (30^2 \cdot d^d \cdot n^{2\ell})/d^d \\ &\geq (30(k-1))^2 n^{2\ell} \\ &\geq (27k)^2 n^{2\ell}, \end{aligned}$$

which finishes the proof. \square

[Theorems 6.5](#) and [6.6](#) construct low-degree dispersers for sources generated by constant-degree polynomials and polynomial-size $\text{AC}[\oplus]$ circuits. These two classes of sources are incomparable. Indeed, $\text{AC}[\oplus]$ computes $\text{AND}(x_1, \dots, x_m)$ which is not a constant-degree polynomial, while constant-degree polynomials compute polynomials in m inputs which do not admit circuits of size polynomial in n . We remark that the techniques of [Theorems 6.5](#) and [6.6](#) can be used to conclude the same result for a class of sources that generalizes both $\text{AC}[\oplus]$ and constant-degree polynomials. This is the class of polynomial-size circuits which extends $\text{AC}[\oplus]$ with gates computing arbitrary polynomials in m inputs of a fixed constant degree. For ease of exposition, we present only the results for more natural sources in [Theorems 6.5](#) and [6.6](#).

7 Random Low-Degree Polynomials Extract from Fixed Sources

In this section, we use our bounds on the values of Hilbert functions to prove the existence of a low-degree extractor for a fixed high min-entropy source. Specifically, in [Theorem 7.3](#) we show that for every source \mathbf{X} of high min-entropy, a random low-degree polynomial p has bias $\leq \varepsilon$, i.e., $\Pr_{x \in_u \mathcal{X}}[f(x) = 1] \in 1/2 \pm \varepsilon$ with high probability. One special case of interest is the case of *k-flat* sources \mathbf{X} which are uniform distributions over sets of size 2^k . In [Section 8](#), we will use [Theorem 7.3](#) to prove the existence of low-degree extractors for various expressive families of sources.

We start this section by using our bounds on the degree- d closure of sets in order to lower-bound the probability that a random somewhat large subset T of a set S has “full Hilbert dimension”, i.e., $h_T(d, \mathbb{F}_2) = |T|$. We then use this to prove [Lemma 7.2](#) which states that for a large enough set $S \subseteq \{0, 1\}^n$, a random subset $T \subseteq S$ of full Hilbert dimension will contain each element $x \in S$ with almost the same probability. Finally, we present a proof of [Theorem 7.3](#) which crucially relies on [Lemma 7.2](#).

Claim 7.1. *Let $1 \leq d \leq n$, $d \leq \ell$, and $S \subseteq \{0, 1\}^n$. Let T be a uniformly random subset of S of size $\binom{\ell}{\leq d}$. Then*

$$\Pr_T \left[h_T(d, \mathbb{F}_2) = |T| = \binom{\ell}{\leq d} \right] \geq 1 - \binom{\ell}{\leq d} \cdot 2^\ell / |S|.$$

Proof. For a random subset $T = \{a_1, a_2, \dots, a_{\binom{\ell}{\leq d}}\}$ and $j \in \left[\binom{\ell}{\leq d} \right]$, let $T^j = \{a_1, \dots, a_j\}$. We prove that with high probability, $h_{T^{j+1}}(d, \mathbb{F}_2) = h_{T^j}(d, \mathbb{F}_2) + 1$ holds for every j . Indeed, for every $j < \binom{\ell}{\leq d}$, we have that

$$\begin{aligned} \Pr [h_{T^{j+1}}(d, \mathbb{F}_2) = h_{T^j}(d, \mathbb{F}_2) + 1] &= \Pr [a_{j+1} \notin \text{cl}_d(T^j)] \\ &= 1 - |\text{cl}_d(T^j)| / |S| \\ &\geq 1 - 2^\ell / |S|, \end{aligned}$$

where the last inequality uses [Corollary 1.7](#). The claim now follows by the union bound over all j . \square

Lemma 7.2. *Let $1 \leq d \leq n$, $d \leq \ell$, and $S \subseteq \{0, 1\}^n$. Let T be a uniformly random subset of S of size $\binom{\ell}{\leq d}$. Then for every $x \in S$,*

$$(1 - \delta) \cdot \frac{\binom{\ell}{\leq d}}{|S|} \leq \Pr_T [x \in T \mid h_T(d, \mathbb{F}_2) = |T|] \leq \frac{1}{(1 - \delta)} \cdot \frac{\binom{\ell}{\leq d}}{|S|},$$

where $\delta = \binom{\ell}{\leq d} \cdot 2^\ell / |S|$.

