Two Structural Results for Low Degree Polynomials and Applications

Gil Cohen and Avishay Tal

Weizmann Institute of Science
Rehovot, Israel
{gil.cohen,avishay.tal}@weizmann.ac.il

Abstract

In this paper, two structural results concerning low degree polynomials over finite fields are given. The first states that over any finite field $\mathbb{F}$, for any polynomial f on n variables with degree $d \leq \log(n)/10$, there exists a subspace of $\mathbb{F}^n$ with dimension $\Omega(d \cdot n^{1/(d-1)})$ on which f is constant. This result is shown to be tight. Stated differently, a degree d polynomial cannot compute an affine disperser for dimension smaller than $\Omega(d \cdot n^{1/(d-1)})$. Using a recursive argument, we obtain our second structural result, showing that any degree d polynomial f induces a partition of $\mathbb{F}^n$ to affine subspaces of dimension $\Omega(n^{1/(d-1)!})$, such that f is constant on each part.

We extend both structural results to more than one polynomial. We further prove an analog of the first structural result to sparse polynomials (with no restriction on the degree) and to functions that are close to low degree polynomials. We also consider the algorithmic aspect of the two structural results.

Our structural results have various applications, two of which are:

- Dvir [11] introduced the notion of extractors for varieties, and gave explicit constructions of such extractors over large fields. We show that over any finite field any affine extractor is also an extractor for varieties with related parameters. Our reduction also holds for dispersers, and we conclude that Shaltiel's affine disperser [26] is a disperser for varieties over $\mathbb{F}_2$.
- Ben-Sasson and Kopparty [6] proved that any degree 3 affine disperser over a prime field is also an affine extractor with related parameters. Using our structural results, and based on the work of Kaufman and Lovett [19] and Haramaty and Shpilka [17], we generalize this result to any constant degree.

1998 ACM Subject Classification F.1.0 Computation by Abstract Devices – General

Keywords and phrases low degree polynomials, affine extractors, affine dispersers, extractors for varieties, dispersers for varieties

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2015.680

1 Introduction

In this paper, we consider the following question concerning polynomials on n variables over the field with q elements, $\mathbb{F}_q$, where q is some prime power:

> What is the largest number $k = k_q(n, d)$, such that any polynomial on n variables over $\mathbb{F}_q$, with degree[1] at most d, is constant on some affine subspace of $\mathbb{F}_q^n$ with dimension k?

[1] Here, and throughout the paper, by degree we mean total degree.

18th Int'l Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX'15) / 19th Int'l Workshop on Randomization and Computation (RANDOM'15).
Editors: Naveen Garg, Klaus Jansen, Anup Rao, and José D. P. Rolim; pp. 680–709

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

A related question was studied by Tardos and Barrington ([28], Lemma 3) who proved that for any prime power q and for any degree d polynomial f on n variables over the *ring* $\mathbb{Z}_q$, there exists a "cube" with dimension $k = \Omega(n^{1/d})$, on which f is constant. That is, there exist linearly independent vectors $\Delta_1, \ldots, \Delta_k \in \mathbb{Z}_q^n$ such that for every $\alpha \in \{0,1\}^k$, $f(\sum_{i=1}^k \alpha_i \Delta_i) = f(0)$. Although the problem studied in [28] is different than the problem mentioned above in several respects, one can make use of the proof idea of Tardos and Barrington and show that $k_2(n,d) = \Omega(n^{1/(d-1)})$ for all n, d (see Appendix B).

The proof idea of [28] seems to be applicable to our problem only for $q = 2$, and new ideas are required for larger fields. For any q, the case $d = 1$ is trivial – $k_q(n,1) = n-1$. The case $d = 2$, at least over fields of characteristic 2, is also well understood. By Dickson's theorem ([9], Theorem 199), $k_q(n,2) \geq \lfloor n/2 \rfloor$ for fields of characteristic 2. This is tight, as can be seen by considering the inner product function $x_1x_2 + x_3x_4 + \cdots + x_{n-1}x_n$.

1.1 Our Results

Our first result is an asymptotically tight upper and lower bounds on $k_q(n,d)$ for any q and $d < \log(n)/10$. The following theorem gives a lower bound for $k_q(n,d)$. In fact, it has a stronger guarantee which is required by one of our applications (see Theorem 6). Informally, for any degree d polynomial f and a point $u_0 \in \mathbb{F}_q^n$, there exists a large subspace U such that f is constant on $u_0 + U$. Note that this is equivalent to saying that there exists a large *linear* subspace on which f is constant (namely, the affine shift is by the zero vector).

▶ **Theorem 1** (Structural Result I). *For any n, d, let k be the least integer such that*

$$n \leq k + (d+1) \cdot \sum_{j=0}^{d-1} (d-j) \cdot \binom{k+j-1}{j}. \tag{1}$$

Let q be a prime power. Let $f\colon \mathbb{F}_q^n \to \mathbb{F}_q$ be a degree d polynomial, and let $u_0 \in \mathbb{F}_q^n$. Then, there exists a subspace $U \subseteq \mathbb{F}_q^n$ of dimension k such that $f|_{u_0+U}$ is constant.

In particular, there exists a universal constant $c_1 \in (0,1)$ such that for all n, d, q, it holds that $k_q(n,d) \geq c_1 \cdot n^{1/(d-1)}$. Moreover, for $d \leq \log(n)/10$ it holds that $k_q(n,d) = \Omega(d \cdot n^{1/(d-1)})$.

Few remarks are in order:

Tightness. Theorem 1 is tight for $d \leq \log(n)/10$. Indeed, one can show that, with probability at most $q^{-\binom{k}{d}}$,[2] a random degree d polynomial on n variables over $\mathbb{F}_q$ is constant on any fixed affine subspace of dimension k. There are at most $q^{(k+1)n}$ affine subspaces of dimension k, so by the union bound, $k_q(n,d)$ must be smaller than any k such that $\binom{k}{d} > (k+1)n$. Hence, $k_q(n,d) < d^{1+1/(d-1)} \cdot n^{1/(d-1)}$. For $d \leq \log(n)/10$, the ratio between our upper and lower bound is $d^{O(1/d)} = 1 + O(\log(d)/d)$.

Low degree polynomials, affine dispersers and affine extractors. An *affine disperser* for dimension k is a function $f\colon \mathbb{F}_q^n \to \mathbb{F}_q$ with the following property. For every affine subspace $u_0 + U \subseteq \mathbb{F}_q^n$ of dimension k, f restricted to $u_0 + U$ is not constant [3]. Thus, in the language of pseudorandomness, Theorem 1 states that a degree $d \leq \log(n)/10$

[2] The expression $\binom{k}{d}$ in the exponent can be replaced by the number of solutions to the equation $r_1 + \cdots + r_k \leq d$, where $r_i \in \{0, \ldots, q-1\}$.

[3] An alternative definition requires that almost all field elements are obtained by f on $u_0 + U$.

polynomial is not an affine disperser for dimension $o(d \cdot n^{1/(d-1)})$, and in particular, polynomials with constant degree are not affine dispersers for sub-polynomial dimension. For the special case $q = 2$, based on the work of Ben-Eliezer *et al.* [2], one can say something stronger regarding the tightness of Theorem 1. A function $f\colon \mathbb{F}_q^n \to \mathbb{F}_q$ is called an *affine extractor* for dimension k with bias ε, if for every affine subspace $u_0 + U \subseteq \mathbb{F}_q^n$ of dimension k, it holds that $f(x)$, where x is sampled uniformly from $u_0 + U$, is ε-close in statistical distance, to the uniform distribution over $\mathbb{F}_q$. By [2] it holds that for every $d \geq 1$, there exists a degree d affine extractor $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$ for any $k \geq \Omega(d \cdot n^{1/(d-1)})$, with $\varepsilon = 2^{-\Omega(k/d)}$(see Section 3.3).

The case of unbounded degree. Theorem 1 yields a non-trivial bound only for $d \leq O(\log n)$. When the degree of the polynomial is unbounded things behave differently. For example, it is considered a folklore that any function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is constant on some affine subspace with dimension $\Omega(\log n)$. Namely, $k_2(n, \infty) = \Omega(\log n)$ (this is, in fact, tight). On the other hand, Gabizon and Raz [12] noted that the polynomial $x_1^1 + x_2^2 + \cdots + x_n^n$ over the field with $n + 1$ elements is not constant on any dimension 1 affine subspace (see also [8]). Thus, $k_{n+1}(n, \infty) = 1$.

The independence of the field size. Note that the bound on $k_q(n, d)$ in Theorem 1 is independent of q. That is, when considering bounded degree polynomials, the field size does not affect $k_q(n, d)$. Throughout the paper we focus on low degree polynomials – polynomials of degree up to $\log(n)/10$. In this range of parameters, Theorem 1 and the fact that it is tight, allow us to suppress the field size and write $k(n, d)$ instead of $k_q(n, d)$, as we do from here on.

Partition of $\mathbb{F}^n$ to affine subspaces, induced by a low degree polynomial

Theorem 1 states that for any degree d polynomial f on n variables, there exists at least one large affine subspace, restricted to which, f is constant. However, for some of our applications we need a stronger structural result. More specifically, we ask what is the maximum number $\mathcal{K} = \mathcal{K}_q(n, d)$, such that any degree d polynomial on n variables over $\mathbb{F}_q$ induces a *partition* of $\mathbb{F}_q^n$ to dimension $\mathcal{K}$ affine subspaces, on each of which f is constant. Using Theorem 1, we show that $\mathcal{K}_q(n, d) = \Omega(n^{1/(d-1)!})$. That is, we obtain the following result.

▶ **Theorem 2** (Structural Result II). *There exists a universal constant $c_2 > 0$ such that the following holds. Let q be a prime power. Let $f\colon \mathbb{F}_q^n \to \mathbb{F}_q$ be a degree d polynomial. Then, there exists a partition of $\mathbb{F}_q^n$ to affine subspaces (not necessarily shifts of the same subspace), each of dimension $c_2 \cdot n^{1/(d-1)!}$, such that f is constant on each part.*

We do not know whether the lower bound in Theorem 2 for $\mathcal{K}_q(n, d)$ is tight or not for all d (note that it is tight for $d \leq 3$), and leave this as an open problem. More precisely, we ask what is the asymptotic behavior of $\mathcal{K}_q(n, d)$? Does it depend on q for, say, constant d?

Generalization of the structural results to many polynomials

Being a natural generalization and also necessary for some of our applications, we generalize the two structural results to the case of any number of polynomials (see Section 3.4). Let $f_1, \ldots, f_t\colon \mathbb{F}_q^n \to \mathbb{F}_q$ be polynomials of degree at most d. The generalization of the first structural result states that there exists an affine subspace of dimension $\Omega((n/t)^{1/(d-1)})$ on which *each* of the t polynomials is constant (see Theorem 19). By applying a probabilistic argument, one can show that the dependence in t is tight. For the second structural result, the guaranteed dimension in Theorem 2 is replaced by $\Omega(n^{1/(d-1)!}/t^e)$, where e is the base of the natural logarithm (see Theorem 20).

The algorithmic aspect

We further study the algorithmic aspect of the structural results (see Section 4). We devise a poly(n)-time deterministic algorithm (see Theorem 22), that given a degree d polynomial $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$ as a black-box, performs poly(n) queries, and outputs a subspace of dimension $\Omega(k(n,d))$, restricted to which, f has degree at most $d-1$. By applying this algorithm recursively d times, one can efficiently obtain a subspace of dimension $\Omega(n^{1/(d-1)!})$ on which f is constant. Our algorithm only works for the binary field. Devising an algorithm for general fields is a natural problem.

Note that there is a gap between $k(n,d)$ and the dimension of the affine subspace that our algorithm produce. A natural open problem is whether this gap can be eliminated. Specifically, we ask whether there is a poly(n)-time algorithm that, given a black-box access to a degree d polynomial $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$, finds an affine subspace with dimension $k(n,d)$ on which f is constant?

Whether there exists an algorithm as in the problem above is not at all clear to us. Verifying that a degree d polynomial is constant on a given affine subspace with dimension $k(n,d)$ can be done in time $O(k(n,d)^d) \leq O(n^2)$, and it might be the case that this problem is expressive enough to be **NP**-hard. We show that the latter scenario is unlikely, at least for constant d, by devising an $\exp(n^{1-\frac{1}{d-1}}) \cdot n^d$-time algorithm that outputs an affine subspace with dimension $\Omega(k(n,d))$ on which f is constant (see Theorem 24). We note that the naive algorithm iterates over all $\binom{2^n}{k(n,d)} = \exp(n^{1+\frac{1}{d-1}})$ affine subspaces with dimension $k(n,d)$. It is also worth mentioning that this algorithm works for all finite fields.

Sparse polynomials

We further give an analog of the first structural result to sparse polynomials (regardless of their degree) over any finite field. We have the following.

▶ **Theorem 3.** *Let q be a prime power. For any integer $c \geq 1$ the following holds. Let f be a polynomial on n variables over $\mathbb{F}_q$, with at most n^c monomials. Then, there exists an affine subspace of dimension $\Omega\left(n^{1/(4(q-1)c)}\right)$ on which f is constant.*

We note that unlike in the case of low degree polynomials, the field size q does affect the dimension of the affine subspace promised by Theorem 3. Some sort of dependency cannot be avoided. Indeed, as mentioned above, the polynomial $x_1^1 + x_2^2 + \cdots + x_n^n$ over the field with $n+1$ elements is not constant on any dimension 1 affine subspace, even though it has only n monomials. On the other hand, Theorem 3 gives no guarantee already for $q = \Omega(\log n)$, while the example above requires fields of size $\Omega(n)$. We leave open the problem of improving upon the dependence of Theorem 3 in the field size q, or proving that this dependence is optimal.

We note that for the special case $q = 2$, the lower bound in Theorem 3 is $\Omega\left(n^{1/(4c)}\right)$, which is essentially tight up to the constant 4 in the exponent, as implied by our tightness result for degree d polynomials. We do not know whether the constant 4 is necessary. Indeed, for degree d polynomials (which may have n^d monomials), the guarantee given by Theorem 1 is stronger, namely, $\Omega\left(n^{1/(d-1)}\right)$.

Functions that are close to low degree polynomials

Theorem 1 implies that any function that is close to a low degree polynomial, is constant on some large affine subspace.

▶ **Corollary 4.** *Let q be a prime power. Let $g : \mathbb{F}_q^n \to \mathbb{F}_q$ be a function that agrees with some degree d polynomial $f : \mathbb{F}_q^n \to \mathbb{F}_q$ on all points but for some subset $B \subseteq \mathbb{F}_q^n$. Then, there exists an affine subspace with dimension $\Omega((n - \log_q(|B|))^{1/(d-1)})$ on which g is constant.*

To see that, note that by averaging argument there is an affine subspace $w + W$ of dimension $n - \log_q(|B|) - 1$ on which f and g agrees. Applying Theorem 1 to $f|_{w+W}$ gives an affine subspace $u + U \subseteq w + W$ on which f, and thus g, is constant on. We suspect that better parameters can be achieved.

