Electronic Colloquium on Computational Complexity, Report No. 145 (2013)

Two Structural Results for Low Degree Polynomials and Applications

Gil Cohen[*] Avishay Tal[†]

October 20, 2013

Abstract

In this paper, two structural results concerning low degree polynomials over the field $\mathbb{F}_2$ are given. The first states that for any degree d polynomial f in n variables, there exists a subspace of $\mathbb{F}_2^n$ with dimension $\Omega(n^{1/(d-1)})$ 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 $\Omega(n^{1/(d-1)})$. Using a recursive argument, we obtain our second structural result, showing that any degree d polynomial f induces a partition of $\mathbb{F}_2^n$ 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, and consider the algorithmic aspect of these results.

Our structural results have various applications:

- Dvir [CC 2012] introduced the notion of extractors for varieties, and gave explicit constructions of such extractors over large fields. We show that over $\mathbb{F}_2$, 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 [FOCS 2011] is a disperser for varieties over $\mathbb{F}_2$.
- Ben-Sasson and Kopparty [SIAM J. C 2012] proved that any degree 3 affine disperser is also an affine extractor with related parameters. Using our structural results, and based on the work of Kaufman and Lovett [FOCS 2008] and Haramaty and Shpilka [STOC 2010], we generalize this result to any constant degree.
- Implicit in Razborov's work [CAAML 1988], the existence of a depth 3 $\mathsf{AC}^0[\oplus]$ circuit that computes an optimal affine extractor was shown. We complement this result by showing that depth 2 $\mathsf{AC}^0[\oplus]$ circuits cannot compute affine dispersers for sub-polynomial dimension. This can be interpreted as a generalization of our structural results to sparse polynomials (regardless of their degree). We also give an alternative proof for the depth 3 case.

We deduce several other corollaries from the structural results, one of which states that any excellent affine extractor has small correlation with low degree polynomials. Another is a lower bound on the granularity of the Fourier spectrum of low degree polynomials.

[*] Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, ISRAEL. gil.cohen@weizmann.ac.il. Supported by an ISF grant and by the I-CORE Program of the Planning and Budgeting Committee.

[†] Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, ISRAEL. avishay.tal@weizmann.ac.il. Supported by an Adams Fellowship of the Israel Academy of Sciences and Humanities, by an ISF grant and by the I-CORE Program of the Planning and Budgeting Committee.

ISSN 1433-8092

Contents

1 Introduction

In this paper we consider the following question concerning polynomials on n variables over the field $\mathbb{F}_2$. What is the largest number $k = k(n,d)$ such that any degree d [1] polynomial is constant on some affine subspace of $\mathbb{F}_2^n$ with dimension k?

This question concerning the structure of low degree polynomials can be rephrased, in the language of pseudorandomness, as whether a low degree polynomial can be a good affine disperser. Recall that, over $\mathbb{F}_2$, an *affine disperser* for dimension k is a function $f \colon \mathbb{F}_2^n \to \mathbb{F}_2$ with the following property. For every affine subspace $u_0+U \subseteq \mathbb{F}_2^n$ of dimension k, f restricted to u_0+U is not constant. A function $f \colon \mathbb{F}_2^n \to \mathbb{F}_2$ is called an *affine extractor* for dimension k with bias ε, if for every affine subspace $u_0+U \subseteq \mathbb{F}_2^n$ of dimension k, it holds that

$$\operatorname{bias}(f|_{u_0+U}) \triangleq \left| \mathbb{E}_{u\sim u_0+U}\left[(-1)^{f(u)}\right] \right| \le \varepsilon.$$

It is worth mentioning that several explicit constructions of affine dispersers (and affine extractors) are in fact low degree polynomials [Bou07, BSG12, BSK12]. Examples for this fact can be found in the literature for other types of dispersers and extractors as well [CG88, BIW06, Dvi12].

Clearly, $k(n,1) = n-1$. The case $d=2$ is also well understood. By considering the inner product function $q(x_1,\dots,x_n) = x_1x_2 + x_3x_4 + \cdots + x_{n-1}x_n$, one obtains the upper bound $k(n,2) \le n/2$, which is tight as implied by Dickson's theorem [Dic01]. The extreme case $d=n$ boils down to the question of understanding the parameters of an optimal affine disperser, and it is not hard to show that $\log n - O(1) \le k(n,n) \le \log n + \log\log n + O(1)$.[2]

To the best of our knowledge, the value of $k(n,d)$ for $d>2$ has not received a formal treatment in the literature, although a variant of this natural question was previously raised by Trevisan [Tre06]. For biased polynomials, or for polynomials with large Gowers norm, one can obtain non-trivial lower bounds on $k(n,d)$ for $d=3,4$ as a corollary of the structural results of Haramaty and Shpilka [HS10]. Assuming low degree and bounded spectral norm, lower bounds on $k(n,d)$ follow by the structural result of [TWXZ13].

1.1 Our Results

In this paper we give an asymptotically tight upper and lower bounds on $k(n,d)$ for all d, and present several applications of this result to complexity theory and in particular to pseudorandomness.