Proof. By the Bayes rule, we have

$$\begin{aligned} \Pr_T [x \in T \mid h_T(d, \mathbb{F}_2) = |T|] &= \frac{\Pr[h_T(d, \mathbb{F}_2) = |T| \mid x \in T] \cdot \Pr[x \in T]}{\Pr[h_T(d, \mathbb{F}_2) = |T|]} \\ &= \frac{\Pr[h_T(d, \mathbb{F}_2) = |T| \mid x \in T]}{\Pr[h_T(d, \mathbb{F}_2) = |T|]} \cdot \frac{\binom{\ell}{\leq d}}{|S|}. \end{aligned}$$

For the upper bound, by [Claim 7.1](#), $\Pr[h_T(d, \mathbb{F}_2) = |T|] \geq 1 - \binom{\ell}{\leq d} \cdot 2^\ell / |S| = 1 - \delta$. Thus,

$$\Pr[x \in T \mid h_T(d, \mathbb{F}_2) = |T|] \leq \frac{1}{(1 - \delta)} \cdot \frac{\binom{\ell}{\leq d}}{|S|}.$$

For the lower bound, first include x in T , and then randomly pick the remaining $\binom{\ell}{\leq d} - 1$ elements in T . Analogous to the analysis in [Claim 7.1](#), we have $\Pr_T[h_T(d, \mathbb{F}_2) = |T| \mid x \in T] \geq 1 - \binom{\ell}{\leq d} \cdot 2^\ell / |S| = 1 - \delta$. Thus,

$$\Pr_T[x \in T \mid h_T(d, \mathbb{F}_2) = |T|] \geq (1 - \delta) \cdot \frac{\binom{\ell}{\leq d}}{|S|},$$

which concludes the proof. \square

Equipped with [Lemma 7.2](#), we are ready to present the proof of [Theorem 7.3](#).

Theorem 7.3. *Let $n, d, k \geq 1$, and $\varepsilon > 0$ be a real. Then for every distribution \mathbf{X} over $\{0, 1\}^n$ with $H_\infty(\mathbf{X}) \geq k$, a uniformly random degree- d polynomial f is an ε -extractor for \mathbf{X} ,*

$$\Pr_{x \sim \mathbf{X}}[f(x) = 1] = \frac{1}{2} \pm \varepsilon$$

with probability at least $1 - e^{3n - \varepsilon^2 \binom{\ell}{\leq d} / (Cn^2)}$ where $\ell = k/2 - \log(32n/\varepsilon)$ and $C = 7 \cdot (32)^2$.

Proof. We will assume that $\ell := k/2 - \log(32n/\varepsilon) > 0$ (and, in particular, that $\varepsilon \geq 2^{-n}$) as otherwise the theorem statement holds trivially. Similarly, we will assume that $3n - \varepsilon^2 \binom{\ell}{\leq d} / (Cn^2) < 0$, and, in particular, that $2^{n+1} \cdot e^{-(\varepsilon/(32n))^2 \cdot \binom{\ell}{\leq d} / 7} < 1$.

Let \mathbf{X} be a distribution over $\{0, 1\}^n$ with $H_\infty(\mathbf{X}) \geq k$, and let $S = \text{support}(\mathbf{X}) \subseteq \{0, 1\}^n$. For $x \in S$, let $p_x = \Pr[\mathbf{X} = x]$, so that $p_x \leq 2^{-k}$ and $\sum_{x \in S} p_x = 1$. For every $i \geq k$, define

$$S_i := \{x \in S : 2^{-i-1} < p_x \leq 2^{-i}\},$$

and note that $|S_i| \leq 2^{i+1}$. We will show that with high probability a random degree- d polynomial f satisfies

$$\sum_{i \geq k} \left| \sum_{x \in S_i} p_x f(x) - \frac{1}{2} \sum_{x \in S_i} p_x \right| \leq \varepsilon. \quad (13)$$

Given [Equation \(13\)](#), we finish the proof of the theorem as follows.

$$\begin{aligned} \left| \Pr_{x \sim \mathbf{X}}[f(x) = 1] - 1/2 \right| &= \left| \sum_{x \in S} p_x f(x) - 1/2 \right| \\ &= \left| \sum_{x \in S} p_x f(x) - \frac{1}{2} \sum_{x \in S} p_x \right| \\ &\leq \sum_{i \geq k} \left| \sum_{x \in S_i} p_x f(x) - \frac{1}{2} \sum_{x \in S_i} p_x \right| \\ &\leq \varepsilon. \end{aligned}$$

It remains to prove [Equation \(13\)](#), and we will do this by analyzing the terms for different values of i separately.