1.2 Applications

We now present several applications of our structural results.

Extractors and Dispersers for Varieties over all Finite Fields

Let $\mathbb{F}$ be some finite field. An affine subspace of $\mathbb{F}^n$ can be thought of as the set of common zeros of one or more degree 1 polynomials with coefficients in $\mathbb{F}$. Recall that an affine extractor over the field $\mathbb{F}$ is a function $f\colon \mathbb{F}^n \to \mathbb{F}$ that has small bias on every large enough affine subspace. In [11], the study of the following natural generalization was initiated: construct a function that has small bias on the set of common zeros of one or more degree $d > 1$ polynomials. In general, the set of common zeros of one or more polynomials is called a *variety*. For a set of polynomials $g_1, \ldots, g_t$ on n variables over $\mathbb{F}$, we denote their variety by $\mathbf{V}(g_1, \ldots, g_t) = \{x \in \mathbb{F}^n : g_1(x) = \cdots = g_t(x) = 0\}$. A function $f\colon \mathbb{F}^n \to \mathbb{F}$ as above is called an extractor for varieties.

In [11], two explicit constructions of extractors for varieties were given. For simplicity, we suppress here both the bias of the extractor and the number of output bits. Dvir's first construction works under no assumption on the variety size (more precisely, some assumption is made, but that assumption is necessary). The downside of this construction is that the underlining field is assumed to be quite large, more precisely, $|\mathbb{F}| > d^{\Omega(n^2)}$. The second construction works for fields with size as small as $\mathrm{poly}(d)$, however the construction is promised to work only for varieties with size at least $|\mathbb{F}|^{n/2}$. Dvir applies tools from algebraic geometry for his constructions.

Even the construction of affine extractors, which is a special case of extractors for varieties, is extremely challenging. Indeed, the (far from optimal) constructions known today use either very sophisticated exponential sum estimates [4, 33] or involved composition techniques [20], where the correctness relies, among other results, on deep structural results from additive combinatorics [30] and on XOR lemmas for low degree polynomials [32, 3]. The same can be said about the constructions of affine dispersers.

Given the difficulties in constructing affine extractors and dispersers, one may suspect that the construction of extractors and dispersers for varieties will be substantially more challenging, especially for small fields that seem to be immune against algebraic geometry based techniques. Nevertheless, based on our structural results, the following theorem states that any affine extractor is also an extractor for varieties with related parameters.

▶ **Theorem 5.** *Let q be a prime power. For any integers n, d, t the following holds. Let $f\colon \mathbb{F}_q^n \to \mathbb{F}_q$ be an affine extractor for dimension $\Omega(n^{1/(d-1)!}/t^e)$ with bias ε, where e is the base of the natural logarithm. Then, f is an extractor with bias ε for varieties that are the common zeros of any t polynomials, each of degree at most d.*

In fact, one can view Theorem 5 as an explanation for the difficulty of constructing affine extractors for dimension n^δ for constant $\delta < 1$.

We also obtain a reduction that does not depend on the number of polynomials defining the variety, but rather on the variety size (see Theorem 26). The proof idea in this case is to "approximate" the given variety by a variety induced by a small number of low degree polynomials, and then apply Theorem 5.

The state of the art explicit constructions of affine extractors for the extreme case $q = 2$, work only for dimension $\Omega(n/\sqrt{\log\log n})$ [4, 33, 20], and thus the reduction in Theorem 5 only gives an explicit construction of an extractor for varieties defined by quadratic polynomials (and in fact, up to $(\log\log n)^{1/(2e)}$ quadratic polynomials). However, a similar reduction to that in Theorem 5 also holds for dispersers.

▶ **Theorem 6.** *Let n, d, t be integers such that $d < \log(n/t)/10$. Let $f : \mathbb{F}_q^n \to \mathbb{F}_q$ be an affine disperser for dimension $\Omega(d \cdot (n/t)^{1/(d-1)})$. Then, f is a disperser for varieties that are the common zeros of any t polynomials of degree at most d.*

Over $\mathbb{F}_2$, an explicit construction of an affine disperser for dimension as small as $2^{\log^{0.9} n}$ is known [26]. Thus, we obtain the first disperser for varieties over $\mathbb{F}_2$.

▶ **Theorem 7.** *For any n, d, t such that $d < (1 - o_n(1)) \cdot \frac{\log(n/t)}{\log^{0.9} n}$, there exists an explicit construction of an affine disperser for varieties which are the common zeros of any t polynomials of degree at most d. In particular, when $t \leq n^\alpha$ for some constant $\alpha < 1$, the requirement on the degree is $d < (1 - \alpha - o_n(1)) \cdot \log^{0.1} n$.*

A few words regarding the limitation of the reduction in Theorem 6 are in order. Note that even if f is an optimal affine disperser, that is, a disperser for dimension $O(\log n)$, Theorem 6 only guarantees that f is a disperser for varieties defined by degree $O(\log n)$ polynomials. One cannot expect much more from the reduction. Indeed, there exists a degree $O(\log n)$ polynomial that computes an optimal affine disperser (this can be proven via a probabilistic argument. See also Theorem 36). However, this affine disperser is clearly not a disperser for varieties defined by even a single degree $O(\log n)$ polynomial.

Thus, the reduction in Theorem 6 is useful only for varieties defined by degree $o(\log n)$ polynomials. A recent work of Hrubeš and Rao [16] shows that it would be challenging to construct an explicit f which is an extractor (or even a disperser) for varieties of size $2^{\rho n}$ defined by degree n^ε polynomials over $\mathbb{F}_2$, for any constants $0 < \varepsilon, \rho < 1$. Indeed, such a function would solve Valiant's problem [29], since f cannot be computed by Boolean circuits of logarithmic depth and linear size.

From Affine Dispersers to Affine Extractors

Constructing an affine disperser is, by definition, an easier task than constructing an affine extractor. Nevertheless, Ben-Sasson and Kopparty [6] proved (among other results) that any degree 3 affine disperser is also an affine extractor with comparable parameters. [4] Using the extension of Theorem 1 to many polynomials, we are able to generalize the reduction of Ben-Sasson and Kopparty, over prime fields, to any degree $d \geq 3$.

▶ **Theorem 8.** *Let p be a prime number. For all $d \geq 3$ and $\delta > 0$, there exists $c = c(d, \delta)$ such that the following holds. Let $f : \mathbb{F}_p^n \to \mathbb{F}_p$ be an affine disperser for dimension k, which has degree d as a polynomial over $\mathbb{F}_p$. Then, f is also an affine extractor for dimension $k' \triangleq c \cdot k^{d-2}$ with bias δ.*

[4] A reduction from "low rank" extractors to dispersers in the context of two sources was also obtained, by Ben-Sasson and Zewi [7], conditioned on the well-known Polynomial Freiman-Ruzsa conjecture from additive combinatorics.

Note that Theorem 8 is only interesting in the case where $k^{d-2} < n$. However, this case is achievable since a random polynomial of degree d is an affine disperser for dimension $O(d \cdot n^{1/(d-1)})$. In particular, Theorem 8 implies that an explicit construction of an optimal affine disperser that has a constant degree as a polynomial, suffices to break the current natural barrier in the construction of affine extractors, namely, constructing affine extractors for dimension $n^{1-\delta}$ for some constant $\delta > 0$ (here $\delta = 1/(d-1)$).

On top of Theorem 1, the key ingredient we use in the proof of Theorem 8 is the work of Kaufman and Lovett [19], generalizing a result by Green and Tao [13] (see Section 6). For $d = 4$, we get a better dependency between k and k' based on the work of Haramaty and Shpilka [17] (see Theorem 28).

$\mathsf{AC}^0[\oplus]$ Circuits and Affine Extractors / Dispersers

Constructing affine dispersers, and especially affine extractors, is a challenging task. As mentioned, the state of the art explicit constructions for affine extractors over $\mathbb{F}_2$ work only for dimension $\Omega(n/\sqrt{\log\log n})$. By a probabilistic argument however, one can show the existence of affine extractors for dimension $(1 + o(1)) \log n$ (see Lemma 31). Thus, there is an exponential gap between the non-explicit construction and the explicit ones.

It is therefore tempting to try and utilize this situation and prove circuit lower bounds for affine extractors. This idea works smoothly for AC^0 circuits. Indeed, by applying the work of Håstad [14], one can easily show that an AC^0 circuit on n inputs cannot compute an affine disperser for dimension $o(n/\text{polylog}(n))$ (see Corollary 30). However, strong lower bounds for AC^0 circuits are known, even for much simpler and more explicit functions such as Parity and Majority. Thus, it is far more interesting to prove lower bounds against circuit families for which the known lower bounds are modest. One example would be to show that a De Morgan formula of size $O(n^3)$ cannot compute a good affine extractor, improving upon the best known lower bound [15]. [5]

Somewhat surprisingly, we show that even depth 3 $\mathsf{AC}^0[\oplus]$ circuit (that is, AC^0 circuits with XOR gates) can compute an optimal affine extractor over $\mathbb{F}_2$. In fact, the same construction can also be realized by a polynomial-size De Morgan formula and has degree $(1 + o(1)) \log n$ as polynomial over $\mathbb{F}_2$ (see Theorem 36).

Theorem 36 is implicit in the works of [22, 24] who studied a similar problem in the context of bipartite Ramsey graphs (that is, two-source dispersers). We give an alternative proof in Appendix A, which can be extended to work also in the context of bipartite Ramsey graphs.

Given that depth 3 $\mathsf{AC}^0[\oplus]$ circuits exhibit the surprising computational power mentioned above, it is natural to ask whether depth 2 $\mathsf{AC}^0[\oplus]$ circuit can compute a good affine extractor. We stress that even depth 2 $\mathsf{AC}^0[\oplus]$ circuits should not be disregarded easily! For example, such circuits *can* compute, in a somewhat different setting, optimal Ramsey graphs (see [18], Section 11.7). Moreover, any degree d polynomial $f : \mathbb{F}_2^n \to \mathbb{F}_2$ can be computed by a depth 2 $\mathsf{AC}^0[\oplus]$ circuit with size n^d. Nevertheless, we complement the above result by showing that a depth 2 $\mathsf{AC}^0[\oplus]$ circuit cannot compute an affine disperser for sub-polynomial dimension. The proof is based on the following reduction.

▶ **Lemma 9.** *Let C be a depth 2 $\mathsf{AC}^0[\oplus]$ circuit on n inputs, with size n^c. Let $k < n/10 - c\log(n)$. If C computes an affine disperser for dimension k, then there exists a degree $2c$ polynomial over $\mathbb{F}_2$ on $\sqrt{n}/5$ variables which is an affine disperser for dimension k.*

[5] The property of being an affine extractor meets the largeness condition of the natural proof barrier [23]. However, it does not necessarily get in the way of improving existing polynomial lower bounds.

The proof of Lemma 9 uses ideas from our proof of the structural result for sparse polynomials (see Lemma 21). Lemma 9 together with Theorem 1 imply the following theorem.

▶ **Theorem 10.** *Let C be a depth 2 $\mathsf{AC}^0[\oplus]$ circuit on n inputs, with size n^c, which is an affine disperser for dimension k. Then, $k > k(\sqrt{n}/5, 2c) = \Omega(n^{1/4c})$.*

Good Affine Extractors are Hard to Approximate by Low Degree Polynomials

Using our second structural result, Theorem 2, we obtain an average-case hardness result, or in other words, correlation bounds for low degree polynomials. Namely, we show that any affine extractor with very good parameters cannot be approximated by low degree polynomials over $\mathbb{F}_2$.

▶ **Corollary 11.** *Let $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$ be an affine extractor for dimension k with bias ε. Then, for any polynomial $g\colon \mathbb{F}_2^n \to \mathbb{F}_2$ of degree d such that $k = \Omega(n^{1/(d-1)!})$, it holds that*

$$\mathrm{Cor}(f,g) \triangleq \mathop{\mathbf{E}}_{x\sim\mathbb{F}_2^n}\left[(-1)^{f(x)}\cdot(-1)^{g(x)}\right] \le \varepsilon.$$

Proof. Let g be a degree d polynomial over $\mathbb{F}_2$ on n variables. By Theorem 2, there exists a partition of $\mathbb{F}_2^n$ to affine subspaces $P_1, P_2, \ldots, P_\ell$, each of dimension $k = \Omega(n^{1/(d-1)!})$, such that for all $i \in [\ell]$, $g|_{P_i}$ is some constant $g(P_i)$. Thus,

$$\begin{aligned}\mathrm{Cor}(f,g) &= \left|\mathop{\mathbf{E}}_{x\sim\mathbb{F}_2^n}[(-1)^{f(x)+g(x)}]\right| = \left|\mathop{\mathbf{E}}_{i\sim[\ell]}\mathop{\mathbf{E}}_{x\sim P_i}[(-1)^{f(x)+g(P_i)}]\right| \\ &\le \mathop{\mathbf{E}}_{i\sim[\ell]}\left|(-1)^{g(P_i)}\cdot\mathop{\mathbf{E}}_{x\sim P_i}[(-1)^{f(x)}]\right| ,\end{aligned}$$

which is at most ε since f is an affine extractor for dimension k with bias ε. ◀

As mentioned, explicit constructions of affine extractors for dimension $\Omega(n/\sqrt{\log\log n})$ are known. Corollary 11 implies that these extractors cannot be approximated by quadratic polynomials. Corollary 11 also implies that for any constant $\beta \in (0,1)$, affine extractors for dimension $k \le 2^{(\log n)^\beta}$ with bias ε have correlation ε with degree $d \le O_\beta(\log\log n/\log\log\log n)$ polynomials.[6] Unfortunately, an explicit construction for extractors with such parameters has not yet been achieved.

We also note that stronger correlation bounds are known in the literature for explicit (and simple) functions (see [31] and references therein). Nevertheless, we find the fact that *any* affine extractor has small correlation with low degree polynomials interesting.

The Granularity of the Fourier Spectrum of Low-Degree Polynomials over $\mathbb{F}_2$

The bias of an arbitrary function $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$ is clearly some integer multiplication of 2^{-n}. Theorem 2 readily implies that the bias of a degree d polynomial on n variables has a somewhat larger granularity – the bias is a multiplication of $2^{\Omega(n^{1/(d-1)!})}/2^n$ by some integer.[7]

[6] This is the best d we can guarantee for any k, and we gain nothing more by taking $k = O(\log n)$.