Theorem 1.1 (Structural Result I)**.** *For all $d \ge 1$, $k(n,d)$ is bounded below by the least integer k such that*

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

In particular, there exists a universal constant $\alpha \in (0,1)$ such that $k(n,d) \ge \alpha\cdot n^{1/(d-1)}$ for all d. Moreover, if $d \le \log(n)/3$, then $k(n,d) \ge \alpha d\cdot n^{1/(d-1)}$.

Theorem 1.1 is tight – by applying a probabilistic argument, one can show that $k(n,d)$ is bounded above by the least integer k such that

$$\binom{k}{\le d} > n(k+1).$$

[1] Throughout this paper, by degree we mean total degree.

[2] In fact, we show in Appendix A that $k(n,(1+o(1))\log n)$ is at most $\log n + \log\log n + O(1)$.

Thus, we have an asymptotically matching upper bound on $k(n, d)$. [3] Based on the work of Ben-Eliezer *et al.* [BEHL09], one can say something even stronger regarding the tightness of Theorem 1.1. Namely, for every $d \geq 1$, there exists a degree d polynomial $f \colon \mathbb{F}_2^n \to \mathbb{F}_2$ with bias $2^{-\Omega(k/d)}$ on any affine subspace of dimension $k \geq \Omega(d \cdot n^{1/(d-1)})$. (see Section 3.3).

In the language of pseudorandomness, Theorem 1.1 states that a degree $d \leq \log{(n)}/3$ polynomial is not an affine disperser for dimension smaller than $\alpha d \cdot n^{1/(d-1)}$, and in particular, polynomials with constant degree are not affine dispersers for sub-polynomial dimension. The tightness result mentioned above, implies that there exists a degree d polynomial which is an affine extractor for dimension $k = O(d \cdot n^{1/(d-1)})$ with bias $2^{-\Omega(k/d)}$. In fact, most degree d polynomials share this property.

As a corollary from Theorem 1.1, we obtain a second structural result for low degree polynomials.

Theorem 1.2 (Structural Result II). *Let $f \colon \mathbb{F}_2^n \to \mathbb{F}_2$ be a degree d polynomial. Then, there exists a partition of $\mathbb{F}_2^n$ to affine subspaces (not necessarily shifts of the same subspace), each of dimension $\Omega(n^{1/(d-1)!})$, such that f is constant on each part.*

We do not know if the lower bound on the dimension in Theorem 1.2 is tight or not, and leave this as an open problem.

Generalization of the structural results to many polynomials. Being a natural generalization and also useful for some of our applications, we generalize the two structural results to the case of any number of polynomials (see Section 3.4). Let $f_1, \ldots, f_t \colon \mathbb{F}_2^n \to \mathbb{F}_2$ 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 3.7). In the second structural result, the promised dimension in Theorem 1.2 is replaced by $\Omega(n^{1/(d-1)!}/t^e)$, where e is the base of the natural logarithm (see Theorem 3.8).

The algorithmic aspect. We further study the algorithmic aspect of the structural results (see Section 4). We devise a $\mathrm{poly}(n)$-time algorithm (see Theorem 4.1), that given a degree d polynomial f on n variables as a black-box, performs $\mathrm{poly}(n)$ 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. We also give a $2^{o(n)}$-time algorithm that, for $d = o(\log n)$, outputs a subspace with an optimal dimension $\Omega(k(n, d))$ on which f is constant (see Theorem 4.5).

1.2 Applications

We now present several applications of our structural results.

Extractors and Dispersers for Varieties over $\mathbb{F}_2$

Let $\mathbb{F}$ be some 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 $\mathbb{F}$. An affine extractor over the field $\mathbb{F}$ is a function $f \colon \mathbb{F}^n \to \mathbb{F}$ that has small bias (defined appropriately) on every large enough affine subspace. In

[3] In fact, one can verify that the ratio between the upper and lower bounds we have on $k(n, d)$ is $1 + o_d(1)$ for all d.

[Dvi12], 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 $g_1, \dots, g_t$ in n variables over $\mathbb{F}$, we denote their variety by

$$\mathbf{V}(g_1, \dots, g_t) = \{x \in \mathbb{F}^n : g_1(x) = \cdots = g_t(x) = 0\}.$$

A function $f \colon \mathbb{F}^n \to \mathbb{F}$ as above is called an extractor for varieties.

In [Dvi12], 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.

In this paper we consider the extreme case of constructing extractors for varieties over the smallest field $\mathbb{F}_2$, which seems to be immune against algebraic geometry based techniques. We apply the generalization of Theorem 1.2 to many polynomials, and deduce a reduction from the problem of constructing extractors for varieties over $\mathbb{F}_2$ to the special case of constructing affine extractors.

Theorem 1.3. *Let $f \colon \mathbb{F}_2^n \to \mathbb{F}_2$ be an affine extractor for dimension $\Omega(n^{1/(d-1)!}/t^e)$ with bias ε. Then, f is an extractor with bias ε for varieties that are the common zeros of any t polynomials of degree at most d.*

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 5.1). 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 1.3.