Large i . Let S_L be the union of all S_i with $i \geq 2n + 1$. Note that

$$\sum_{x \in S_L} p_x \leq 2^n \cdot 2^{-2n-1} \leq \varepsilon/2,$$

where we used that $\varepsilon \geq 2^{-n}$. Thus, in the rest of the proof we will only deal with $k \leq i \leq 2n$. Specifically, we will show that for each i in this interval, $|\sum_{x \in S_i} p_x f(x) - \frac{1}{2} \sum_{x \in S_i} p_x| \leq 8\varepsilon'$ for

$$\varepsilon' := \frac{\varepsilon}{32n}.$$

We will assume that $\varepsilon \leq 1/2$ (as the theorem statement is trivial otherwise), and, thus, $\varepsilon' \leq 1/64$.

Small S_i . If $|S_i| \leq 2^{2\ell+1}/\varepsilon'$, then

$$\sum_{x \in S_i} p_x \leq 2^{2\ell+1-i}/\varepsilon' \leq 2^{2\ell+1-k}/\varepsilon' \leq \varepsilon',$$

where we used that $i \geq k$ and $\ell = k/2 - \log(32n/\varepsilon) = k/2 - \log(1/\varepsilon')$.

Small i and large S_i . Let I be the set of $i \leq 2n$ such that $|S_i| > 2^{2\ell+1}/\varepsilon'$.

Fix an $i \in I$ and let $t = |S_i|$. We independently and uniformly at random pick subsets T_1, \dots, T_t of S_i of size $|T_j| = \binom{\ell}{\leq d}$ satisfying $h_{T_j}(d, \mathbb{F}_2) = |T_j|$. For any $x \in S_i$, let n_x denote the number of sets T_j containing x : $n_x = |\{j \in [t]: x \in T_j\}|$. Note that by [Lemma 7.2](#),

$$(1 - \delta) \cdot \binom{\ell}{\leq d} \leq \mathbb{E}[n_x] \leq \frac{1}{(1 - \delta)} \cdot \binom{\ell}{\leq d},$$

where $\delta = \binom{\ell}{\leq d} \cdot 2^\ell / |S_i|$. Since $|S_i| > 2^{2\ell+1}/\varepsilon'$, we have that $\delta \leq \varepsilon'/2 \leq 1/2$, and

$$\mathbb{E}[n_x] = (1 \pm \varepsilon') \cdot \binom{\ell}{\leq d}.$$

Applying the Chernoff bound for the concentration of n_x and the union bound over all t choices of $x \in S_i$, we have that

$$\Pr_{T_1, \dots, T_t} \left[\exists x \in S_i, n_x \notin (1 \pm 2\varepsilon') \cdot \binom{\ell}{\leq d} \right] \leq 2t \cdot e^{-\varepsilon'^2 \cdot \binom{\ell}{\leq d} / 7} \leq 2^{n+1} \cdot e^{-(\varepsilon/(32n))^2 \cdot \binom{\ell}{\leq d} / 7} < 1.$$

In particular, there exists a choice of T_1, \dots, T_t such that $h_{T_j}(d, \mathbb{F}_2) = |T_j|$ for all $j \in [t]$, and each $n_x = (1 \pm 2\varepsilon') \cdot \binom{\ell}{\leq d}$. Let us fix this choice of sets T_1, \dots, T_t .

For a fixed set T_j such that $h_{T_j}(d, \mathbb{F}_2) = |T_j|$, a random degree- d polynomial f satisfies

$$\Pr_{f \in \mathcal{U}\mathcal{P}_2(n, d)} \left[\left| \sum_{x \in T_j} p_x f(x) - \frac{1}{2} \sum_{x \in T_j} p_x \right| \geq \frac{\varepsilon' \binom{\ell}{\leq d}}{2^i} \right] \leq 2e^{-2\varepsilon'^2 \binom{\ell}{\leq d}}. \quad (14)$$

Indeed, $h_{T_j}(d, \mathbb{F}_2) = |T_j|$ implies that a random degree- d polynomial f induces a random map $f|_{T_j}: T_j \rightarrow \{0, 1\}$ (see, e.g., the proof of [Theorem 6.1](#)). [Equation \(14\)](#) now follows from the

Hoeffding bound, as the random variables $2^i p_x f(x)$ are $\binom{\ell}{\leq d}$ independent $[0, 1]$ -valued random variables with mean $2^{i-1} p_x$.