[7] Throughout the paper, for readability, we supress flooring and ceiling. In the last expression, however, it should be noted that we mean 2^{k-n}, where k is some integer such that $k = \Omega(n^{1/(d-1)!})$.

In fact, Theorem 2 implies that *all* Fourier coefficients of a low degree polynomial has this granularity. To see this, apply Theorem 2 to obtain a partition $P_1, \ldots, P_\ell$ of $\mathbb{F}_2^n$ to affine subspaces of dimension $k = \Omega(n^{1/(d-1)!})$, such that for each $i \in [\ell]$, $f|_{P_i}$ is some constant $f(P_i)$. Let $\beta \in \mathbb{F}_2^n$. Then,

$$\begin{aligned} 2^n \cdot \widehat{f}(\beta) &= \sum_{x \in \mathbb{F}_2^n} (-1)^{\langle \beta, x \rangle} \cdot (-1)^{f(x)} = \sum_{i=1}^{\ell} \sum_{x \in P_i} (-1)^{\langle \beta, x \rangle} \cdot (-1)^{f(x)} \\ &= \sum_{i=1}^{\ell} (-1)^{f(P_i)} \cdot \sum_{x \in P_i} (-1)^{\langle \beta, x \rangle}. \end{aligned}$$

The proof then follows as for all $i \in [\ell]$, the inner sum $\sum_{x \in P_i} (-1)^{\langle \beta, x \rangle}$ is either 0 or $\pm 2^k$.

1.3 Proof Overview

In this section we give proof sketches for some of our structural results. We start with Theorem 1, and consider first the special case $q = 2$. As mentioned, the proof for this special case follows the proof idea of [28]. We then consider general finite fields and present the new ideas required for this case.

We are given a point $u_0 \in \mathbb{F}_2^n$ and assume, without loss of generality, that $f(u_0) = 0$. We iteratively construct affine subspaces, restricted to which, f is zero. We start with affine subspaces of dimension 0, which are just the singletons $\{x\}$, where $x \in \mathbb{F}_2^n$ is such that $f(x) = 0$. Assume that we were able to find basis vectors $\Delta_1, \ldots, \Delta_k$ for a subspace U such that f restricted $u_0 + U$ is constantly 0. Consider all cosets $x + U$, restricted to which f is constantly 0. We call such cosets *good*. Clearly the coset $u_0 + U$ is good. If at least one more good coset $x + U$ exists, then we can pick a new direction Δ_{k+1} to be $x + u_0$, and get that f is zero on $u_0 + \text{span}\{\Delta_1, \ldots, \Delta_{k+1}\}$, as indeed

$$\begin{aligned} u_0 + \text{span}\{\Delta_1, \ldots, \Delta_{k+1}\} &= (u_0 + \text{span}\{\Delta_1, \ldots, \Delta_k\}) \cup (u_0 + \Delta_{k+1} + \text{span}\{\Delta_1, \ldots, \Delta_k\}) \\ &= (u_0 + \text{span}\{\Delta_1, \ldots, \Delta_k\}) \cup (x + \text{span}\{\Delta_1, \ldots, \Delta_k\}). \end{aligned}$$

The main observation used to derive Theorem 1 is the following. Given $\Delta_1, \ldots, \Delta_k$, there exists a degree $D \leq d^2 \cdot k^{d-1}$ polynomial $t : \mathbb{F}_2^n \to \mathbb{F}_2$, such that $x + U$ is a good coset if and only if $t(x) = 1$. Since we know that t is not the constant 0 function (as $t(u_0) = 1$), the DeMillo-Lipton-Schwartz-Zippel lemma (see Lemma 13) implies that there are at least 2^{n-D} x's such that $t(x) = 1$, namely, 2^{n-D} good cosets. So in each iteration, by our choice of Δ_{k+1}, we ensure that one coset in the next iteration is good, and then use DeMillo-Lipton-Schwartz-Zippel to claim that many other cosets are good as well. One can continue expanding the subspace U until $n \leq D$, which completes the proof.

For a general finite field, $\mathbb{F}_q$, we similarly define a polynomial $t(x)$ over $\mathbb{F}_q$ that attains only the values 0 and 1, and whose 1's capture the good cosets. The polynomial $t(x)$ is of degree at most $(q-1) \cdot d^2 \cdot k^{d-1}$. We wish to find a new direction Δ_{k+1}, linearly independent of $\Delta_1, \ldots, \Delta_k$, such that all cosets along the line $\{u_0 + \Delta_{k+1} \cdot a\}_{a \in \mathbb{F}_q}$, i.e. $\{u_0 + \Delta_{k+1} \cdot a + U\}_{a \in \mathbb{F}_q}$, are good. Over $\mathbb{F}_2$ this task was easy since $u_0 + U$ and $x + U$ define such a line.

The main new idea needed over $\mathbb{F}_q$ is to consider a polynomial

$$s(y) = \prod_{a \in \mathbb{F}_q} t(u_0 + y \cdot a),$$

whose variable represents a direction in $\mathbb{F}_q^n$ rather than a point. Note that $s(y)$ has degree at most $q \cdot \deg(t)$ and that $s(y) = 1$ if and only if $t(u_0 + y \cdot a) = 1$ for all $a \in \mathbb{F}_q$. Thus, $s(y) = 1$ iff f is zero on all cosets $\{u_0 + y \cdot a + U\}_{a \in \mathbb{F}_q}$, whose union is a dimension $k+1$ affine subspace as long as $y \notin U$. As before, since $s(0) = 1$, by a generalized DeMillo-Lipton-Schwartz-Zippel lemma, it holds that $s(\cdot)$ has many 1's, and as long as $k \ll n^{1/(d-1)}$ there is some $y \in s^{-1}(1)$ such that $y \notin U$. We can now pick such a y as Δ_{k+1}. A slightly more careful argument shows that actually there is no dependency of the dimension k in the field size q.

The proof of the second structural result (Theorem 2) can be described informally as follows. Consider a degree d polynomial f. Theorem 1 implies the existence of an affine subspace $u_0 + U$ with dimension $\Omega(n^{1/(d-1)})$ on which f is constant. One can then show (see Lemma 16) that restricting f to any affine shift of U yields a degree (at most) $d-1$ polynomial. Thus, one can partition each such affine subspace recursively to obtain a partition of $\mathbb{F}_q^n$ to affine subspaces (not necessarily shifts of one another), such that f is constant on each one of them.

In fact, to prove Theorem 2, one is not required to find an affine subspace on which f is constant, and it suffices to find an affine subspace on which the degree of f decreases. In order to obtain the first algorithmic result (Theorem 22), we devise an algorithm that finds such an affine subspace and proceed similarly to the proof of Theorem 2. To obtain the second algorithmic result (Theorem 24), we observe that the polynomial t described above has many linear factors. This structure of t allows us to save on the running time.

The generalization of Theorems 1 and 2 to more than one polynomial is quite straightforward.

2 Preliminaries

We shall denote prime numbers with the letter p and prime powers with q. The set $\{1, \ldots, n\}$ is denoted by $[n]$. We denote by $\log(\cdot)$ the logarithm to the base 2. Throughout the paper, for readability sake, we suppress flooring and ceiling. For $x, y \in \mathbb{F}_q^n$ we denote by $\langle x, y \rangle$ their scalar product over $\mathbb{F}_q$, i.e., $\langle x, y \rangle = \sum_{i=1}^n x_i \cdot y_i$. The vector e_i is the unit vector defined as having 1 in the i^{th} entry and 0 elsewhere. For a set $T \subseteq [n]$, we denote by $\mathbf{1}_T$ the indicating vector of T with 1 in the i^{th} entry if $i \in T$ and 0 otherwise. For a vector $\alpha \in \mathbb{N}^m$, we denote its *weight* by $\mathrm{wt}(\alpha) \triangleq \sum_i \alpha_i$.

The statistical distance between two random variables X, Y, over the same domain D, denoted by $\mathrm{SD}(X, Y)$, is defined as $\mathrm{SD}(X, Y) = \max_{A \subseteq D} |\Pr[X \in A] - \Pr[Y \in A]|$. It is known that $\mathrm{SD}(X, Y)$ is a metric. More precisely, it is (up to a multiplicative constant factor of 2) the ℓ_1 norm of the vector $(\Pr[d \in X] - \Pr[d \in Y])_{d \in D} \in \mathbb{R}^{|D|}$. In particular, we have the triangle inequality: for X, Y, Z over D, $\mathrm{SD}(X, Z) \leq \mathrm{SD}(X, Y) + \mathrm{SD}(Y, Z)$. Moreover, if X can be written as a convex combination of two random variables Y, Z as follows $X = (1 - \gamma) \cdot Y + \gamma \cdot Z$, where $\gamma \in [0, 1]$, then $\mathrm{SD}(X, Y) \leq \gamma$. We sometimes abuse notation, and for a set $S \subseteq D$, consider S also as the random variable that is uniformly distributed over the set S.

Restriction to an affine subspace

Let $f \colon \mathbb{F}_q^n \to \mathbb{F}_q$ be a function, $U \subseteq \mathbb{F}_q^n$ a subspace of dimension k and $u_0 \in \mathbb{F}_q^n$ some vector. We denote by $f|_{u_0+U} : (u_0 + U) \to \mathbb{F}_q$ the restriction of f to $u_0 + U$. The degree of $f|_{u_0+U}$ is defined as the minimal degree of a polynomial (from $\mathbb{F}_q^n$ to $\mathbb{F}_q$) that agrees with f on $u_0 + U$. For recursive arguments, it will be very useful to fix some basis $u_1, \ldots, u_k$ for U and

to consider the function $g : \mathbb{F}_q^k \to \mathbb{F}_q$ defined by

$$g(x_1, \ldots, x_k) = f\left(u_0 + \sum_{i=1}^{k} x_i \cdot u_i\right).$$

Note that the $\deg(g) = \deg(f|_{u_0+U})$ regardless of the choice for the basis.

Polynomials

We review some definitions and known facts about polynomials that we use.

The degree of a function $f : \mathbb{F}_q^n \to \mathbb{F}_q$, denoted by $\deg(f)$, is the degree of the unique multivariate polynomial over $\mathbb{F}_q$, where each individual degree is at most $q-1$, which agrees with f on $\mathbb{F}_q^n$. In the special case $q = 2$, such polynomials are called multi-linear. We will abuse notation and interchange between a function and its unique polynomial over $\mathbb{F}_q$ that agrees with f on $\mathbb{F}_q^n$.

▶ **Definition 12.** Let $f\colon \mathbb{F}_q^n \to \mathbb{F}_q$ be a polynomial of degree d, and let $\Delta \in \mathbb{F}_q^n$. The polynomial $\frac{\partial f}{\partial \Delta}(x) \triangleq f(x+\Delta) - f(x)$ is called the *derivative of f in direction Δ*.

It is easy to verify that $\deg\left(\frac{\partial f}{\partial \Delta}\right) \leq \deg(f) - 1$. Let $\Delta_1, \ldots, \Delta_k \in \mathbb{F}_q^n$ then

$$\frac{\partial^k f}{\partial \Delta_1 \ldots \partial \Delta_k}(x) = \sum_{S \subseteq [k]} (-1)^{1+|S|} \cdot f\left(x + \sum_{i \in S} \Delta_i\right)$$

is a degree $\leq \deg(f) - k$ polynomial.

The following lemma is a variant of the well-known DeMillo-Lipton-Schwartz-Zippel lemma [10, 25, 34].

▶ **Lemma 13** (DeMillo-Lipton-Schwartz-Zippel)**.** *Let q be a prime power. Let $f \in \mathbb{F}_q[x_1, \ldots, x_n]$ be a degree d non-zero polynomial. Then,* $\Pr_{x \sim \mathbb{F}_q^n}[f(x_1, \ldots, x_n) \neq 0] \geq q^{-d/(q-1)}$.

For completeness, we give the proof of Lemma 13 in Appendix C. The following folklore fact about polynomials over $\mathbb{F}_2$ is easy to verify.

▶ **Lemma 14** (Möbius inversion formula)**.** *Let $f(x_1, \ldots, x_n) = \sum_{S \subseteq [n]} a_S \cdot \prod_{i \in S} x_i$ be a polynomial over $\mathbb{F}_2$. Then, its coefficients are given by the formula: $a_S = \sum_{T \subseteq S} f(\mathbf{1}_T)$.*

Circuits

A Boolean circuit is an unbounded fan-in circuit composed of OR and AND gates, and literals x_i, $\neg x_i$. The size of such a circuit is the number of gates in it. A Boolean formula is a Boolean circuit such that every OR and AND gate has fan-out 1. De Morgan formula is a Boolean formula where each gate has fan-in at most 2. We recall that an AC^0 circuit is a Boolean circuit of polynomial size and constant depth. An $\mathsf{AC}^0[\oplus]$ circuit is an AC^0 circuit with unbounded fan-in XOR gates as well.

3 Structural Results

This section contains the proofs of all the structural results in this paper. In Section 3.1 we give a proof for Theorem 1. Section 3.2 contains the proof for Theorem 2. The tightness of the first structural result is given in Section 3.3. In Section 3.4 we describe the generalization of the two structural results to many polynomials. In Section 3.5 we prove Theorem 3.

3.1 Proof of Theorem 1

In this section we prove Theorem 1. For a slightly simpler proof, for the special case $q = 2$, we refer the reader to Appendix B. The proof of Theorem 1 is based on the following lemma.