The state of the art explicit constructions of affine extractors work only for dimension $\Omega(n/\sqrt{\log\log n})$ [Bou07, Yeh11, Li11], and thus the reduction in Theorem 1.3 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 1.3 also holds for dispersers (see Theorem 5.2), and an explicit construction of an affine disperser for dimension as small as $2^{\log^{0.9} n}$ is known [Sha11]. Thus, we obtain the first disperser for varieties over $\mathbb{F}_2$.

Theorem 1.4. *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 $t \le n^\alpha$ for some constant $\alpha < 1$, the requirement on the degree is $d < (1 - \alpha - o_n(1)) \cdot \log^{0.1} n$.*

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 [BSK12] proved (among other results) that any degree 3 affine disperser is also an affine extractor with comparable parameters. [4] Using the extension

[4] A reduction from extractors to dispersers in the context of two sources was also obtained, by Ben-Sasson and Zewi [BSZ11], conditioned on the well-known Freiman-Ruzsa conjecture from additive combinatorics.

of Theorem 1.1 to many polynomials, we are able to generalize the reduction of Ben-Sasson and Kopparty, over the field $\mathbb{F}_2$, to any degree $d \geq 3$.

Theorem 1.5. *For all $d \geq 3$ and $\delta > 0$, there exists $c = c(d, \delta)$ such that the following holds. Let $f \colon \mathbb{F}_2^n \to \mathbb{F}_2$ be an affine disperser for dimension k, which has degree d as a polynomial over $\mathbb{F}_2$. Then, f is also an affine extractor for dimension $k' \triangleq c \cdot k^{d-2}$ with bias δ.*

Note that Theorem 1.5 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 $O(d \cdot n^{1/(d-1)})$. On top of Theorem 1.1, the key ingredient we use in the proof of Theorem 1.5 is the work of Kaufman and Lovett [KL08] (see Section 6). For $d = 4$ we get a better dependency between k and k' based on the work of [HS10] (see Theorem 6.2).

$\mathsf{AC}^0[\oplus]$ 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 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 Claim A.1). 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 [Hås86], one can easily show that an AC^0 circuit on n inputs cannot compute an affine disperser for dimension $o(n/\mathrm{polylog}(n))$ (see Corollary 7.2). 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 [Hås98]. [5]

Somewhat surprisingly, we show that even depth 3 $\mathsf{AC}^0[\oplus]$ circuit can compute an optimal affine extractor. In fact, the same construction can be also realized by a polynomial-size De Morgan formula and has degree $(1 + o(1)) \log n$ as a polynomial over $\mathbb{F}_2$ (see Theorem A.6).

Theorem A.6 is implicit in the works of [Raz88, Sav95] who studied a similar problem in the context of Ramsey graphs (that is, two-source dispersers). We give a different proof in Appendix A, which can be extended to work also in the context of 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 [Juk12], Section 11.7). Moreover, any degree d polynomial $f \colon \mathbb{F}_2^n \to \mathbb{F}_2$ 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 1.6. *Let C be a depth 2 $\mathsf{AC}^0[\oplus]$ circuit on n inputs, with size n^d. Let $k < n/10 - d\log(n)$. If C computes an affine disperser for dimension k, then there exists a degree $2d$ 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 [RR94]. However, it does not necessarily get in the way of improving existing polynomial lower bounds.

Hence, by Theorem 1.2, the following theorem readily follows.

Theorem 1.7. *Let C be a depth 2 $\mathsf{AC}^0[\oplus]$ circuit on n inputs, with size n^d which is an affine disperser for dimension k, then $k > k(\sqrt{n}/5, 2d) = \Omega(n^{1/4d})$.*

We note that an $\mathsf{AC}^0[\oplus]$ circuit with size s on n inputs, can simulate a polynomial with s monomials (having no bound on the degree). Thus, Theorem 1.7 can be thought of as a generalization of Theorem 1.1 to sparse polynomials.

Good Affine Extractors are Hard to Approximate by Low Degree Polynomials

Using Theorem 1.2, 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 1.8. *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 1.2, 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,

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

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 1.8 implies that these extractors cannot be approximated by quadratic polynomials. Corollary 1.8 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 [Vio09] 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 1.2 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] In fact, Theorem 1.2 implies that *all* Fourier coefficients of a low degree polynomial has this granularity. To

[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 $k = \Omega(n^{1/(d-1)!})$.

see this, apply Theorem 1.2 to obtain a partition $P_1, \dots, P_\ell$ 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,

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

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 our results. We start with Theorem 1.1. Our proof is rather elementary, in spite of what one should expect considering previous works in this area, which apply machinery from additive combinatorics and Fourier analysis. Assume without loss of generality that $f(x) = 0$ for some $x \in \mathbb{F}_2^n$. 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 $f(x) = 0$. After selecting basis vectors $\Delta_1, \dots, \Delta_k$ for a subspace U, we consider all cosets $x + U$, restricted to which, f is constantly 0. We call such cosets *good*. If at least two good cosets exist, $x + U$ and $y + U$, then we can pick a new direction Δ_{k+1} to be $y - x$, and get that f is zero on $x + \operatorname{span}\{\Delta_1, \dots, \Delta_{k+1}\}$.