Taking the union bound over all $j \in [t]$, we have that

$$\Pr_{f \in_u \mathcal{P}_2(n, d)} \left[\forall j \in [t], \left| \sum_{x \in T_j} p_x f(x) - \frac{1}{2} \sum_{x \in T_j} p_x \right| \leq \frac{\varepsilon' \binom{\ell}{\leq d}}{2^i} \right] \geq 1 - 2te^{-2\varepsilon'^2 \binom{\ell}{\leq d}}.$$

Therefore, using $t \leq 2^n$, with probability at least

$$1 - 2^{n+1} \cdot e^{-2\varepsilon'^2 \binom{\ell}{\leq d}},$$

we simultaneously have that (i) for every $x \in S_i$, $n_x = (1 \pm 2\varepsilon') \cdot \binom{\ell}{\leq d}$, and (ii) for every $j \in [t]$,

$$\left| \sum_{x \in T_j} p_x f(x) - \frac{1}{2} \sum_{x \in T_j} p_x \right| \leq \frac{\varepsilon' \binom{\ell}{\leq d}}{2^i}.$$

Conditioning on this good event, we get

$$\begin{aligned} \sum_{x \in S_i} p_x f(x) &= \sum_{x \in S_i} \frac{n_x}{n_x} p_x f(x) \\ &= \frac{1}{(1 \pm 2\varepsilon') \cdot \binom{\ell}{\leq d}} \cdot \sum_{j \in [t]} \sum_{x \in T_j} p_x f(x) \\ &= \frac{1}{(1 \pm 2\varepsilon') \cdot \binom{\ell}{\leq d}} \cdot \sum_{j \in [t]} \left(\left(\frac{1}{2} \sum_{x \in T_j} p_x \right) \pm \frac{\varepsilon' \binom{\ell}{\leq d}}{2^i} \right) \\ &= \frac{1}{(1 \pm 2\varepsilon') \cdot \binom{\ell}{\leq d}} \cdot \left(\frac{1}{2} \sum_{x \in S_i} n_x p_x \right) \pm \frac{\varepsilon' t}{(1 \pm 2\varepsilon') \cdot 2^i} \\ &= \frac{(1 \pm 2\varepsilon')}{(1 \pm 2\varepsilon')} \cdot \left(\frac{1}{2} \sum_{x \in S_i} n_x p_x \right) \pm \frac{2\varepsilon'}{(1 \pm 2\varepsilon')} \\ &= \left(\frac{1}{2} \sum_{x \in S_i} p_x \right) \pm 8\varepsilon', \end{aligned}$$

where the penultimate equality uses $t = |S_i| \leq 2^{i+1}$.

Finally, we apply the union bound over all $i \in I$. Since $|I| \leq 2n$, combining the contributions from small S_i 's and large i 's, we conclude that

$$\left| \frac{1}{2} \sum_{x \in S_i} p_x - \sum_{x \in S_i} p_x f(x) \right| \leq \frac{\varepsilon}{2} + 8\varepsilon' \cdot |I| \leq \varepsilon$$

with probability at least

$$1 - n \cdot 2^{n+2} \cdot e^{-2\varepsilon'^2 \binom{\ell}{\leq d}} \geq 1 - e^{3n - \varepsilon^2 \binom{\ell}{\leq d} / (512n^2)} \geq 1 - e^{3n - \varepsilon^2 \binom{\ell}{\leq d} / (Cn^2)}. \quad \square$$

8 Low-Degree Extractors

In this section, we extend the results of [Section 6](#) to the setting of extractors. We start with the extractors version of [Corollary 6.2](#) in [Theorem 8.1](#), where we show that low-degree polynomials extract from small families of sources. Then, in [Theorem 8.3](#), we use [Theorem 8.1](#) to prove the existence of low-degree extractors for a number of families of sources. Finally, in [Section 8.1](#), we prove the existence of low-degree extractors with multi-bit outputs.

Theorem 8.1. *Let \mathcal{X} be a family of distributions of min-entropy $k \geq 5 \log n$ over $\{0, 1\}^n$ for large enough n . Let \mathcal{Y} be a family of distributions each of which is ε' -close to a convex combination of distributions from \mathcal{X} . Then for every $d \geq 6$, a uniformly random polynomial $p \in \mathcal{P}_2(n, d)$ is an ε -extractor for \mathcal{Y} with probability at least*

$$1 - |\mathcal{X}| \cdot e^{3n - 30k^{d/2}/n^2}$$

for $\varepsilon = (2d/k^{1/4})^d + \varepsilon'$.

Proof. Let $\delta = \varepsilon - \varepsilon' = (2d/k^{1/4})^d$. We first bound the probability that a random polynomial is a δ -extractor for a fixed source $\mathbf{X} \in \mathcal{X}$, and then apply the union bound over all sources from \mathcal{X} .