▶ **Lemma 15.** *Let $f : \mathbb{F}_q^n \to \mathbb{F}_q$ be some function, and let U be a subspace of $\mathbb{F}_q^n$ with basis vectors $\Delta_1, \ldots, \Delta_k$. Then, there exist polynomials $(f_\alpha)_{\alpha \in \{0,1,\ldots,q-1\}^k}$ such that*
1. *$\deg(f_\alpha) \leq \deg(f) - \mathrm{wt}(\alpha)$ for all $\alpha \in \{0, 1, \ldots, q-1\}^k$.*
2. *Let $x \in \mathbb{F}_q^n$, then $f|_{x+U} \equiv 0$ if and only if $f_\alpha(x) = 0$ for all $\alpha \in \{0, 1, \ldots, q-1\}^k$.*

Proof. Complete $\Delta_1, \ldots, \Delta_k$ into a basis of $\mathbb{F}_q^n$ by picking vectors $\Delta_{k+1}, \ldots, \Delta_n \in \mathbb{F}_q^n$. Let A be the linear transformation which maps the standard basis into $\Delta_1, \ldots, \Delta_n$, and let $g(y) := f(Ay)$ (alternatively, $f(x) = g(A^{-1}x)$). Write g as a polynomial over $\mathbb{F}_q$:

$$g(y) = \sum_{\gamma \in \{0,1,\ldots,q-1\}^n} c_\gamma \cdot \prod_{i=1}^{n} y_i^{\gamma_i} .$$

Since both f and g can be obtained from one another by applying a linear transformation to the inputs, we have $\deg(f) = \deg(g)$. Think of the input to g as a concatenation of two parts $y = z \circ w$, where $z \in \mathbb{F}_q^k$, $w \in \mathbb{F}_q^{n-k}$. Let $P_z : \mathbb{F}_q^n \to \mathbb{F}_q^k$ be the projection of a vector of length n to the first k coordinates and let $P_w : \mathbb{F}_q^n \to \mathbb{F}_q^{n-k}$ be the projection to the last $n - k$ coordinates. We may rewrite g as

$$g(z \circ w) = \sum_{\alpha \in \{0,1,\ldots,q-1\}^k} \sum_{\beta \in \{0,1,\ldots,q-1\}^{n-k}} c_{\alpha \circ \beta} \cdot \prod_{i=1}^{k} z_i^{\alpha_i} \cdot \prod_{i=1}^{n-k} w_i^{\beta_i} .$$

By reordering the summations we get

$$g(z \circ w) = \sum_{\alpha \in \{0,1,\ldots,q-1\}^k} g_\alpha(w) \cdot \prod_{i=1}^{k} z_i^{\alpha_i} ,$$

where

$$g_\alpha(w) = \sum_{\beta \in \{0,1,\ldots,q-1\}^{n-k}} c_{\alpha \circ \beta} \cdot \prod_{i=1}^{n-k} w_i^{\beta_i} .$$

Note that $\deg(g_\alpha) \leq \deg(g) - \mathrm{wt}(\alpha)$. We have

$$\begin{aligned} f|_{x+U} \equiv 0 \quad &\iff \quad g|_{A^{-1}x + A^{-1}U} \equiv 0 \\ &\iff \quad g|_{A^{-1}x + \mathrm{span}\{e_1,\ldots,e_k\}} \equiv 0 \ (*) . \end{aligned}$$

Writing $(z, w) = (P_z(A^{-1}x), P_w(A^{-1}x))$ gives

$$\begin{aligned} (*) \quad &\iff \quad \forall z' \in \mathbb{F}_q^k : g(z' \circ w) = 0 \\ &\iff \quad \forall \alpha : g_\alpha(w) = 0 \\ &\iff \quad \forall \alpha : g_\alpha(P_w(A^{-1}x)) = 0 . \end{aligned}$$

Taking f_α to be the composition $g_\alpha \circ P_w \circ A^{-1}$ we obtain Item 2. As $P_w \circ A^{-1}$ is simply a linear transformation, it is clear that $\deg(f_\alpha) \leq \deg(g_\alpha) \leq \deg(g) - \mathrm{wt}(\alpha) \leq \deg(f) - \mathrm{wt}(\alpha)$, which completes the proof. ◀

Proof of Theorem 1. Assume without loss of generality that $f(u_0) = 0$, as otherwise we can look at the polynomial $g(x) = f(x) - f(u_0)$ which is of the same degree. The proof is by induction. Let k be such that

$$n > k + (d+1) \cdot \sum_{j=0}^{d-1} (d-j) \cdot \binom{k+j-1}{j} . \tag{2}$$

We assume by induction that there exists an affine subspace $u_0 + \text{span}\{\Delta_1, \ldots, \Delta_k\} \subseteq \mathbb{F}_q^n$, where the Δ_i's are linearly independent vectors, on which f evaluates to 0. Assuming Equation 2 holds, we show there exists a vector Δ_{k+1}, linearly independent of $\Delta_1, \ldots, \Delta_k$, such that $f \equiv 0$ on $u_0 + \text{span}\{\Delta_1, \ldots, \Delta_{k+1}\}$. To this aim, consider the set

$$A = \left\{ x \in \mathbb{F}_q^n \;\middle|\; f|_{x+\text{span}\{\Delta_1,\ldots,\Delta_k\}} \equiv 0 \right\} .$$

By the induction hypothesis, $u_0 \in A$. By Lemma 15, for any $x \in \mathbb{F}_q^n$,

$$f|_{x+\text{span}\{\Delta_1,\ldots,\Delta_k\}} \equiv 0 \quad \iff \quad \forall \alpha \in \{0,1,\ldots,q-1\}^k : f_\alpha(x) = 0 ,$$

where f_α is of degree at most $d - \text{wt}(\alpha)$. Thus $f_\alpha \equiv 0$ for $\text{wt}(\alpha) > d$, and we may write A as

$$A = \left\{ x \in \mathbb{F}_q^n \;\middle|\; \forall \alpha : \text{wt}(\alpha) \leq d,\ f_\alpha(x) = 0 \right\} .$$

Hence, A is the set of solutions to a system of $\leq \binom{k+d}{d}$ polynomial equations, where there are at most $\binom{k+j-1}{j}$ equations which correspond to α's of weight j and thus to degree (at most) $d - j$ polynomials. One can also write A as the set of non-zeros to the single polynomial

$$t(x) := \prod_{\alpha : \text{wt}(\alpha) \leq d} (1 - f_\alpha(x)^{q-1}) ,$$

which is of degree

$$\deg(t) \leq (q-1) \cdot \sum_{j=0}^{d-1} (d-j) \cdot \binom{k+j-1}{j} .$$

Note that $t(x)$ obtains only the values 0 and 1. Let $R \subseteq \mathbb{F}_q$ be an arbitrary subset of $\mathbb{F}_q$ with size $|R| = \min(q, d+1)$. Define a polynomial $s(y) := \prod_{r \in R} t(u_0 + r \cdot y)$. We claim that any non-zero of s not in the span of $\{\Delta_1, \ldots, \Delta_k\}$ can be taken to be the desired Δ_{k+1}. Indeed, if y is such that $s(y) = 1$, then $t(u_0 + r \cdot y) = 1$ for all $r \in R$. That is, for every $z \in \text{span}(\Delta_1, \ldots, \Delta_k)$ and any $r \in R$ it follows that $f(u_0 + z + r \cdot y) = 0$. Namely, f obtains $|R|$ roots on the affine line with offset $u_0 + z$ and direction y. If $R = \mathbb{F}_q$ then clearly this implies that f is the zero function restricted to the line. Otherwise, $|R| = d + 1$ and thus f, which is a degree d polynomial, obtains $d + 1$ zeros on the line. Thus, again f is the zero function on this line. Hence, $f(u_0 + z + r \cdot y) = 0$ for all $r \in \mathbb{F}_q$.

Thus, we just have to show that there exists some non-zero of s which is linearly independent of $\{\Delta_1, \ldots, \Delta_k\}$. Since the trivial solution $y = 0$ is a non-zero of s, we get that s is not the constant 0 function. Thus, by Lemma 13 it holds that $\Pr[s(y) \neq 0] \geq q^{-\deg(s)/(q-1)}$. The above equation implies that s has at least $q^{n - \deg(s)/(q-1)}$ ones. Since we need to avoid q^k linear combinations of the previous $\Delta_1, \ldots, \Delta_k$, it is enough to have

$$n - \frac{\deg(s)}{q-1} > k . \tag{3}$$

Since

$$\deg(s) \leq (d+1)\cdot(q-1)\cdot\sum_{j=0}^{d-1}(d-j)\cdot\binom{k+j-1}{j}$$

and by the assumption on k in Equation (2) we have that Equation (3) holds. ◀

3.2 Proof of Theorem 2

In this section we prove Theorem 2. To this end we use the following lemma.

▶ **Lemma 16.** *Let q be a prime power. Let $f\colon \mathbb{F}_q^n \to \mathbb{F}_q$ be a degree d polynomial. Assume there exists an affine subspace u_0+U, restricted to which f has degree at most $d-1$. Then, the degree of f restricted to any affine shift of U is at most $d-1$.*

Proof. Fix $u_1 \in \mathbb{F}_q^n$. Now, for any $u \in U$

$$f(u_1+u) = f(u_1+u) - f(u_0+u) + f(u_0+u) = \frac{\partial f}{\partial(u_1-u_0)}(u_0+u) + f(u_0+u).$$

Since the degree of the partial derivative of f is at most $d-1$ and the degree of $f|_{u_0+U}$ is also at most $d-1$, we get that $f|_{u_1+U}$ has degree at most $d-1$. ◀

Proof of Theorem 2. Let $c_1 \in (0,1)$ be the constant from Theorem 1. Define the sequence $\{\beta_d\}_{d=1}^{\infty}$ as follows.

$$\beta_d = \begin{cases} 1/2, & d=1; \\ \beta_{d-1}\cdot c_1^{\frac{1}{(d-2)!}}, & d>1. \end{cases}$$

We will prove by induction on d, the degree of a given polynomial f, that there exists a partition of $\mathbb{F}_q^n$ to affine subspaces of dimension $\geq \beta_d \cdot n^{1/(d-1)!}$, such that f restricted to each part is constant. The proof then follows by noting that for all $d \geq 1$,

$$\beta_d = \frac{1}{2}\cdot c_1^{\frac{1}{(d-2)!}+\cdots+\frac{1}{1!}+\frac{1}{0!}} \geq \frac{c_1^e}{2},$$

and thus one can take $c_2 = c_1^e/2$ to be the constant in the theorem statement.

The base case of the induction, namely $d=1$, trivially follows as f is an affine function, and we can partition $\mathbb{F}_q^n$ to q affine subspaces of dimension $n-1 \geq n/2 = \beta_1 n$, such that on each of which f is constant. Assume now that f is a degree $d>1$ polynomial. By Theorem 1 and Lemma 16, there exists a partition of $\mathbb{F}_q^n$ to affine subspaces of dimension $k \geq c_1 \cdot n^{1/(d-1)}$, such that f restricted to any affine subspace in the partition has degree at most $d-1$. Fix some affine subspace u_0+U in this partition, and apply the induction hypothesis to the polynomial $f' = f|_{u_0+U}$, which has degree $d' \leq d-1$. [8] By the induction hypothesis, we obtain a partition of u_0+U such that f is constant on each part. Moreover, the dimension of each such part is at least

$$\beta_{d'}\cdot k^{\frac{1}{(d'-1)!}} \geq \beta_{d-1}\cdot k^{\frac{1}{(d-2)!}} \geq \beta_{d-1}\cdot\left(c_1\cdot n^{\frac{1}{d-1}}\right)^{\frac{1}{(d-2)!}} = \beta_{d-1}\cdot c_1^{\frac{1}{(d-2)!}}\cdot n^{\frac{1}{(d-1)!}} = \beta_d\cdot n^{\frac{1}{(d-1)!}},$$

where the first inequality follows since $\{\beta_d\}_{d=1}^{\infty}$ is monotonically decreasing and $d' \leq d-1$, and the last equality follows by the definitions of the β_d's. ◀

[8] We may apply the induction because there exists a linear bijection from U to $\mathbb{F}_q^{\dim U}$. More precisely, if A is an $n\times k$ matrix over $\mathbb{F}_q$ that maps U to $\mathbb{F}_q^k$ bijectively, then one can apply the induction to the polynomial $f''(x) = f'(u_0+Ax)$, defined on k variables, and then induce a partition of u_0+U from the partition of $\mathbb{F}_q^k$ obtained by the induction. The induction can be carried on f'' since $\deg f'' \leq \deg f' \leq d-1$, where the first inequality holds because the variables of f'' are linear combinations of the variables of f'.

3.3 On the Tightness of Structural Result I

Roughly speaking, Theorem 1 states that for any prime power q, a degree d polynomial over $\mathbb{F}_q$ in n variables is not an affine disperser for dimension $k = \Omega(n^{1/(d-1)})$. We mentioned that this result is tight in the sense that by increasing k a bit, there exists a degree d polynomial which is an affine disperser. In this section we show, that in the special case $q = 2$, a stronger claim can be proven. Namely, by increasing k a bit, there exists a degree d polynomial which is an affine extractor.

▶ **Theorem 17.** *There exists a constant c such that the following holds. Let n, d be such that $d < n/2$. There exists a degree d polynomial $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$, such that for every affine subspace $u_0 + U \subseteq \mathbb{F}_2^n$ of dimension $k \geq cd \cdot n^{1/(d-1)}$, $\mathrm{bias}(f|_{u_0+U}) \leq 2^{-\Omega(k/d)}$.*

To prove Theorem 17 we apply the following lemma due to Ben-Eliezer, Hod and Lovett [2].

▶ **Lemma 18** ([2], Lemma 2). *Fix $\varepsilon > 0$ and let $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$ be a random degree d polynomial*[9] *for $d \leq (1-\varepsilon)n$. Then,*

$$\Pr_f\left[\mathrm{bias}(f) > 2^{-c_1 n/d}\right] \leq 2^{-c_2\binom{n}{\leq d}},$$

where $0 < c_1, c_2 < 1$ are constants depending only on ε.

Proof of Theorem 17. Let $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$ be a random polynomial of degree at most d. Fix an affine subspace $u_0 + U \subseteq \mathbb{F}_2^n$ of dimension k. One can easily show that $f|_{u_0+U}$ is equidistributed as a random polynomial on k variables, of degree at most d. Therefore, by Lemma 18,

$$\Pr_f\left[\mathrm{bias}(f|_{u_0+U}) > 2^{-c_1 k/d}\right] \leq 2^{-c_2\binom{k}{\leq d}},$$

where c_1, c_2 are the constants from Lemma 18 suitable for the (somewhat arbitrary) choice $\varepsilon = 1/2$. By taking the union bound over all $\leq 2^n \cdot \binom{2^n}{k}$ affine subspaces of $\mathbb{F}_2^n$ of dimension k, it is enough to require that

$$2^{-c_2\binom{k}{\leq d}} \cdot 2^n \cdot \binom{2^n}{k} < 1$$

so to conclude the proof of the theorem. It is easy to verify that one can choose c, as a function of c_2, such that the above equation does hold for k as defined in the theorem statement. ◀

3.4 Generalization of the Structural Results to Many Polynomials

▶ **Theorem 19** (Structural Result I for many polynomials). *Let q be a prime power. Let $f_1, \ldots, f_t : \mathbb{F}_q^n \to \mathbb{F}_q$ be polynomials of degree $d_1, \ldots, d_t$ respectively. Let k be the least integer satisfying the inequality*

$$n \leq k + \sum_{i=1}^{t} (d_i + 1) \cdot \sum_{j=0}^{d_i - 1} (d_i - j) \cdot \binom{k+j-1}{j}.$$

Then, for every $u_0 \in \mathbb{F}_q^n$ there exists a subspace $U \subseteq \mathbb{F}_q^n$ of dimension k, such that for all $i \in [t]$, f_i restricted to $u_0 + U$ is a constant function. In particular, if $d_1, \ldots, d_t \leq d$ then $k = \Omega((n/t)^{1/(d-1)})$. Moreover, for $d \leq \log(n/t)/10$, $k = \Omega(d \cdot (n/t)^{1/(d-1)})$.