The main observation that allows us to derive Theorem 1.1 is the following. Given $\Delta_1, \dots, \Delta_k$, there exists a degree $D \sim \binom{k}{d-1}$ polynomial $t : \mathbb{F}_2^n \to \mathbb{F}_2$, such that $x + U$ is a good coset if and only if $t(x) = 0$. The Schwartz-Zippel lemma then assures us that if a single zero exists to a degree D polynomial then the polynomial has in fact many zeros – at least 2^{n-D}. 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 Schwartz-Zippel to claim that many other cosets are good as well. We can continue expanding our subspace U until $n \leq D$, which completes the proof.

The proof of the second structural result (Theorem 1.2) can be described informally as follows. Consider a degree d polynomial f. Theorem 1.1 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 Claim 3.4) 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 $\mathbb{F}_2^n$ to affine subspaces (not necessarily shifts of one another), such that f is constant on each one of them.

In fact, to prove Theorem 1.2, 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 4.1), we devise an algorithm that finds such an affine subspace and proceed similarly to the proof of Theorem 1.2. To obtain the second algorithmic result (Theorem 4.5), 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 Theorem 1.1 and Theorem 1.2 to more than one polynomial is quite straightforward.

2 Preliminaries

The set $\{1, \dots, n\}$ 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 $x, y \in \mathbb{F}_2^n$ we denote by $\langle x, y \rangle$

their scalar product over $\mathbb{F}_2$, i.e., $\langle x, y\rangle = \sum_{i=1}^{n} x_i \cdot y_i \mod 2$. 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. The degree of a Boolean function $f : \mathbb{F}_2^n \to \mathbb{F}_2$, denoted by $\deg(f)$, is the degree of the unique multivariate multi-linear polynomial over $\mathbb{F}_2$ which agrees with f on $\mathbb{F}_2^n$. We will abuse notation and interchange between a Boolean function and the unique multi-linear polynomial over $\mathbb{F}_2$ that agrees with f on $\mathbb{F}_2^n$.

The following folklore fact about polynomials over $\mathbb{F}_2$ is easy to verify.

Fact 2.1 (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: $a_S = \sum_{T\subseteq S} f(\mathbf{1}_T)$.*

Restriction to an affine subspace. Let $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$ be a Boolean function, $U \subseteq \mathbb{F}_2^n$ a subspace of dimension k and $u_0 \in \mathbb{F}_2^n$ some vector. We denote by $f|_{u_0+U} : (u_0+U) \to \mathbb{F}_2$ 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 $\mathbb{F}_2^n$ to $\mathbb{F}_2$) 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 $g : \mathbb{F}_2^k \to \mathbb{F}_2$ 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).

Definition 2.2 (Discrete Partial Derivative)**.** *For a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ and a direction $\Delta \in \mathbb{F}_2^n$, we define $\frac{\partial f}{\partial \Delta}(x) \triangleq f(x+\Delta) - f(x)$ to be the discrete partial derivative of f in direction Δ at the point x.*

Since addition and substraction are the same over $\mathbb{F}_2$, we may also write $\frac{\partial f}{\partial \Delta}(x) = f(x+\Delta) + f(x)$. If f is a polynomial of degree d over $\mathbb{F}_2$, then the degree of its partial derivative in direction Δ is at most $d-1$. Taking multiple derivatives in the directions $\Delta_1, \ldots, \Delta_k \in \mathbb{F}_2^n$ yields

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

which is symmetric with respect to the Δ_i's. It follows that $\frac{\partial^k f}{\partial \Delta_1 \ldots \partial \Delta_k}(x)$ is of degree at most $\deg(f) - k$ [8] as a polynomial over $\mathbb{F}_2$. Throughout the paper, if $\Delta_1, \ldots, \Delta_k \in \mathbb{F}_2^n$ are clear from the context, then for $S \subseteq [k]$, we denote

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

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.

[8] We consider the degree of the zero polynomial as $-\infty$.

3 Structural Results

We start this section by proving Theorem 1.1. In fact, we prove the following slightly stronger result.

3.1 Proof of Structural Result I

Theorem 3.1 (Structural Result I)**.** *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. In particular, $k(n,d) \geq \alpha \cdot n^{1/(d-1)}$, for some universal constant $\alpha \in (0,1)$. Moreover, for $d \leq \log(n)/3$ it holds that $k(n,d) \geq \alpha d \cdot n^{1/(d-1)}$.

The proof of Theorem 3.1 uses a folklore variant of the Schwartz-Zippel lemma for small fields.