Let $\ell = k/2 - \log(32n/\delta) \geq k/3$ for all large enough n . Let $C = 7 \cdot (32)^2$. By [Theorem 7.3](#), we have that the probability that a random degree- d polynomial is *not* a δ -extractor for a fixed source $\mathbf{X} \in \mathcal{X}$ is at most

$$\begin{aligned} e^{3n - \delta^2 \binom{\ell}{\leq d} / (Cn^2)} &\leq e^{3n - (2d)^{2d} \cdot k^{-d/2} \cdot (\ell/d)^d / (Cn^2)} \\ &\leq e^{3n - 30k^{d/2}/n^2} . \end{aligned}$$

Now the union bound over all sources from \mathcal{X} gives us that a random degree- d polynomial is a δ -extractor for every source in \mathcal{X} with probability at least

$$1 - |\mathcal{X}| \cdot e^{3n - 30k^{d/2}/n^2} .$$

Finally, each polynomial that δ -extracts from \mathcal{X} is also an ε -extractor for all sources in \mathcal{Y} . □

We will use the following input-reduction result from [\[CGG24\]](#).

Theorem 8.2 ([\[CGG24, Theorem 4.1\]](#)). *Let $m, n, k \in \mathbb{N}$, $k > 1$, and $f: \mathbb{F}_2^m \rightarrow \mathbb{F}_2^n$ be a function. If $H_\infty(f(\mathbf{U}_m)) \geq k$, then there exist affine maps $L_1, \dots, L_t: \mathbb{F}_2^{11k} \rightarrow \mathbb{F}_2^m$ such that the distribution $f(\mathbf{U}_m)$ is 2^{-k} -close to a convex combination of distributions $f(L_i(\mathbf{U}_{11k}))$. Moreover, for each $i \in [t]$,*

$$H_\infty(f(L_i(\mathbf{U}_{11k}))) \geq k - 1 .$$

We are now ready to prove that low-degree polynomials extract from many sources of interest.

Theorem 8.3. *For all $\ell, d \geq 1$, and all large enough n , there exists $p \in \mathcal{P}_2(n, d)$ that is an ε -extractor for the following families of sources over $\{0, 1\}^n$ of min-entropy $k \geq 5 \log n$ for $\varepsilon = 2(2d/k^{1/4})^d$.*

- ℓ -local sources for $k \geq (2^\ell n^3 \log n)^{2/d}$.

- depth- ℓ decision forest sources for $k \geq (2^\ell n^3 (\log n + \ell))^{2/d}$.
- degree- ℓ sources for $k \geq (3^\ell n)^{\frac{6}{d-2\ell}}$.
- n^ℓ -size circuit sources for $k \geq 3n^{\frac{4(\ell+1)}{d-4}}$.

Proof. By [Proposition 2.7](#), the number of ℓ -local sources over $\{0, 1\}^n$ is at most $2^{2^\ell n + 2\ell n \log n}$, and the number of depth- ℓ decision forest sources is at most $2^{(\ell + \log n)2^{\ell+1}n}$. Now [Theorem 8.1](#) implies the result for these classes of sources.

Let \mathcal{Y} be the family of degree- ℓ sources. We first apply [Theorem 8.2](#), and obtain the family \mathcal{X} of degree- ℓ sources such that each source in \mathcal{Y} is 2^{-k} -close to a convex combination of sources in \mathcal{X} . Moreover, each source in \mathcal{X} is a degree- ℓ polynomial in $11k$ variables and has entropy $\geq k - 1$. By [Proposition 2.8](#), the number of such sources is bounded from above by $|\mathcal{X}| \leq 2^{n \cdot \binom{11k}{\leq \ell}}$. Now, [Theorem 8.1](#) guarantees the existence of a degree- d ε' -extractor for \mathcal{Y} for $k \geq (3^\ell n)^{\frac{6}{d-2\ell}}$ and $\varepsilon' = (2d/k^{1/4})^d + 2^{-k} \leq \varepsilon$, where the last inequality uses $d \leq k^{1/4}$.

When \mathcal{Y} is the family of n^ℓ -size circuit sources, using [Theorem 8.2](#) and the argument from [Theorem 6.6](#), we obtain the family \mathcal{X} of $(11kn^\ell)$ -size circuits with $11k$ inputs and entropy $k - 1$. Moreover, each source from \mathcal{Y} is 2^{-k} -close to a convex combination of sources from \mathcal{X} . By [Proposition 2.8](#), $|\mathcal{X}| \leq 2^{(27k)^2 n^{2\ell}}$. Now [Theorem 8.1](#) guarantees the existence of a degree- d ε' -extractor for \mathcal{Y} for $k \geq 3n^{\frac{4(\ell+1)}{d-4}}$ and $\varepsilon' = (2d/k^{1/4})^d + 2^{-k} \leq \varepsilon$. \square

8.1 Extractors Outputting Multiple Bits

In [Theorem 8.4](#), we show how to extend our single-bit extractors for small families of sources to the multi-bit setting, which combined with input-reduction lemma, will extend all our single-bit extractors from [Theorem 8.3](#) to $O(k)$ -bit extractors.