[9] That is, every monomial of degree at most d appears in f with probability $1/2$, independently of all other monomials.

Before proving Theorem 19 we note that by applying a probabilistic argument, it can be shown that the theorem is tight. In particular, it has the right dependency in the number of polynomials t.

Proof. The proof is very similar to that of Theorem 1, so we only highlight the differences. As in the proof of Theorem 1, we may assume that $f_1, \ldots, f_t$ evaluate to 0 at u_0. We build by induction an affine subspace $u_0 + U$ on which all the t polynomials evaluate to 0. Given we already picked basis vectors $\Delta_1, \ldots, \Delta_k$, we consider the set A to be the following:

$$A = \left\{ x \in \mathbb{F}_q^n \;\middle|\; \forall i \in t, \;\; f_i|_{x+\mathrm{span}\{\Delta_1,\ldots,\Delta_k\}} \equiv 0 \right\}.$$

As in the proof of Theorem 1, A can be written as the set of solutions to a single polynomial equation $t(x) = 1$, where

$$\deg(t) \leq (q-1) \cdot \sum_{i=1}^{t} (d_i + 1) \sum_{j=0}^{d_i - 1} (d_i - j) \cdot \binom{k+j-1}{j},$$

Similarly to Theorem 1, the polynomial s is now defined, where $\deg(s) \leq (d+1) \cdot \deg(t)$ and such that any non-zero of s, that is independent of $\Delta_1, \ldots, \Delta_k$, can be taken to be Δ_{k+1}. By DeMillo-Lipton-Schwartz-Zippel lemma, it follows that as long k is not too large, such a root can be found. ◀

Similarly to the way we deduced Theorem 2 from Theorem 1, one can deduce the following theorem from Theorem 19. We omit the proof.

▶ **Theorem 20** (Structural Result II for many polynomials)**.** *Let q be a prime power. Let $f_1, \ldots, f_t : \mathbb{F}_q^n \to \mathbb{F}_q$ be polynomials of degree at most d. Then, there exists a partition of $\mathbb{F}_q^n$ to affine subspaces, each of dimension $\Omega(n^{1/(d-1)!}/t^e)$, such that $f_1, \ldots, f_t$ are all constant on each part.*

3.5 Sparse Polynomials

In this section we prove Theorem 3. To this end, we prove the following lemma.

▶ **Lemma 21.** *Let f be a polynomial on n variables over $\mathbb{F}_q$, with n^c monomials. If f is an affine disperser for dimension k, then there exists a subspace U of dimension $\Omega(\sqrt{n})$ on which $f|_U$ is of degree at most $2(q-1)c$.*

Lemma 21 implies Theorem 3. Indeed, the above lemma states that for any polynomial f on n variables and n^c monomials over $\mathbb{F}_q$, there exists an affine subspace of $\mathbb{F}_q^n$, with dimension $k(\Omega(\sqrt{n}), 2(q-1)c)$, on which f is constant. By Theorem 1, $k(\Omega(\sqrt{n}), 2(q-1)c) = \Omega(n^{1/(4(q-1)c)})$, as desired.

Proof of Lemma 21. We perform a random restriction to all variables $x_1, \ldots, x_n$. For each $i \in [n]$, independently, with probability $1 - (2 \cdot n^c)^{-1/(2c)}$, we set x_i to 0. Consider a monomial that has at least $2c$ distinct variables. The probability that such a monomial survives the restriction is at most $1/(2 \cdot n^c)$. Thus, by the union bound, with probability at least $1/2$, no monomial with more than $2c$ distinct variables survived the restriction. Restricting ourselves to this event, since we may assume that the individual degree of each variable in the original polynomial is at most $q-1$, any surviving monomial has degree at most $2(q-1)c$.

The expected number of variables that survived the random restriction is $n \cdot (2 \cdot n^c)^{-1/(2c)} = \Omega(\sqrt{n})$. Thus, by the Chernoff bound, with probability at least, say, 3/4, the number of surviving variables is $\Omega(\sqrt{n})$.

Thus, there exists a restriction of the variables that keeps $\Omega(\sqrt{n})$ of them alive, and such that the resulting polynomial has degree at most $2(q-1)c$. ◀

4 The Algorithmic Aspect

4.1 Efficient Algorithm for Finding a Somewhat Large Subspace

▶ **Theorem 22.** *Let $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$ be a polynomial of degree $d \leq \log(n)/3$ given as a black-box. Then, there exists an algorithm that makes $\mathrm{poly}(n)$ queries to f, runs in time $\mathrm{poly}(n)$, and finds an affine subspace U of dimension $\Omega(d \cdot n^{1/(d-1)})$ such that $\deg(f|_U) \leq d-1$.*

The proof of Theorem 22 is deferred to Appendix B.1 as it relies on notations and ideas from the proof of the first structural result for the binary field, which can be found in Appendix B. We advise the reader to look at the latter section before reading the proof of Theorem 22.

Theorem 22 yields the following corollary.

▶ **Corollary 23.** *There exists an algorithm that given a degree d polynomial $f : \mathbb{F}_2^n \to \mathbb{F}_2$ as a black box, runs in $\mathrm{poly}(n)$-time and finds an affine subspace of dimension $\Omega(n^{1/(d-1)!})$ on which f is constant.*

4.2 Subexponential-Time Algorithm for Finding an Optimal Subspace

▶ **Theorem 24.** *There exists a constant $\beta > 0$ such that the following holds. There is an algorithm that given $f\colon \mathbb{F}_q^n \to \mathbb{F}_q$, a degree d polynomial (as a list of monomials), where $3 \leq d \leq \log(n)/10$, and $u_0 \in \mathbb{F}_q^n$ as inputs, finds an affine subspace $u_0 + U$ of dimension $\Omega(k(n,d))$, restricted to which f is constant. The algorithm runs in time $q^{\beta \cdot n^{(d-2)/(d-1)}} \cdot \mathrm{poly}(n^d)$, and uses $\mathrm{poly}(n^d, \log q)$ space.*

We obtain the following corollary.

▶ **Corollary 25.** *There exists a $q^{n-k} \cdot \mathrm{poly}(n^d)$-time $\mathrm{poly}(n^d, \log q)$-space algorithm that given $f\colon \mathbb{F}_q^n \to \mathbb{F}_q$, a degree d polynomial, partitions $\mathbb{F}_q^n$ to affine subspace of dimension k on each of which f is constant, where $k = \Omega(n^{1/(d-1)!})$.*

In particular, one can compute the number of satisfying assignments for f using Corollary 25.

Proof. We follow the proof of Theorem 1. Again, we may assume $f(u_0) = 0$. Given the previously chosen vectors $\Delta_1, \ldots, \Delta_k$ such that f is the constant 0 on $u_0 + \mathrm{span}\{\Delta_1, \ldots, \Delta_k\}$, we show how to find a new vector Δ_{k+1} which is linearly independent of $\Delta_1, \ldots, \Delta_k$, such that f is constantly zero on $u_0 + \mathrm{span}\{\Delta_1, \ldots, \Delta_{k+1}\}$. The set A is the set of solutions to the following set of polynomial equations:

$$\{f_\alpha(x) = 0 \ : \ \alpha \in \{0, 1, \ldots, q-1\}^k, \mathrm{wt}(\alpha) \leq d-1\} \ ,$$

and by our assumptions, u_0 is a solution to all of these equations. By treating the polynomial f as a formal sum of monomials we can calculate each f_α in $\mathrm{poly}(n^d)$ time. Let R be some arbitrary subset of $\mathbb{F}_q$ of size $\min(q, d+1)$ then any solution y to the following set of equations which is linearly independent of $\Delta_1, \ldots, \Delta_k$ can be the new direction Δ_{k+1}:

$$\{f_\alpha(u_0 + r \cdot y) = 0 \ : \ \alpha \in \{0, 1, \ldots, q-1\}^k, \mathrm{wt}(\alpha) \leq d-1, \ r \in R\} \ .$$

It is therefore enough to find more than q^k different solutions to this set of equations, in order to guarantee that one of them will be linearly independent of the previous Δ_i's. In order to do so, we partition the set of equations into the set of linear equations and the set of non-linear equations:

$$L = \{f_\alpha(u_0 + r \cdot y) = 0 \ : \ \alpha \in \{0,1,\ldots,q-1\}^k, \mathrm{wt}(\alpha) \leq d-1, \deg(f_\alpha) = 1, \ r \in R\} \ .$$
$$NL = \{f_\alpha(u_0 + r \cdot y) = 0 \ : \ \alpha \in \{0,1,\ldots,q-1\}^k, \mathrm{wt}(\alpha) \leq d-1, \deg(f_\alpha) > 1, \ r \in R\} \ .$$

Let $m = \sum_{f_\alpha \in NL} \deg(f_\alpha)$. Since 0^n is a solution to all equations in $L \cup NL$, we can impose new linear equations which hold for 0^n, keeping the system consistent. More specifically, we define a new set L', which initially is equal to L, and iteratively add equations of the form $\{y_i = 0\}$ to L' until $\dim(L') = n - m - k - 1$. [10]

The set of solutions to both L' and NL is non-empty as it contains the all zeros vector. Furthermore, the sum of the degrees of equations in $L' \cup NL$ is exactly $(n-m-k-1)+m = n-k-1$. Therefore, by Lemma 13, there are at least q^{k+1} solutions to the equations in $L' \cup NL$, which guarantees that one of the solutions is linearly independent of $\Delta_1, \ldots, \Delta_k$.

Next, we show how to find all solutions to the equations in $L' \cup NL$. We find a basis for the set of solutions to L' using Gaussian elimination, and iterate over all vectors in the affine subspace this basis spans. For each vector y in this affine subspace we verify that all the equations in NL are satisfied by y. The running time of this process is $O(q^{n-\dim(L')} \cdot |NL| \cdot n^d)$, which is $O(q^{m+k+1} \cdot n \cdot n^d)$.

As $m \leq \min(d+1, q) \cdot \sum_{i=0}^{d-2} (d-i) \cdot \binom{k+i-1}{i}$, an elementary calculation shows that for $k \leq \frac{d}{10e} \cdot n^{1/(d-1)}$ and $3 \leq d \leq \log(n)/10$ we have $m + k \leq \beta \cdot n^{(d-2)/(d-1)}$ for some universal constant β. Thus, the total running time of the algorithm is $q^{\beta \cdot n^{(d-2)/(d-1)}} \cdot \mathrm{poly}(n^d)$. The algorithm uses $O((|NL| + |L|) \cdot n^d \cdot \mathrm{polylog}(q))$ space to store and manipulate the polynomials f_α. In addition, $O(n^2 \cdot \mathrm{polylog}(q))$ space is used to perform the Gaussian elimination. Overall the space used by the algorithm is $O(n^{d+1} \cdot \mathrm{polylog}(q))$. ◀

5 Extractors and Dispersers for Varieties

We start this section by proving Theorem 5.

Proof of Theorem 5. Let $g_1, \ldots, g_t \colon \mathbb{F}_q^n \to \mathbb{F}_q$ be degree d polynomials. By Theorem 20, there exists a partition of $\mathbb{F}_q^n$ to affine subspaces $P_1, \ldots, P_\ell$, each of dimension $\Omega(n^{1/(d-1)!}/t^e)$, such that $g_j|_{P_i}$ is constant for all $i \in [\ell]$ and $j \in [t]$. Since f is an affine extractor for such dimension, with bias ε, then for all $i \in [\ell]$ it holds that $\mathrm{SD}(f(P_i), \mathbb{F}_q) \leq \varepsilon$.

Let $I \subseteq [\ell]$ be the set of indices of affine subspaces in the partition such that $i \in I$ if and only if $g_j|_{P_i} = 0$ for all $j \in [t]$. In other words, we consider the partition of $\mathbf{V}(g_1, \ldots, g_t)$ to affine subspaces, induced by the partition of $\mathbb{F}_q^n$ to $P_1, \ldots, P_\ell$. Since the P_i's are disjoint, the random variable $f(\mathbf{V}(g_1, \ldots, g_t)) = f(\cup_{i \in I} P_i)$ is a convex combination of the random variables $\{f(P_i)\}_{i \in I}$. Thus, $\mathrm{SD}(f(\mathbf{V}(g_1, \ldots, g_t)), \mathbb{F}_q) \leq \max_{i \in I} \mathrm{SD}(f(P_i), \mathbb{F}_q) \leq \varepsilon$. ◀

We now give a formal statement and proof for the reduction from extractors for varieties to affine extractors, which does not depend on the number of polynomials defining the variety, but rather on the variety size.

[10] We add these constraints as concentrating at finding a solution of this form (that is, a solution that satisfies all equations in $L' \cup NL$ rather than only the equations in $L \cup NL$) is easier from the computational aspect.

▶ **Theorem 26.** *For every $d \in \mathbb{N}$ and $\delta, \rho \in (0,1)$ the following holds. Let $f\colon \mathbb{F}_q^n \to \mathbb{F}_q$ be an affine extractor for dimension $\Omega(n^{1/(d-1)!}/\ell^e)$ with bias ε, where $\ell = \log_q(1/(\rho\delta))$. Then, f is an extractor with bias $\varepsilon + \delta$ for varieties with density at least ρ (i.e., size at least $\rho \cdot q^n$), that are the common zeros of any degree (at most) d polynomials.*

Proof. Let $g_1, \ldots, g_t\colon \mathbb{F}_q^n \to \mathbb{F}_q$ be degree (at most) d polynomials. First, we prove the existence of ℓ polynomials $h_1, \ldots, h_\ell\colon \mathbb{F}_q^n \to \mathbb{F}_q$, each of degree at most d, with a variety that approximates $\mathbf{V}(g_1, \ldots, g_t)$. More precisely, we will have

$$\mathbf{V}(g_1, \ldots, g_t) \subseteq \mathbf{V}(h_1, \ldots, h_\ell) \quad \text{and} \quad \Pr_{x \sim \mathbb{F}_q^n}[x \in \mathbf{V}(h_1, \ldots, h_\ell) \setminus \mathbf{V}(g_1, \ldots, g_t)] \leq q^{-\ell}, \tag{4}$$

The proof of this claim follows by a standard argument, like the one that appears in [21, 27]: Let $\alpha_1, \ldots, \alpha_\ell$ be random vectors, sampled uniformly and independently from $\mathbb{F}_q^t$. For each $i \in [\ell]$, define the (random) polynomial

$$H_i(x) = \sum_{j=1}^{t} (\alpha_i)_j \cdot g_j(x),$$

where the summation and multiplications are taken over $\mathbb{F}_q$. Clearly, if $x \in \mathbf{V}(g_1, \ldots, g_t)$ then $H_i(x) = 0$ with probability 1 (where the probability is taken over $\alpha_1, \ldots, \alpha_\ell$). Otherwise, for each $i \in [\ell]$, $\Pr[H_i(x) = 0] = 1/q$. By an averaging argument, one can fix $\alpha_1, \ldots, \alpha_\ell$ and obtain fixed polynomials $h_1, \ldots, h_\ell$, of degree at most d, that satisfy the conditions in Equation (4).