Claim 3.2. *Let $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$ be a non-zero degree d polynomial. Then,*

$$\mathop{\mathbf{Pr}}_{x \sim \mathbb{F}_2^n} [f(x) \neq 0] \geq 2^{-d}.$$

Proof of Theorem 3.1. 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

$$n > k + \sum_{j=0}^{d-1} (d-j) \cdot \binom{k}{j} . \tag{3.1}$$

We assume by induction that there exists an affine subspace $u_0 + \mathrm{span}\{\Delta_1, \ldots, \Delta_k\} \subseteq \mathbb{F}_2^n$, where the Δ_i's are linearly independent vectors on which f evaluates to 0. Assuming Equation 3.1 holds, we show there exists a vector Δ_{k+1}, linearly independent of $\Delta_1, \ldots, \Delta_k$, such that $f \equiv 0$ on $u_0 + \mathrm{span}\{\Delta_1, \ldots, \Delta_{k+1}\}$. 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 we recall (see Preliminaries) that f_S is defined by

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

In particular, $\deg(f_S) \le d - |S|$. Thus $f_S \equiv 0$ for $|S| > d$, and we may write A as

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

Hence, A is the set of solutions to a system of $\binom{k}{\le d}$ 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. [9] One can also write A as the set of solutions to the single polynomial equation

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

which is of degree

$$D \le \sum_{j=0}^{d-1} (d-j)\cdot\binom{k}{j}\,.$$

Since A is non-empty, by Claim 3.2 we have that

$$|A| \ge 2^{n-D} \ge 2^{n-\sum_{j=0}^{d-1}(d-j)\cdot\binom{k}{j}}. \tag{3.2}$$

This, together with Equation (3.1) implies that $|A| > 2^k$. Hence, there exists a point $y \in A$ such that $y - u_0 \notin \operatorname{span}\{\Delta_1, \Delta_2, \ldots, \Delta_k\}$. Pick such a point y arbitrarily and denote by $\Delta_{k+1} \triangleq y - u_0$. Since both u_0 and y are in A we have that $f \equiv 0$ on

$$u_0 + \operatorname{span}\{\Delta_1, \ldots, \Delta_{k+1}\}\,.$$

The inductive proof shows that there exists a subspace U of dimension k such that f is constant on $u_0 + U$ and

$$n \le k + \sum_{j=0}^{d-1} (d-j)\cdot\binom{k}{j}\,, \tag{3.3}$$

since otherwise we could have continue this process and pick a bigger subspace U'. We now complete the proof by showing that $k = \Omega(d\cdot n^{1/(d-1)})$ for $d \le \log(n)/3$ and $k = \Omega(n^{1/(d-1)})$ for any d. The right hand side of Equation (3.3) is bounded above by $d \cdot 2^k$, hence $k \ge \log(n/d)$. Under the assumption $d \le \log(n)/3$ we get $k \ge 2\log(n)/3 \ge 2d$. We return to Equation (3.3) and deduce that

$$n \le k + d\cdot\sum_{j=0}^{d-1}\binom{k}{j} \underset{2d\le k}{\le} (d^2+1)\cdot\binom{k}{d-1} \le (d^2+1)\cdot\left(\frac{ke}{d-1}\right)^{d-1}$$

and so

$$k \ge \left(\frac{n}{d^2+1}\right)^{\frac{1}{d-1}} \cdot \frac{d-1}{e} > \frac{1}{28}\cdot d\cdot n^{1/(d-1)}\,.$$

For $d \ge \log(n)/3$ the proof follows since $n^{1/(d-1)} \le 64$. □

[9] 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.

3.2 Proof of Structural Result II

In this section we prove the following theorem, which is a slightly more formal restatement of Theorem 1.2.

Theorem 3.3 (Structural Result II). *There exists a constant $\alpha' \in (0,1)$ such that the following holds. Let $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$ be a degree d polynomial, then, there exists a partition of $\mathbb{F}_2^n$ to affine subspaces, each of dimension $\alpha' \cdot n^{1/(d-1)!}$, such that f is constant on each part.*

We use the following claim for the proof of Theorem 3.3.

Claim 3.4. *Let $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$ be a degree d polynomial. Assume there exists an affine subspace $u_0 + U$ of dimension k, 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 $u_1 \in \mathbb{F}_2^n$. Now, for any $u \in U$

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

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 3.3. Let α be the constant from Theorem 3.1. Define the sequence $\{\beta_d\}_{d=1}^{\infty}$ as follows.

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

We will prove by induction on d, the degree of a given polynomial f, that there exists a partition of $\mathbb{F}_2^n$ 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 $d \geq 1$,

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

and thus one can take $\alpha' = \alpha^e/2$ 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 $\mathbb{F}_2^n$ to two 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 3.1 and Claim 3.4, there exists a partition of $\mathbb{F}_2^n$ to affine subspaces of dimension $k \geq \alpha \cdot n^{1/(d-1)}$, 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 $d' \leq d-1$. [10] 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

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

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. □

[10] We may apply the induction because there exists a linear bijection from U to $\mathbb{F}_2^{\dim U}$. More precisely, if A is an $n \times k$ matrix over $\mathbb{F}_2$ that maps U to $\mathbb{F}_2^k$ 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 $\mathbb{F}_2^k$ 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

Theorem 3.5. *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 3.5 we apply the following lemma due to Ben-Eliezer, Hod and Lovett [BEHL09].