Theorem 8.4. *Let \mathcal{X} be a family of distributions of min-entropy $k \geq 5 \log n$ over $\{0, 1\}^n$ for large enough n . Let \mathcal{Y} be a family of distributions each of which is ε' -close to a convex combination of distributions from \mathcal{X} . Then for every $d \geq 6$ and $t < k$, let $p_1, \dots, p_t \in \mathcal{P}_2(n, d)$ be independent and uniformly random polynomials. Then $p = (p_1, \dots, p_t)$ is a $t\varepsilon$ -extractor for \mathcal{Y} with probability at least*

$$1 - |\mathcal{X}| \cdot e^{3n+t+1-30(k-2t)^{d/2}/n^2}$$

for $\varepsilon = (2d/k^{1/4})^d + \varepsilon'$, assuming $\varepsilon \leq 1/4$.

Proof. Define \mathcal{X}_i to be the family of sources resulting by conditioning sources $\mathbf{X} \in \mathcal{X}$ on $(p_1, \dots, p_i) = (b_1, \dots, b_i) \in \mathbb{F}_2^n$ for any $(b_1, \dots, b_i) \in \{0, 1\}^i$, so that $\mathcal{X}_0 = \mathcal{X}$, and $|\mathcal{X}_i| = 2^i |\mathcal{X}|$. Let E_i be the event that p_i is an ε -extractor for \mathcal{X}_i , and let $E_{\leq i} = E_1 \wedge \dots \wedge E_i$.

Note that, conditioned on $E_{\leq i}$, every source in \mathcal{X}_{i+1} consists of $\leq 2^{i+1} |\mathcal{X}|$ sources of min-entropy $\geq k - 2i$. This can be shown by an induction. The base case is true since $\mathcal{X}_1 = \mathcal{X}$, and \mathcal{X} has min-entropy $\geq k$. For the inductive step, let $\mathbf{X}' \in \mathcal{X}_{i+1}$ be obtained by conditioning $\mathbf{X} \in \mathcal{X}_i$ on $p_i = b$ for some $b \in \{0, 1\}$. Now by the Bayes rule, we have for every x in the support of \mathbf{X}'

$$\Pr[\mathbf{X}' = x] = \Pr[\mathbf{X} = x | p_i(X) = b] = \frac{\Pr[p_i(\mathbf{X}) = b | \mathbf{X} = x] \cdot \Pr[\mathbf{X} = x]}{\Pr[p_i(\mathbf{X}) = b]} \leq \frac{2^{-(k-2(i-1))}}{\frac{1}{2} - \varepsilon} \leq 2^{-(k-2i)},$$

as long as $\varepsilon \leq \frac{1}{4}$, which shows that \mathbf{X}' has min-entropy at least $k - 2i$ as desired.

Thus, by [Theorem 8.1](#),

$$\Pr[E_{i+1}|E_{\leq i}] \geq 1 - |\mathcal{X}_{i+1}| \cdot e^{3n-30(k-2t)^{d/2}/n^2} \geq 1 - 2^{i+1}|\mathcal{X}| \cdot e^{3n-30(k-2t)^{d/2}/n^2}.$$

Thus by the chain rule, we have

$$\begin{aligned} \Pr[E_1 \wedge \dots \wedge E_t] &= \prod_{i=0}^{t-1} \Pr[E_{i+1}|E_{\leq i}] \\ &\geq \prod_{i=0}^{t-1} \left(1 - 2^{i+1}|\mathcal{X}| \cdot e^{3n-30(k-2t)^{d/2}/n^2}\right) \\ &\geq 1 - \sum_{i=0}^{t-1} \left(2^{i+1}|\mathcal{X}| \cdot e^{3n-30(k-2t)^{d/2}/n^2}\right) \\ &\geq 1 - |\mathcal{X}| \cdot e^{t+1+3n-30(k-2t)^{d/2}/n^2}. \end{aligned}$$

Next, we analyze the error of the extractor conditioned on the above event $E_{\leq t}$. Let $\mathbf{X} \in \mathcal{X}$ be a fixed source. Let (u_1, u_2, \dots, u_t) be uniformly distributed over $\{0, 1\}^t$ independent of \mathbf{X} and p_1, \dots, p_t , and define random variables D_0, \dots, D_t as follows. For every i , define

$$D_i = (p_1(\mathbf{X}), p_2(\mathbf{X}), \dots, p_i(\mathbf{X}), u_{i+1}, \dots, u_t).$$