Since f is an affine extractor with bias ε for dimension $\Omega(n^{1/(d-1)!}/\ell^e)$, Theorem 5 implies that $\mathrm{SD}(f(\mathbf{V}(h_1, \ldots, h_\ell)), \mathbb{F}_q) \leq \varepsilon$. To conclude the proof, we show that

$$\mathrm{SD}(f(\mathbf{V}(h_1, \ldots, h_\ell)), f(\mathbf{V}(g_1, \ldots, g_t))) \leq \delta.$$

To see this, observe that $\mathbf{V}(h_1, \ldots, h_\ell)$ can be written as a convex combination

$$\mathbf{V}(h_1, \ldots, h_\ell) = \frac{|\mathbf{V}(g_1, \ldots, g_t)|}{|\mathbf{V}(h_1, \ldots, h_\ell)|} \cdot \mathbf{V}(g_1, \ldots, g_t) + \left(1 - \frac{|\mathbf{V}(g_1, \ldots, g_t)|}{|\mathbf{V}(h_1, \ldots, h_\ell)|}\right) \cdot \mathcal{E},$$

where $\mathcal{E}$ is some random variable over $\mathbb{F}_q$. Thus, by Equation (4),

$$\mathrm{SD}(\mathbf{V}(h_1, \ldots, h_\ell), \mathbf{V}(g_1, \ldots, g_t)) \leq 1 - \frac{|\mathbf{V}(g_1, \ldots, g_t)|}{|\mathbf{V}(h_1, \ldots, h_\ell)|} \leq \frac{q^{-\ell}}{\rho} = \delta.$$

This implies that $\mathrm{SD}(f(\mathbf{V}(h_1, \ldots, h_\ell)), f(\mathbf{V}(g_1, \ldots, g_t))) \leq \delta$, as claimed. ◀

Next, we prove Theorem 6 which gives an analog reduction from dispersers for varieties to affine dispersers.

Proof of Theorem 6. Let $g_1, \ldots, g_t\colon \mathbb{F}_q^n \to \mathbb{F}_q$ be degree (at most) d polynomials. Let $u_0 \in \mathbf{V}(g_1, \ldots, g_t)$ (if $\mathbf{V}(g_1, \ldots, g_t) = \emptyset$, there is nothing to prove). By Theorem 19, there exists a subspace U of dimension $\Omega(d \cdot (n/t)^{1/(d-1)})$ such that $u_0 + U \subseteq \mathbf{V}(g_1, \ldots, g_t)$. The proof then follows as f is an affine disperser for dimension $\Omega(d \cdot (n/t)^{1/(d-1)})$. ◀

6 From Affine Dispersers to Affine Extractors

To prove Theorem 8, we use the following theorem of Kaufman and Lovett [19].

▶ **Theorem 27** ([19]). *Let p be a prime number and let $f : \mathbb{F}_p^n \to \mathbb{F}_p$ be a degree (at most) d polynomial with* $\mathrm{bias}(f) \geq \delta$. *Then, there exist $c = c(d, \delta)$ polynomials $f_1, \ldots, f_c$ of degree at most $d-1$ such that $f = G(f_1, \ldots, f_c)$, for some function $G : \mathbb{F}_p^c \to \mathbb{F}_p$. Moreover, $f_1, \ldots, f_c$ are derivatives of the form $\frac{\partial f}{\partial y}$ where $y \in \mathbb{F}_p^n$.*

Proof of Theorem 8. We show by a counter-positive argument that if f is not an affine extractor for dimension k' with bias δ, then f is not an affine disperser for dimension k. Let $f : \mathbb{F}_p^n \to \mathbb{F}_p$ be a function which is not an affine extractor for dimension k' with bias δ. Then, there exists an affine subspace $u_0 + U$, with $\dim(U) = k'$ such that $\mathrm{bias}(f|_{u_0+U}) > \delta$. Let $u_1, \ldots, u_{k'}$ be a basis for U and let $g : \mathbb{F}_p^{k'} \to \mathbb{F}_p$ be the function defined by $g(y_1, \ldots, y_{k'}) = f(u_0 + \sum_{i=1}^{k'} u_i \cdot y_i)$. Then, g is a δ-biased polynomial of degree $\leq d$. Applying Theorem 27 to g, we can write it as $G(g_1, \ldots, g_c)$, where the g_i's are of degree at most $d-1$, and $c = c(d, \delta)$ as defined in Theorem 27.

By Theorem 19, there is an affine subspace W of $\mathbb{F}_p^{k'}$ with dimension $c_1 \cdot (k'/c)^{1/(d-2)}$ for which all the g_i's are constant, for some constant $c_1 > 0$. In particular $g|_W$ is constant, which implies that there exists a subspace of $\mathbb{F}_p^n$, with the same dimension, on which the original function f is constant. Taking $k' = k^{d-2} \cdot \frac{c(d,\delta)}{c_1^{d-2}}$ completes the proof. ◀

For degree 3 and 4, we rely on stronger results from [17]. Although degree 3 was treated in [6], we present it here for completeness.

▶ **Theorem 28.** *Let $f : \mathbb{F}_p^n \to \mathbb{F}_p$ be an affine disperser for dimension k of degree d. If $d = 3$ then f is an affine extractor for dimension $k' = k + O(\log(1/\delta)^2)$ with bias δ. If $d = 4$ then f is an affine extractor for dimension $k' = k \cdot \mathrm{poly}(1/\delta)$ with bias δ.*

Proof. As in the proof of Theorem 8, it is enough to show that if g is a degree 3 or 4 polynomial over $\mathbb{F}_p$ with k' variables and bias $\geq \delta$ then there exists a subspace of dimension k on which g is constant. We consider the two cases $\deg(f) = 3, 4$ separately.

Cubic ($\deg(g) = 3$)

Implicit in [17], any polynomial of degree 3 with bias $\geq \delta$, in particular g, can be represented as $\sum_{i=1}^{r} \ell_i(x) \cdot q_i(x) + q_0(x)$, where the ℓ_i's are linearly independent linear functions (with no constant term), $\deg(q_i) \leq 2$ and $r = O(\log^2(1/\delta))$. Restricting to the subspace W defined by $\{x : \ell_i(x) = 0\}$ reduces the degree of g to at most 2, and by Lemma 16, this is also true for any coset of this subspace. By averaging, there is a coset on which $\mathrm{bias}(g|_{w+W}) \geq \delta$. By Dickson's theorem [9], there is an affine subspace $w' + W'$ of $w + W$ of co-dimension $O(\log(1/\delta))$ on which g is constant. Setting $k' = k + O(\log^2(1/\delta))$ ensures that $\dim(W')$ is at least k.

Quartic ($\deg(g) = 4$)

Theorem 4 in [17] states that any polynomial of degree 4 with bias $\geq \delta$, in particular g, can be represented as

$$\sum_{i=1}^{r} \ell_i(x) \cdot g_i(x) + \sum_{i=1}^{r} q_i(x) \cdot q_i'(x) + g_0(x),$$

where $\deg(\ell_i) \leq 1, \deg(q_i) \leq 2, \deg(q_i') \leq 2, \deg(g_i) \leq 3$ and $r = \mathrm{poly}(1/\delta)$. By Theorem 19, there exists a subspace W of dimension $\Omega(n/r)$ on which all ℓ_i's, q_i's and q_i''s are constants. By Lemma 16, in any coset of W the degrees of ℓ_i, q_i and q_i' for $i = 1, \ldots, r$ are decreased

by at least 1, hence $g|_{w+W}$ is of degree at most 3 for any coset $w+W$. Since $\text{bias}(g) \geq \delta$, by averaging there is a coset on which $\text{bias}(g|_{w+W}) \geq \delta$. Using the earlier case of biased cubic polynomials, there is an affine subspace $w'+W'$ of dimension $\Omega(n/r) - O(\log^2(1/\delta))$ on which g is constant. Setting $k' = k \cdot \text{poly}(1/\delta)$ ensures that the dimension of W' is at least k. ◀

Remark

It may be tempting to think that the polynomial loss of parameters in our reduction from affine extractors to affine dispersers, $k' = O_{\delta,d}(k^{d-2})$, is not necessary. Indeed, Theorem 28 shows that for degree 3 and 4 one can take the dimension k' of the affine extractor (for a constant error, say) to be linear in k – the dimension of the affine disperser. However, this linear dependency breaks for $d \geq 6$, as pointed up to us by Shachar Lovett. To see this, take $f : \mathbb{F}_2^n \to \mathbb{F}_2$ to be the product of two random degree 3 polynomials. It is easy to check that, with high probability, f is an affine disperser for dimension $\Theta(\sqrt{n})$, whereas $\Pr[f=1] = 1/4 + o(1)$. Namely, f is not even an (n,n) affine extractor.

Nonetheless, a better polynomial dependency may still be possible. Perhaps $k' = O_{\delta,d}(k^{(d-2)/2})$ (which is not ruled out by similar counterexamples).

7 $\mathsf{AC}^0[\oplus]$ Circuits and Affine Extractors / Dispersers

In Section 7.1 we (easily) derive lower bounds on the dimension for which an AC^0 circuit can be affine disperser. In Section 7.2 we prove that a depth 2 $\mathsf{AC}^0[\oplus]$ circuit on n inputs cannot compute an affine disperser for dimension $n^{o(1)}$. We do so by a reduction to Theorem 1.

7.1 AC^0 Circuits Cannot Compute Affine Dispersers for Dimension $o(n/\text{polylog}(n))$

The next lemma, following Håstad's work [14], appears in [5].

▶ **Lemma 29** ([5], Corollary 3.7, restated). *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a function computable by a depth d and size s Boolean circuit. Then, there is a restriction ρ leaving $\frac{n}{10(10\log(s))^{d-2}} - \log(s)$ variables alive, under which $f|_\rho$ is constant.*

Lemma 29 readily implies the following corollary.

▶ **Corollary 30.** *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a function computable by a Boolean circuit of depth d and size s. Then, f cannot be a bit fixing disperser (and, in particular, f cannot be an affine disperser) for min-entropy $k < \frac{n}{10(10\log(s))^{d-2}} - \log(s)$.*

7.2 Depth 2 $\mathsf{AC}^0[\oplus]$ Circuits Cannot Compute Good Affine Dispersers

As mentioned in the introduction, to prove Theorem 10, one only needs to prove Lemma 9.

Proof of Lemma 9. During the proof we will exploit the fact that if a function f on n inputs is an affine disperser for dimension k, then fixing the values of m inputs or even the values of m linear functions on the inputs, one gets an affine disperser on $n-m$ inputs for the same dimension k.

We assume that the top gate is an XOR gate. Afterwards we justify this assumption by showing that if the top gate is not an XOR gate, then the circuit C could not have computed an affine disperser with the claimed parameters to begin with.

Note that one might as well assume that there are no XOR gates at the bottom level. Indeed, assume there are t XOR gates at the bottom level, and denote by $\ell_1, \ldots, \ell_t$ the linear functions computed by these gates, respectively. Define the linear function $\ell = \ell_1 \oplus \cdots \oplus \ell_t$. Note that if ℓ is the constant 1 then by removing all the t gates from C and wiring the constant 1 as an input to the top gate, one gets an equivalent circuit with no XOR gates at the bottom layer. Assume therefore that ℓ is not the constant 1. Then, by removing all the XOR gates at the bottom layer, we get a circuit, with no XOR gates at the bottom layer, that is equivalent to the original circuit on the affine subspace $\{x : \ell(x) = 0\}$. Hence, the resulting circuit is an affine disperser on $n-1$ inputs for dimension k.

We perform a random restriction to all variables, leaving a variable alive with probability $p = \frac{1}{4\sqrt{n}}$ and otherwise setting the value of a variable uniformly and independently at random. We show that the restriction shrinks all OR, AND gates to have fan-in smaller than $2c$ with positive probability. We consider AND gates, but our arguments may be carried to OR gates similarly. The restriction shrinks every AND gate in the following way: if one of the literals which is an input to the AND gate is false under the restriction, the AND gate is eliminated. Otherwise, the AND gate shrinks to be the AND of all the remaining live variables. We wish to bound the probability that each AND gate is of fan-in greater than $2c$ after the restriction. Let m be the fan-in of the AND gate before the restriction, and m' its fan-in afterwards. We have

$$\begin{aligned}\Pr[m' \geq 2c] &= \sum_{i=2c}^{m} \binom{m}{i} \cdot p^i \cdot \left(\frac{1-p}{2}\right)^{m-i} \leq \sum_{i=2c}^{m} \binom{m}{i} \cdot p^i \cdot (1/2)^{m-i} \\ &= (1/2)^m \cdot \sum_{i=2c}^{m} \binom{m}{i} \cdot (2p)^i \ .\end{aligned}$$

Since $2p$ is smaller than 1, the right hand side of the above inequality is at most $(1/2)^m \cdot 2^m \cdot (2p)^{2c} = (2p)^{2c}$. Thus, $\Pr[m' \geq 2c] \leq (2p)^{2c}$. By our choice of parameter p, this is at most $1/(4n)^c$. By union bound over all $\leq n^c$ AND and OR gates, with probability at least $1 - 1/4^c \geq 3/4$ over the random restrictions, the fan-in of all AND and OR gates, under the restriction, is smaller than $2c$. Furthermore, by Chernoff bound, with probability greater than $1/2$ over the random restrictions, the number of surviving variables is at least $\sqrt{n}/5$. Therefore, there exists a restriction where the number of surviving variables is $\sqrt{n}/5$ and all AND and OR gates in the resulting circuit, under the restriction, have fan-in smaller than $2c$. Expressing the resulting circuit as a polynomial over $\mathbb{F}_2$ we get a polynomial on at least $\sqrt{n}/5$ variables with degree at most $2c$ which is an affine disperser for dimension k.