Lemma 3.6 ([BEHL09], Lemma 2)**.** *Fix $\varepsilon > 0$ and let f be a random degree d polynomial [11] 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 3.5. 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 3.6,

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

where c_1, c_2 are the constants from Lemma 3.6 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 3.7 (Structural Result I for many polynomials)**.** *Let $f_1, \ldots, f_t\colon \mathbb{F}_2^n \to \mathbb{F}_2$ be polynomials of degree $d_1, \ldots, d_t$ respectively. Let k be the least integer satisfying the inequality*

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

Then, for every $u_0 \in \mathbb{F}_2^n$ there exists a subspace $U \subset \mathbb{F}_2^n$ 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 $d \leq \log(n)/3$, $k = \Omega(d \cdot (n/t)^{1/(d-1)})$.

Proof. The proof is very similar to that of Theorem 3.1, so we only highlight the differences. As in the proof of Theorem 3.1, 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:

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

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

As in the proof of Theorem 3.1, A can be written as the set of solutions to a single polynomial equation of degree

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

and is non-empty as it contains u_0. By Claim 3.2, $|A| \geq 2^{n-D}$, and we can choose a new linearly independent Δ_{k+1} as long as $n - D > k$, which completes the proof. □

Similarly to the way we deduced Theorem 3.3 from Theorem 3.1, one can deduce the following theorem from Theorem 3.7. We omit the proof.

Theorem 3.8 (Structural Result II for many polynomials)**.** *Let $f_1, \ldots, f_t \colon \mathbb{F}_2^n \to \mathbb{F}_2$ be polynomials of degree at most d. Then, there exists a partition of $\mathbb{F}_2^n$ 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.*

4 The Algorithmic Aspect

4.1 Efficient Algorithm for Finding a Somewhat Large Subspace

Theorem 4.1. *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 $\deg(f|_U) \leq d - 1$.*

We obtain the following corollaries.

Corollary 4.2. *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.*

Corollary 4.3. *Let $f \colon \mathbb{F}_2^n \to \mathbb{F}_2$ be a degree d polynomial given as a black-box, then there is a $2^{n-k} \cdot \mathrm{poly}(n)$-time $\mathrm{poly}(n)$-space algorithm, which partitions $\mathbb{F}_2^n$ 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 4.3. The proof of Theorem 4.1 uses the following lemma.

Lemma 4.4. *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.*

Proof of Lemma 4.4. 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 2.1), $a_S = \sum_{T \subseteq S} g(\mathbf{1}_T)$. 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,

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

which completes the proof. □

Proof of Theorem 4.1. Similarly to the proof of Theorem 3.1, we find by induction basis vectors $\Delta_1, \dots, \Delta_k$ for the subspace U. We assume by induction that $\deg(f|_U) \le d-1$, and we wish to find a new vector Δ_{k+1}, linearly independent of $\Delta_1, \dots, \Delta_k$, for which $\deg(f|_{U'}) \le d-1$, where $U' = \mathrm{span}\{\Delta_1, \dots, \Delta_{k+1}\}$. We continue doing so as long as $\binom{k}{d-1} + k < n$.[12]

By Lemma 4.4, 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 $\Delta_1, \Delta_2, \dots, \Delta_k$. Since, $\mathsf{dim}(W^\perp) = n - \mathsf{dim}(W) \ge n - \binom{k}{d-1}$, which by our 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 4.4 we have shown that our choice of Δ_{k+1} gives a linear subspace $U' = \mathrm{span}\{\Delta_1, \dots, \Delta_{k+1}\}$ for which $f|_{U'}$ is of degree $\le d-1$.

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 $0, e_1, e_2, \dots, e_n$ 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 $f_S(0), f_S(e_1), \dots, f_S(e_n)$ 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 $\Delta_1, \dots, \Delta_k$. Using Gaussian elimination we find a basis for $W^\perp$. We check for each basis vector if it is not in the span of $\Delta_1, \dots, \Delta_k$; 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 $\binom{k}{d-1} + k \ge n$. By a simple calculation, under the assumption that $d \le \log(n)/3$ 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 $v \in \{0, e_1, \dots, e_n\}$ and $T \subseteq [k]$ of size $\le d-1$. Hence, if we make sure not to query f more than once on the same point, the number of queries is $(n+1) \cdot \binom{k}{\le d-1}$ which is at most $O(n^2)$ for $d \le \log(n)/3$.

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 $\mathrm{span}\{\Delta_1, \dots, \Delta_{k+1}\}$. 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 $d \le \log(n)/3$. Therefore, the total running time is $O(n^3 \cdot k)$. □

[12] Note that this is slightly better than the expression we had in Theorem 3.1.