Note that D_0 is the uniform distribution and $D_t = p(\mathbf{X})$ is the output of the random degree- d polynomial conditioned on $E_{\leq t}$. By the triangle inequality, we have

$$\Delta(D_0, D_t) \leq \sum_{i=0}^{t-1} \Delta(D_i, D_{i+1}).$$

Thus it suffices to bound each $\Delta(D_i, D_{i+1})$ by ε , which can be obtained by noting that conditioning \mathbf{X} on any value of f_1, \dots, f_i results in a source that belongs to \mathbf{X}_i for which f_{i+1} is an ε -extractor. Since $\Delta(D_i, D_{i+1})$ is a convex combination of the bias of f_{i+1} for such fixings, we obtain $\Delta(D_i, D_{i+1}) \leq \varepsilon$ as desired. \square

References

- [ACG⁺22] Omar Alrabiah, Eshan Chattopadhyay, Jesse Goodman, Xin Li, and João Ribeiro. Low-degree polynomials extract from local sources. In *ICALP, 2022*. [2](#), [24](#)
- [AGMR24] Omar Alrabiah, Jesse Goodman, Jonathan Mosheiff, and João Ribeiro. Low-degree polynomials are good extractors. *Manuscript*, 2024. [10](#)
- [BD18] Peter Beelen and Mrinmoy Datta. Generalized Hamming weights of affine Cartesian codes. *Finite Fields and Their Applications*, 51:130–145, 2018. [1](#), [4](#), [20](#)
- [BEHL12] Ido Ben-Eliezer, Rani Hod, and Shachar Lovett. Random low-degree polynomials are hard to approximate. *computational complexity*, 21(1):63–81, 2012. [1](#), [4](#)

- [BG13] Andrej Bogdanov and Siyao Guo. Sparse extractor families for all the entropy. In *ITCS*, 2013. 2
- [BGY20] 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. 1, 4
- [Blu86] Manuel Blum. Independent unbiased coin flips from a correlated biased source—a finite state Markov chain. *Combinatorica*, 6:97–108, 1986. 2
- [BM93] Dave Bayer and David Mumford. What can be computed in algebraic geometry? *arXiv preprint alg-geom/9304003*, 1993. 3
- [CG88] Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing (SICOMP)*, 17(2):230–261, 1988. 2
- [CGG24] Eshan Chattopadhyay, Jesse Goodman, and Mohit Gurumukhani. Extractors for polynomial sources over \mathbb{F}_2 . In *ITCS*, 2024. 3, 6, 24, 31
- [CL18] Kuan Cheng and Xin Li. Randomness extraction in AC^0 and with small locality. In *RANDOM*, 2018. 2
- [CLO13] David Cox, John Little, and Donal O’Shea. *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra*. Springer, 2013. 13
- [CT15] Gil Cohen and Avishay Tal. Two structural results for low degree polynomials and applications. In *RANDOM*, 2015. 2
- [DOPS04] Yevgeniy Dodis, Shien Jin Ong, Manoj Prabhakaran, and Amit Sahai. On the (im)possibility of cryptography with imperfect randomness. In *FOCS*, 2004. 2
- [DTT22] Dean Doron, Amnon Ta-Shma, and Roei Tell. On hitting-set generators for polynomials that vanish rarely. *Computational Complexity*, 31(2):16, 2022. 2, 4, 5
- [DW11] Anindya De and Thomas Watson. Extractors and lower bounds for locally samplable sources. In *RANDOM*, 2011. 2
- [DY21] Yevgeniy Dodis and Kevin Yeo. Doubly-affine extractors, and their applications. In *ITC*, 2021. 2
- [Fel07] Bálint Felszeghy. *Gröbner theory of zero dimensional ideals with a view toward combinatorics*. PhD thesis, Budapest University of Technology and Economics, 2007. 14
- [GVW15] Oded Goldreich, Emanuele Viola, and Avi Wigderson. On randomness extraction in AC^0 . In *CCC*, 2015. 2
- [Hei99] Petra Heijnen. *Some classes of linear codes: observations about their structure, construction, (non-)existence and decoding*. PhD thesis, Technische Universiteit Eindhoven, 1999. 4, 20