We are left to justify the assumption that the top gate must be an XOR gate. For contradiction, assume that the top gate is an OR gate. The case where the top gate is an AND gate is handled similarly. If there is an XOR gate at the bottom layer of C, we choose such gate and consider the affine subspace of co-dimension 1 on which this XOR gate outputs 1. Since the top gate is an OR gate, the circuit C is the constant 1 on an affine subspace of co-dimension 1. This stands in contradiction as k is (much) smaller than $n-1$. Thus, we obtain a depth 2 AC^0 circuit with size $s = n^c$. However, under the assumption that $k < n/10 - \log(s)$ this is a contradiction to Corollary 30. ◀

Acknowledgement. We wish to thank our advisor Ran Raz for many helpful discussions and for his encouragement. We thank Chaim Even Zohar, Elad Haramaty, Noam Lifshitz and Amir Shpilka for helpful discussions regarding this work. We thank the user who goes by the name david from Stack Exchange for pointing out [2]. We thank the anonymous referees for pointing out [28] and for many helpful comments.

References

1 Sanjeev Arora and Boaz Barak. *Computational Complexity - A Modern Approach*. Cambridge University Press, 2009.

2 I. Ben-Eliezer, R. Hod, and S. Lovett. Random low degree polynomials are hard to approximate. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 366–377. Springer, 2009.

3 A. Bhattacharyya, S. Kopparty, G. Schoenebeck, M. Sudan, and D. Zuckerman. Optimal testing of reed-muller codes. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 488–497. IEEE, 2010.

4 J. Bourgain. On the construction of affine extractors. *GAFA Geometric And Functional Analysis*, 17(1):33–57, 2007.

5 R. B. Boppana and M. Sipser. The complexity of finite functions. In *Handbook of Theoretical Computer Science, Volume A: Algorithms and Complexity (A)*, pages 757–804. 1990.

6 E. Ben-Sasson and S. Kopparty. Affine dispersers from subspace polynomials. *SIAM Journal on Computing*, 41(4):880–914, 2012.

7 E. Ben-Sasson and N. Zewi. From affine to two-source extractors via approximate duality. In *Proceedings of the 43rd annual ACM symposium on Theory of computing*, pages 177–186. ACM, 2011.

8 M. DeVos and A. Gabizon. Simple affine extractors using dimension expansion. In *Computational Complexity (CCC), 2010 IEEE 25th Annual Conference on*, pages 50–57. IEEE, 2010.

9 L. E. Dickson. *Linear groups with an exposition of the Galois field theory*. B.G Teubner's Sammlung von Lehrbuchern auf dem Gebiete der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen. B.G. Teubner, 1901.

10 R. A. DeMillo and R. J. Lipton. A probabilistic remark on algebraic program testing. *Inf. Process. Lett.*, 7(4):193–195, 1978.

11 Z. Dvir. Extractors for varieties. *computational complexity*, 21(4):515–572, 2012.

12 A. Gabizon and R. Raz. Deterministic extractors for affine sources over large fields. *Combinatorica*, 28(4):415–440, 2008.

13 B. Green and T. Tao. The distribution of polynomials over finite fields, with applications to the gowers norms. *Contributions to Discrete Mathematics*, 4(2), 2009.

14 J. Håstad. Almost optimal lower bounds for small depth circuits. In *STOC*, pages 6–20, 1986.

15 J. Håstad. The shrinkage exponent of de Morgan formulas is 2. *SIAM Journal on Computing*, 27(1):48–64, 1998.

16 P. Hrubeš and A. Rao. Circuits with medium fan-in. *Electronic Colloquium on Computational Complexity (ECCC)*, 20, 2014.

17 E. Haramaty and A. Shpilka. On the structure of cubic and quartic polynomials. In *Proceedings of the 42nd ACM symposium on Theory of computing*, pages 331–340. ACM, 2010.

18 S. Jukna. *Boolean function complexity: advances and frontiers*, volume 27. Springerverlag Berlin Heidelberg, 2012.

19 T. Kaufman and S. Lovett. Worst case to average case reductions for polynomials. In *Foundations of Computer Science (FOCS), 2008 49th Annual IEEE Symposium on*, pages 166–175. IEEE, 2008.

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

21 A. Razborov. Lower bounds on the size of bounded depth networks over a complete basis with logical addition (Russian). *Matematicheskie Zametki*, 41(4):598–607, 1987.

22 A. Razborov. Bounded-depth formulas over $\{\wedge, \oplus\}$ and some combinatorial problems. *Complexity of Algorithms and Applied Mathematical Logic (in Russian). Ser. Voprosy Kibernetiky (Problems in Cybernetics), S. I. Adian, Ed., Moscow*, pages 149–166, 1988.

23 A. Razborov and S. Rudich. Natural proofs. In *Proceedings of the twenty-sixth annual ACM symposium on Theory of computing*, pages 204–213. ACM, 1994.

24 P. Savickỳ. Improved Boolean formulas for the Ramsey graphs. *Random Structures & Algorithms*, 6(4):407–415, 1995.

25 J. T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *J. ACM*, 27(4):701–717, 1980.

26 R. Shaltiel. Dispersers for affine sources with sub-polynomial entropy. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pages 247–256. IEEE, 2011.

27 R. Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proceedings of the nineteenth annual ACM symposium on Theory of computing*, STOC'87, pages 77–82, New York, NY, USA, 1987. ACM.

28 G. Tardos and D. A. M. Barrington. A lower bound on the mod 6 degree of the or function. *Computational Complexity*, 7(2):99–108, 1998.

29 L. G. Valiant. Graph-theoretic arguments in low-level complexity. In Jozef Gruska, editor, *MFCS*, volume 53 of *Lecture Notes in Computer Science*, pages 162–176. Springer, 1977.

30 L. A. Vinh. The szemerédi–trotter type theorem and the sum-product estimate in finite fields. *European Journal of Combinatorics*, 32(8):1177–1181, 2011.

31 E. Viola. Guest column: correlation bounds for polynomials over {0,1}. *ACM SIGACT News*, 40(1):27–44, 2009.

32 E. Viola and A. Wigderson. Norms, XOR lemmas, and lower bounds for GF(2) polynomials and multiparty protocols. In *Computational Complexity, 2007. CCC'07. Twenty-Second Annual IEEE Conference on*, pages 141–154. IEEE, 2007.

33 A. Yehudayoff. Affine extractors over prime fields. *Combinatorica*, 31(2):245–256, 2011.

34 R. Zippel. Probabilistic algorithms for sparse polynomials. In Edward W. Ng, editor, *EUROSAM*, volume 72 of *Lecture Notes in Computer Science*, pages 216–226. Springer, 1979.

A Depth 3 $AC^0[\oplus]$ Circuits Can Compute Optimal Affine Extractors

We start this section by giving a proof for the following folklore lemma. We bother doing so because afterwards we argue that the proof implies, in fact, something stronger, which we make use of.

▶ **Lemma 31.** *There exist universal constants n_0, c such that the following holds. For every $\varepsilon > 0$ and $n > n_0$ there exists an affine extractor for dimension k with bias ε, $f : \mathbb{F}_2^n \to \mathbb{F}_2$, where $k = \log \frac{n}{\varepsilon^2} + \log\log \frac{n}{\varepsilon^2} + c$.*

The proof of Lemma 31 makes use of Hoeffding bound.

▶ **Theorem 32** (Hoeffding Bound)**.** *Let $X_1, \ldots, X_n$ be independent random variables for which $X_i \in [a_i, b_i]$. Define $X = \frac{1}{n} \cdot \sum_{i=1}^n X_i$, and let $\mu = \mathbb{E}[X]$. Then,*

$$\Pr[|X - \mu| \geq \varepsilon] \leq 2 \cdot \exp\left(-\frac{2n^2\varepsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

Proof of Lemma 31. Let $F : \mathbb{F}_2^n \to \mathbb{F}_2$ be a random function, that is, $\{F(x)\}_{x \in \mathbb{F}_2^n}$ are independent random bits. Fix an affine subspace $u_0 + U \subseteq \mathbb{F}_2^n$ of dimension k as defined

above. By Hoeffding Bound (Theorem 32),

$$\Pr\left[\frac{1}{2^k}\left|\sum_{u\in u_0+U}(-1)^{F(u)}\right|\geq\varepsilon\right]\leq 2\cdot\exp\left(-\frac{2^k\varepsilon^2}{2}\right).$$

The number of affine subspaces of dimension k is bounded by $2^n\binom{2^n}{k}\leq 2^{(k+1)n}$. Hence, by union bound over all affine subspaces, if $2^{(k+1)n}\cdot 2e^{-2^k\varepsilon^2/2}<1$ then there exists a function $f:\mathbb{F}_2^n\to\mathbb{F}_2$ that is an affine extractor for dimension k with bias ε. It is a simple calculation to show that our choice of k suffices for the above equation to hold. ◀

For the proof of Theorem 36, we introduce the following notion.

▶ **Definition 33.** An (n,k,d) *linear injector* with size m is a family of $d\times n$ matrices $\{A_1,\ldots,A_m\}$ over $\mathbb{F}_2$ with the following property: for every subspace $U\subseteq\mathbb{F}_2^n$ of dimension k, there exists an $i\in[m]$ such that $\ker(A_i)\cap U=\{0\}$.

▶ **Lemma 34.** *For every n,k such that $2\leq k\leq n$, there exists an $(n,k,k+1)$ linear injector with size $m=nk$.*

Proof. Fix a subspace $U\subseteq\mathbb{F}_2^n$ of dimension k. Let A be a $d\times n$ matrix such that every entry of A is sampled from $\mathbb{F}_2$ uniformly and independently at random. For every $u\in U\setminus\{0\}$ it holds that $\Pr[Au=0]=2^{-d}$. By taking the union bound over all elements in $U\setminus\{0\}$, we get that $\Pr[\ker(A)\cap U\neq\{0\}]\leq 2^{k-d}$. Let $A_1,\ldots,A_m$ be $d\times n$ matrices such that the entry of each of the matrices is sampled from $\mathbb{F}_2$ uniformly and independently at random. By the above equation, it holds that $\Pr[\forall i\in[m]\ \ker(A_i)\cap U\neq\{0\}]\leq 2^{m(k-d)}$. The number of linear subspaces of dimension k is bounded above by $\binom{2^n}{k}$, which is bounded above by 2^{nk-1} for $k\geq 2$. Thus, if $2^{nk-1}\cdot 2^{m(k-d)}<1$ there exists an (n,k,d) linear injector with size m. The latter equation holds for $d=k+1$ and $m=nk$. ◀

▶ **Lemma 35.** *Let n_0,c be the constants from Lemma 31. Let $n>n_0$ and let k,ε be such that $k=\log\frac{n}{\varepsilon^2}+\log\log\frac{n}{\varepsilon^2}+c$. Let $\{A_1,\ldots,A_m\}$ be an (n,k,d) linear injector with size m. Then, there exist functions $f_1,\ldots,f_m:\mathbb{F}_2^d\to\mathbb{F}_2$ such that the function $f:\mathbb{F}_2^n\to\mathbb{F}_2$ defined by*

$$f(x)=\bigoplus_{i=1}^{m}f_i(A_ix) \tag{5}$$

is an affine extractor for dimension k with bias ε.

Proof. Recall that in the proof of Lemma 31, we took F to be a random function. We observe however, that the proof did not use the full independence offered by a uniformly sampled random function. In fact, the proof required only that for every affine subspace $u_0+U\subseteq\mathbb{F}_2^n$ of dimension k, $\{f(u)\}_{u\in u_0+U}$ are independent random bits.

Let $F_1,\ldots,F_m:\mathbb{F}_2^d\to\mathbb{F}_2$ be independent random functions, that is, the random bits $\{F_i(x)\}_{i\in[m],x\in\mathbb{F}_2^d}$ are independent. Define the random function $F:\mathbb{F}_2^n\to\mathbb{F}_2$ as follows

$$F(x)=\bigoplus_{i=1}^{m}F_i(A_ix).$$

We claim that for every affine subspace $u_0+U\subseteq\mathbb{F}_2^n$ of dimension k, the random bits $\{F(u)\}_{u\in u_0+U}$ are independent. By the observation above, proving this will conclude the proof. Let $u_0+U\subseteq\mathbb{F}_2^n$ be an affine subspace of dimension k. As $\{A_1,\ldots,A_m\}$ is an (n,k,d)

linear injector, there exists an $i \in [m]$ such that $\ker(A_i) \cap U = \{0\}$. This implies that for every two distinct elements $u, v \in U$ it holds that $A_i(u_0 + u) \neq A_i(u_0 + v)$. Otherwise $A_i(u + v) = 0$ and thus $u + v$, a non-zero vector in U, lies in $\ker(A_i)$. This stands in contradiction to the choice of i. Recall that F_i is a random function, and from the above it follows that A_i behaves as an injection to the domain $u_0 + U$. Hence, the random bits $\{F_i(A_i u)\}_{u \in u_0 + U}$ are independent. Since $F(x)$ is defined to be the XOR of $F_i(A_i x)$ with $m - 1$ other *independent* random variables, we get that $\{F(u)\}_{u \in u_0 + U}$ are also independent random bits, as claimed. ◀

▶ **Theorem 36.** *Let f be the function from Equation (5), where $\{A_1, \ldots, A_m\}$ is the (n, k, d) linear injector from Lemma 34 (that is, $m = nk$ and $d = k+1$). Then, f is an affine extractor for dimension k and bias ε, where $k = \log(n/\varepsilon^2) + \log\log(n/\varepsilon^2) + O(1)$. Moreover,*

1. $\deg(f) = \log(n/\varepsilon^2) + \log\log(n/\varepsilon^2) + O(1)$.
2. *f can be realized by an* XOR−AND−XOR *circuit of size $O((n/\varepsilon)^2 \cdot \log^3(n/\varepsilon))$.*
3. *f can be realized by a De Morgan formula of size $O((n^5/\varepsilon^2) \cdot \log^3(n/\varepsilon))$.*

Proof. To prove the first item, we note that each of the f_i's is a function on $d = k + 1$ inputs, and thus can be computed by a polynomial with degree at most $k + 1$. The proof then follows as in the computation of f, each f_i is composed with linear functions of the variables, and f is the XOR of the f_i's.

To prove the second item, we show an XOR−AND−XOR circuit C with the desired size, that computes the function f. Since each of the functions f_i are degree d polynomials on d inputs, each of them can be computed by an XOR − AND circuit, where the fan-in of the top XOR gate is bounded above by 2^d and the fan-in of each AND gate is at most d. Thus, for $i \in [m]$, each of the functions $f_i(A_i x)$ on n inputs is computable by an XOR − AND − XOR circuit.

By its definition, f is the XOR of these functions and so one can collapse this XOR together with the top m XOR gates. This yields an XOR − AND − XOR circuit C that computes f.