4.2 Subexponential-Time Algorithm for Finding an Optimal Subspace

Theorem 4.5. *There exist constants $\alpha \in (0,1)$ and $\beta > 0$ such that the following holds. There exists an algorithm that given $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$, a degree d polynomial (as a list of monomials), where $3 \le d \le \log(n)/3$, and $u_0 \in \mathbb{F}_2^n$ as inputs, finds an affine subspace $u_0 + U$ of dimension $\Omega(k(n,d))$, restricted to which f is constant. The running time of the algorithm is $2^{\beta \cdot n^{(d-2)/(d-1)}} \cdot \mathrm{poly}(n^d)$ time, and it uses $\mathrm{poly}(n^d)$ space.*

Proof. We follow the proof of Theorem 3.1. Again, we may assume $f(u_0) = 0$. Given the previously chosen vectors $\Delta_1, \dots, \Delta_k$ such that f is the constant 0 on $u_0 + \mathrm{span}\{\Delta_1, \dots, \Delta_k\}$, we show how to find a new vector Δ_{k+1} which is linearly independent of $\Delta_1, \dots, \Delta_k$, such that f is constantly zero on $u_0 + \mathrm{span}\{\Delta_1, \dots, \Delta_{k+1}\}$. The set A is the set of solutions to the following set of polynomial equations:

$$\{f_S(x) = 0 \ : \ S \subseteq [k], |S| \le d-1\} \ ,$$

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_S in $\mathrm{poly}(n^d)$ time. Given y which is a solution to this set of equations, if $\Delta_{k+1} := y - u_0$ is linearly independent of $\Delta_1, \dots, \Delta_k$ then we have achieved our goal. It is therefore enough to find more than 2^k different solutions to this set of equations, in order to guarantee that for one of them $y - u_0$ 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:

$$L = \{f_S(x) = 0 \ : \ S \subseteq [k],\ |S| \le d-1,\ \deg(f_S) = 1\};$$
$$NL = \{f_S(x) = 0 \ : \ S \subseteq [k],\ |S| \le d-1,\ \deg(f_S) > 1\}.$$

Let $m = \sum_{f_S \in NL} \deg(f_S)$. Since we know u_0 is a solution to all equations in $L \cup NL$, we can impose new linear equations which hold for u_0, 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 $\{x_i = (u_0)_i\}$ to L' until $\mathsf{dim}(L') = n - m - k - 1$. [13]

The set of solutions to both L' and NL is non-empty as it contains u_0. Furthermore, the sum of the degrees of equations in $L' \cup NL$ is exactly $(n - m - k - 1) + m = n - k - 1$. Therefore, by Claim 3.2, there are at least 2^{k+1} solutions to the equations in $L' \cup NL$, which guarantees that one of the solutions will yield a new vector Δ_{k+1}, linearly independent of $\Delta_1, \dots, \Delta_k$.

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 x in this affine subspace we verify that all the equations in NL are satisfied by x. The running time of this process is $O(2^{n - \mathsf{dim}(L')} \cdot |NL| \cdot n^d)$, which is $O(2^{m+k} \cdot n \cdot n^d)$.

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

[13] 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.

5 Extractors and Dispersers for Varieties over $\mathbb{F}_2$

We start this section by proving Theorem 1.3.

Proof of Theorem 1.3. Let $g_1, \ldots, g_t \colon \mathbb{F}_2^n \to \mathbb{F}_2$ be degree d polynomials. By Theorem 3.8, there exists a partition of $\mathbb{F}_2^n$ 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

$$\left| \mathop{\mathbf{E}}_{x \sim P_i} \left[(-1)^{f(x)} \right] \right| \le \varepsilon. \tag{5.1}$$

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 $\mathbb{F}_2^n$ to $P_1, \ldots, P_\ell$. Then,

$$\left| \mathop{\mathbf{E}}_{x \sim \mathbf{V}(g_1,\ldots,g_t)} \left[(-1)^{f(x)} \right] \right| = \left| \mathop{\mathbf{E}}_{i \sim I} \mathop{\mathbf{E}}_{x \sim P_i} \left[(-1)^{f(x)} \right] \right| \le \mathop{\mathbf{E}}_{i \sim I} \left| \mathop{\mathbf{E}}_{x \sim P_i} \left[(-1)^{f(x)} \right] \right| \le \varepsilon,$$

where the last inequality follows by Equation (5.1). □

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.

Theorem 5.1. *For every $d \in \mathbb{N}$ and $\delta, \rho \in (0,1)$ the following holds. Let $f \colon \mathbb{F}_2^n \to \mathbb{F}_2$ be an affine extractor for dimension $\Omega(n^{1/(d-1)!}/\ell^e)$ with bias ε, where $\ell = \log(2/(\rho\delta))$. Then, f is an extractor with bias $\varepsilon + \delta$ for varieties with density at least ρ (i.e., size at least $\rho \cdot 2^n$), that are the common zeros of any degree (at most) d polynomials.*