- [HIV22] Xuanguai Huang, Peter Ivanov, and Emanuele Viola. Affine extractors and AC^0 -parity. In *RANDOM*, 2022. [2](#)
- [Hoe63] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. [11](#)
- [HP98] Petra Heijnen and Ruud Pellikaan. Generalized Hamming weights of q -ary Reed-Muller codes. *IEEE Transactions on Information Theory (ToIT)*, 44(1):181–196, 1998. [4](#), [20](#)
- [Ind07] Piotr Indyk. Uncertainty principles, extractors, and explicit embeddings of L_2 into L_1 . In *STOC*, 2007. [2](#)
- [KLP12] Tali Kaufman, Shachar Lovett, and Ely Porat. Weight distribution and list-decoding size of Reed–Muller codes. *IEEE Transactions on Information Theory (ToIT)*, 58(5):2689–2696, 2012. [8](#)
- [KS05] Peter Keevash and Benny Sudakov. Set systems with restricted cross-intersections and the minimum rank of inclusion matrices. *SIAM Journal on Discrete Mathematics*, 18(4):713–727, 2005. [1](#), [4](#)
- [KS18] Swastik Kopparty and Srikanth Srinivasan. Certifying polynomials for $AC^0[\oplus]$ circuits, with applications to lower bounds and circuit compression. *Theory of Computing*, 14(12):1–24, 2018. [4](#)
- [Li23] Xin Li. Two source extractors for asymptotically optimal entropy, and (many) more. *arXiv preprint arXiv:2303.06802*, 2023. [2](#), [7](#)
- [LY22] Jiayu Li and Tianqi Yang. $3.1n - o(n)$ circuit lower bounds for explicit functions. In *STOC*, 2022. [2](#)
- [MR16] Shay Moran and Cyrus Rashtchian. Shattered sets and the Hilbert function. In *MFCS*, 2016. [1](#)
- [MU17] Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis*. Cambridge University Press, 2017. [11](#)
- [NW15] Zipei Nie and Anthony Y. Wang. Hilbert functions and the finite degree Zariski closure in finite field combinatorial geometry. *Journal of Combinatorial Theory, Series A*, 134:196–220, 2015. [2](#), [4](#), [21](#)
- [NZ96] Noam Nisan and David Zuckerman. Randomness is linear in space. *Journal of Computer and System Sciences*, 52(1):43–52, 1996. [2](#)
- [OSS19] Igor Carboni Oliveira, Rahul Santhanam, and Srikanth Srinivasan. Parity helps to compute majority. In *CCC*, 2019. [4](#)
- [Raz87] Alexander A. Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *Mathematical notes of the Academy of Sciences of the USSR*, 41(4):333–338, 1987. [1](#)

- [Rem16] Zachary Remscrim. The Hilbert function, algebraic extractors, and recursive Fourier sampling. In *FOCS*, 2016. 14
- [Smo87] Roman Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *STOC*, 1987. 1
- [Smo93] Roman Smolensky. On representations by low-degree polynomials. In *FOCS*, 1993. 3
- [Sri23] Srikanth Srinivasan. A robust version of Hegedűs’s lemma, with applications. *TheoretCS*, 2, 2023. 4
- [SV86] Miklos Santha and Umesh V. Vazirani. Generating quasi-random sequences from semi-random sources. *Journal of Computer and System Sciences (JCSS)*, 33(1):75–87, 1986. 2
- [TSZ01] Amnon Ta-Shma and David Zucherman. Extractor codes. In *STOC*, 2001. 2
- [TV00] Luca Trevisan and Salil Vadhan. Extracting randomness from samplable distributions. In *FOCS*, 2000. 2
- [Vad04] Salil P. Vadhan. Constructing locally computable extractors and cryptosystems in the bounded-storage model. *Journal of Cryptology*, 17:43–77, 2004. 2
- [Vio05] Emanuele Viola. The complexity of constructing pseudorandom generators from hard functions. *Computational Complexity*, 13(3-4):147–188, 2005. 2
- [Vio12] Emanuele Viola. The complexity of distributions. *SIAM Journal on Computing (SICOMP)*, 41(1):191–218, 2012. 2
- [Vio14] Emanuele Viola. Extractors for circuit sources. *SIAM Journal on Computing (SICOMP)*, 43(2):655–672, 2014. 2
- [Vio16] Emanuele Viola. Quadratic maps are hard to sample. *ACM Transactions on Computation Theory (TOCT)*, 8(4):1–4, 2016. 9
- [Wei91] Victor K. Wei. Generalized Hamming weights for linear codes. *IEEE Transactions on Information Theory (ToIT)*, 37(5):1412–1418, 1991. 4, 20
- [WZ99] Avi Wigderson and David Zuckerman. Expanders that beat the eigenvalue bound: Explicit construction and applications. *Combinatorica*, 19(1):125–138, 1999. 2
- [Zuc07] David Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. *Theory of Computing*, 3(6):103–128, 2007. 2