The size of the circuit C is $O(m \cdot d \cdot 2^d)$ as each of the m functions $f_i(A_i x)$ applies 2^d AND gates, each on d XOR gates (whom in turn compute the linear injector). Since $m = nk$ and $d = k + 1$, $\text{size}(C) = O((n/\varepsilon)^2 \cdot \log^3(n/\varepsilon))$ as stated.

As for the third item, we show a De Morgan formula with the desired size, that computes f. Since each of the functions f_i are on d inputs, each of them can be computed by a De Morgan formula of size $O(2^d)$. Moreover, every XOR operation needed for the computation of the linear injector $\{A_1, \ldots, A_m\}$ can be implemented in size $O(n^2)$. Replacing each leaf in the formula for f_i with the relevant formula computing the corresponding bit of $A_i x$ (or its negation), results in an $O(2^d n^2)$ size De Morgan formula computing $f_i(A_i x)$. Again, since the XOR of bits $y_1, \ldots, y_m$ can be computed by a De Morgan formula of size $O(m^2)$, and one can replace each leaf marked by y_i (or $\neg y_i$) with the formula computing $f_i(A_i x)$ (or its negation), one gets a De Morgan formula computing f of size

$$O(m^2 \cdot 2^d \cdot n^2) = O((nk)^2 \cdot 2^k \cdot n^2) = O((n^5/\varepsilon^2) \cdot \log^3(n/\varepsilon)),$$

as desired. ◀

B A Slightly Simpler Proof of the First Structural Result for $\mathbb{F}_2$

In this section we give a slightly simpler proof for Theorem 1, for the special case $q = 2$, based on ideas in [28]. We prove the following:

▶ **Theorem 37** (Structural Result I for the Binary Field). *Let k be the smallest integer such that*

$$n \leq k + \sum_{j=0}^{d-1} (d-j) \cdot \binom{k}{j} .$$

Let $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$ be a degree d polynomial, and let $u_0 \in \mathbb{F}_2^n$. Then, there exists a subspace $U \subset \mathbb{F}_2^n$ of dimension k such that $f|_{u_0+U}$ is constant.

Proof. Fix $u_0 \in \mathbb{F}_2^n$. We assume without loss of generality that $f(u_0) = 0$, as otherwise we can look at the polynomial $g(x) = f(x) - f(u_0)$ which is of the same degree. The proof is by induction. Let k be such that

$$n > k + \sum_{j=0}^{d-1} (d-j) \cdot \binom{k}{j} . \tag{6}$$

We assume by induction that there exists an affine subspace $u_0 + \operatorname{span}\{\Delta_1, \ldots, \Delta_k\} \subseteq \mathbb{F}_2^n$, where the Δ_i's are linearly independent vectors on which f evaluates to 0. Assuming Equation 6 holds, we show there exists a vector Δ_{k+1}, linearly independent of $\Delta_1, \ldots, \Delta_k$, such that $f \equiv 0$ on $u_0 + \operatorname{span}\{\Delta_1, \ldots, \Delta_{k+1}\}$. To this aim, consider the set

$$A = \left\{ x \in \mathbb{F}_2^n \;\middle|\; \forall S \subseteq [k],\ f\left(x + \sum_{i \in S} \Delta_i\right) = 0 \right\}.$$

By the induction hypothesis, $u_0 \in A$. It can be verified that for any $x \in \mathbb{F}_2^n$

$$\forall S \subseteq [k] : f\left(x + \sum_{i \in S} \Delta_i\right) = 0 \quad \Leftrightarrow \quad \forall S \subseteq [k] : f_S(x) = 0 \,,$$

where f_S is defined by

$$f_S(x) \triangleq \sum_{T \subseteq S} f\left(x + \sum_{i \in T} \Delta_i\right).$$

Namely, f_S is the derivative of f in directions $\{\Delta_i\}_{i \in S}$. In particular, $\deg(f_S) \leq d - |S|$. Thus $f_S \equiv 0$ for $|S| > d$, and we may write A as

$$A = \{x \in \mathbb{F}_2^n \mid \forall S \subseteq [k] : |S| \leq d,\ f_S(x) = 0\} .$$

Hence, A is the set of solutions to a system of $\binom{k}{\leq d}$ polynomial equations, where there are $\binom{k}{j}$ equations which correspond to sets S of size j and thus to degree (at most) $d - j$ polynomials. [11] One can also write A as the set of solutions to the single polynomial equation

$$\prod_{S \subseteq [k] : |S| \leq d} (1 - f_S(x)) \ = 1,$$

[11] In particular, equations that correspond to sets S of size d are of the form $c_S = 0$ for some constant $c_S \in \mathbb{F}_2$. Since A is non-empty, the constants c_S must be 0, making those equations tautologies $0 = 0$ that does not depend on x. Moreover, most of the remaining equations correspond to sets S of size $d - 1$, and are therefore either linear equations or tautologies.

which is of degree $D \leq \sum_{j=0}^{d-1} (d-j) \cdot \binom{k}{j}$. Since A is non-empty, by DeMillo-Lipton-Schwartz-Zippel lemma (Lemma 13, for $q = 2$) we have that

$$|A| \geq 2^{n-D} \geq 2^{n-\sum_{j=0}^{d-1}(d-j)\cdot\binom{k}{j}}. \tag{7}$$

This, together with Equation (6) implies that $|A| > 2^k$. Hence, there exists a point $y \in A$ such that $y - u_0 \notin \text{span}\{\Delta_1, \Delta_2, \ldots, \Delta_k\}$. Pick such a point u arbitrarily and denote by $\Delta_{k+1} \triangleq u - u_0$. Since both u_0 and u are in A we have that $f \equiv 0$ on $u_0 + \text{span}\{\Delta_1, \ldots, \Delta_{k+1}\}$. The inductive proof shows that there exists a subspace U of dimension k such that f is constant on $u_0 + U$ and

$$n \leq k + \sum_{j=0}^{d-1} (d-j) \cdot \binom{k}{j}, \tag{8}$$

since otherwise we could have continue this process and pick a bigger subspace U'. ◀

B.1 Proof of Theorem 22

The proof of Theorem 22 uses the following lemma.

▶ **Lemma 38.** *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a degree d polynomial, and let U be a linear subspace with basis $\Delta_1, \ldots, \Delta_k$. Then, $\deg(f|_U) \leq d-1$ if and only if $f_S(0) = 0$ for all $S \subseteq [k]$ of size d, where $f_S(x) := \sum_{T \subseteq S} f\left(x + \sum_{i \in T} \Delta_i\right)$.*

Proof of Lemma 38. As noted in the Preliminaries section, the degree of $f|_U$ is equal to the degree of $g : \mathbb{F}_2^k \to \mathbb{F}_2$ defined as $g(y_1, \ldots, y_k) = f(\sum_{i=1}^k y_i \Delta_i)$. Since $\deg(g) \leq d$, we may write $g(y) = \sum_{S \subseteq [k], |S| \leq d} a_S \cdot \prod_{i \in S} y_i$, where $a_S \in \mathbb{F}_2$ are constants. By Möbius inversion formula (Fact 14), $a_S = \sum_{T \subseteq S} g(\mathbf{1}_T)$. By the definition of g, we establish the relation $a_S = \sum_{T \subseteq S} f(\sum_{i \in T} \Delta_i) = f_S(0)$. Hence,

$$\begin{aligned}
\deg(f|_U) \leq d-1 \quad &\iff \quad \deg(g) \leq d-1 \\
&\iff \quad \forall S \subseteq [k] \quad \text{s.t. } |S| = d, a_S = 0 \\
&\iff \quad \forall S \subseteq [k] \quad \text{s.t. } |S| = d, f_S(0) = 0,
\end{aligned}$$

which completes the proof. ◀

Proof of Theorem 22. Similarly to the proof of Theorem 37, we find by induction basis vectors $\Delta_1, \ldots, \Delta_k$ for the subspace U. We assume by induction that $\deg(f|_U) \leq d-1$, and we wish to find a new vector Δ_{k+1}, linearly independent of $\Delta_1, \ldots, \Delta_k$, for which $\deg(f|_{U'}) \leq d-1$, where $U' = \text{span}\{\Delta_1, \ldots, \Delta_{k+1}\}$. We continue doing so as long as $\binom{k}{d-1} + k < n$.[12]

By Lemma 38, for any set $S \subseteq [k]$ of size d, $f_S(0) = 0$. We wish to find a new vector Δ_{k+1} such that for all $S \subseteq [k+1]$ of size d, $f_S(0) = 0$. It suffices to consider sets S of size d that contains $k+1$, since the correctness for all other sets is implied by the induction hypothesis.

For sets S of size $d-1$, $f_S(x)$ is an affine function and can be written as $f_S(x) = \langle \ell_S, x \rangle + c_S$, where $\ell_S \in \mathbb{F}_2^n$ and $c_S \in \mathbb{F}_2$. Let W be the linear subspace of $\mathbb{F}_2^n$ spanned by $\{\ell_S : S \subseteq [k], |S| = d-1\}$. Let Δ_{k+1} be any vector orthogonal to W, and linearly independent of $\Delta_1, \Delta_2, \ldots, \Delta_k$. Since, $\dim(W^\perp) = n - \dim(W) \geq n - \binom{k}{d-1}$, which by our

[12] Note that this is slightly better than the expression we had in Theorem 37.

assumption is strictly bigger than k, such a vector Δ_{k+1} exists. Let $S \subseteq [k+1]$ be a set of size d that contains $k+1$ and let $S' = S \cap [k]$, then

$$f_S(0) = f_{S'}(0) + f_{S'}(\Delta_{k+1}) = \langle \ell_{S'}, 0 \rangle + c_{S'} + \langle \ell_{S'}, \Delta_{k+1} \rangle + c_{S'} = 0 \ ,$$

where in the first equality we used the definitions of f_S and $f_{S'}$, and in the last equality we used the fact that Δ_{k+1} is orthogonal to $\ell_{S'}$. Using Lemma 38 we have shown that our choice of Δ_{k+1} gives a linear subspace $U' = \text{span}\{\Delta_1, \ldots, \Delta_{k+1}\}$ for which $f|_{U'}$ is of degree $\leq d-1$.

We now explain how to find, for any set S of size $d-1$, the affine function $f_S(x)$ (that is, ℓ_S and c_S) by performing $2^{d-1} \cdot (n+1)$ queries to f. As f_S is affine, knowing the values of f_S on the inputs $0, e_1, e_2, \ldots, e_n$ determines ℓ_S and c_S: $c_S = f_S(0)$ and $(\ell_S)_i = c_S + f_S(e_i)$ for $i \in [n]$. Each one of the values $f_S(0), f_S(e_1), \ldots, f_S(e_n)$ can be computed using 2^{d-1} queries to f, by the definition of f_S.

We now describe how can one efficiently find the vector Δ_{k+1} given $\Delta_1, \ldots, \Delta_k$. Using Gaussian elimination we find a basis for $W^{\perp}$. We check for each basis vector if it is not in the span of $\Delta_1, \ldots, \Delta_k$; after checking $k+1$ vectors we are promised to find such a vector. Next, we analyze the dimension of the subspace returned by the algorithm, the number of queries it makes to f, and the total running time.

Dimension of subspace

We abuse notation and denote by k the number of rounds in our algorithm, which is also the dimension of the subspace the algorithm returns. Since the algorithm stopped, we know that $\binom{k}{d-1} + k \geq n$. By a simple calculation, under the assumption that $d \leq \log(n)/3$ we get that $k = \Theta(d \cdot n^{1/(d-1)})$.

Number of queries

Overall through the k rounds of the algorithm we query f on all vectors of the form $v + \sum_{i \in T} \Delta_i$ for $v \in \{0, e_1, \ldots, e_n\}$ and $T \subseteq [k]$ of size $\leq d-1$. Hence, if we make sure not to query f more than once on the same point, the number of queries is $(n+1) \cdot \binom{k}{\leq d-1}$ which is at most $O(n^2)$ for $d \leq \log(n)/3$.

Running time

The total running time per round is $O(n^3)$ since we perform Gaussian elimination to calculate the basis for $W^{\perp}$, and another Gaussian elimination to check which of the first $k+1$ vectors of this basis is not in $\text{span}\{\Delta_1, \ldots, \Delta_{k+1}\}$. In addition, in each round we calculate the linear functions ℓ_S, but this only takes $O(n^2 \cdot 2^d)$ time, which is negligible compared to $O(n^3)$ under the assumption that $d \leq \log(n)/3$. Therefore, the total running time is $O(n^3 \cdot k)$. ◀

C Proof of DeMillo-Lipton-Schwartz-Zippel Variant

In this section we provide a proof for Lemma 13. Our proof is adapted from the proof of Lemma A.36 in the book of Arora and Barak [1].

Proof of Lemma 13. Since we only care about the values the polynomial take on $\mathbb{F}_q^n$, we may assume without loss of generality that the individual degree of each variable is at most $q-1$, since $a^q = a$ for all $a \in \mathbb{F}_q$.

We use induction on n. If $n = 1$ then f is a univariate polynomial of degree d for some $d \leq q-1$, since we assumed each individual degree is at most $q-1$. We have $\Pr[f(x_1) \neq 0] \geq 1 - d/q \geq q^{-d/(q-1)}$, where the first inequality follows since a univariate degree d polynomial over a field obtains at most d roots, and the last inequality can be verified for any $d \leq q-1$ using basic calculus. Suppose the statement is true when the number of variables is at most $n-1$. Then f can be written as

$$f(x_1, \ldots, x_n) = \sum_{i=0}^{\min(d,q-1)} x_1^i \cdot f_i(x_2, \ldots, x_n)$$

where f_i is of total degree at most $d-i$. Let k be the largest i such that f_i is a non-zero polynomial. By conditioning we have,

$$\Pr[f(x_1, \ldots, x_n) \neq 0] \geq \Pr[f_k(x_2, \ldots, x_n) \neq 0] \cdot \Pr[f(x_1, \ldots, x_n) \neq 0 \mid f_k(x_2, \ldots, x_n) \neq 0] \,.$$

By the induction hypothesis, the first multiplicand is at least $q^{-(d-k)/(q-1)}$. As for the second multiplicand, for any fixed $(x_2, \ldots, x_n) = (a_2, \ldots, a_n)$ such that $f_k(a_2, \ldots, a_n) \neq 0$, we get that $f(x_1, a_2, \ldots, a_n)$ is a non-zero univariate polynomial, in the variable x_1, of degree k. Hence, $\Pr_{x_1 \sim \mathbb{F}_q}[f(x_1, a_2, \ldots, a_n) \neq 0] \geq q^{-k/(q-1)}$ from the base case. Overall we get

$$\Pr[f(x_1, \ldots, x_n) \neq 0] \geq q^{-(d-k)/(q-1)} q^{-k/(q-1)} = q^{-d/(q-1)} \,. \qquad \blacktriangleleft$$