Proof. Let $g_1, \ldots, g_t \colon \mathbb{F}_2^n \to \mathbb{F}_2$ be degree (at most) d polynomials. First, we prove the existence of ℓ polynomials $h_1, \ldots, h_\ell \colon \mathbb{F}_2^n \to \mathbb{F}_2$, each of degree at most d, with a variety that approximates $\mathbf{V}(g_1, \ldots, g_t)$. More precisely,

$$\mathbf{V}(g_1, \ldots, g_t) \subseteq \mathbf{V}(h_1, \ldots, h_\ell) \quad \text{and} \quad \mathop{\mathbf{Pr}}_{x \sim \mathbb{F}_2^n} [x \in \mathbf{V}(h_1, \ldots, h_\ell) \setminus \mathbf{V}(g_1, \ldots, g_t)] \le 2^{-\ell}, \tag{5.2}$$

The proof of this claim follows by a standard argument, like the one that appears in [Raz87, Smo87]. Let $S_1, \ldots, S_\ell$ be random sets of $[t]$ defined as follows. For each $i \in [\ell]$, independently, an element of $[t]$ is contained in S_i with probability $1/2$, independently of all other elements in $[t]$. For each $i \in [\ell]$, define the (random) polynomial

$$H_i(x) = \sum_{j \in S_i} g_j(x),$$

where the summation is taken over $\mathbb{F}_2$. 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 $S_1, \ldots, S_\ell$). Otherwise, for each $i \in [\ell]$, $\mathbf{Pr}[H_i(x) = 0] = 1/2$. By an averaging argument, one can fix $S_1, \ldots, S_\ell$ and obtain fixed polynomials $h_1, \ldots, h_\ell$, of degree at most d, that satisfy the conditions in Equation (5.2).

Since f is an affine extractor with bias ε for dimension $\Omega(n^{1/(d-1)!}/\ell^e)$, we can apply Theorem 1.3 and get that f is an extractor with bias ε for $V(h_1,\dots,h_\ell)$. Therefore,

$$\begin{aligned}
\varepsilon \cdot |\mathbf{V}(h_1,\dots,h_\ell)| &\geq \left| \sum_{x\in\mathbf{V}(h_1,\dots,h_\ell)} (-1)^{f(x)} \right| \\
&\geq \left| \sum_{x\in\mathbf{V}(g_1,\dots,g_t)} (-1)^{f(x)} \right| - \left| \sum_{x\in\mathbf{V}(h_1,\dots,h_\ell)\setminus\mathbf{V}(g_1,\dots,g_t)} (-1)^{f(x)} \right| \\
&\geq \left| \sum_{x\in\mathbf{V}(g_1,\dots,g_t)} (-1)^{f(x)} \right| - 2^{n-\ell},
\end{aligned}$$

and so

$$\mathop{\mathbf{E}}_{x\sim\mathbf{V}(g_1,\dots,g_t)} \left[(-1)^{f(x)}\right] \leq \frac{\varepsilon\cdot|\mathbf{V}(h_1,\dots,h_\ell)| + 2^{n-\ell}}{|\mathbf{V}(g_1,\dots,g_t)|} \leq \varepsilon + 2\cdot\frac{2^{-\ell}}{\rho}.$$

By our choice of ℓ, f has bias at most $\varepsilon+\delta$ on $\mathbf{V}(g_1,\dots,g_t)$, as claimed. □

The following theorem states an analog reduction from dispersers for varieties to affine dispersers.

Theorem 5.2. *Let $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$ be an affine disperser for dimension $\Omega((n/t)^{1/(d-1)})$. Then, f is a disperser for varieties that are the common zeros of any t polynomials of degree at most d.*

Proof. Let $g_1,\dots,g_t\colon \mathbb{F}_2^n \to \mathbb{F}_2$ be degree (at most) d polynomials. Let $u_0 \in \mathbf{V}(g_1,\dots,g_t)$ (if $\mathbf{V}(g_1,\dots,g_t) = \emptyset$, there is nothing to prove). By Theorem 3.7, there exists a subspace U of dimension $\Omega((n/t)^{1/(d-1)})$ such that $u_0 + U \subseteq \mathbf{V}(g_1,\dots,g_t)$. The proof then follows as f is an affine disperser for dimension $\Omega((n/t)^{1/(d-1)})$. □

In [Sha11] (Theorem 1.2), an explicit construction of an affine disperser $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$ for dimension $2^{\log^{0.9} n}$ is given. Theorem 1.4 readily follows by Theorem 5.2 as indeed, one only needs to make sure that $(n/t)^{1/(d-1)} = \Omega(2^{\log^{0.9} n})$.

6 From Affine Dispersers to Affine Extractors

To prove Theorem 1.5, we use the following theorem of Kaufman and Lovett [KL08].

Theorem 6.1 ([KL08])**.** *Let $f\colon \mathbb{F}^n \to \mathbb{F}$ be a degree (at most) d polynomial with* $\mathrm{bias}(f) \geq \delta$*. Then, there exist $c = c(d,\delta)$ polynomials $f_1,\dots,f_c$ of degree at most $d-1$ such that $f = G(f_1,\dots,f_c)$, for some function $G : \mathbb{F}^c \to \mathbb{F}$.*

Proof of Theorem 1.5. 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 $f : \mathbb{F}_2^n \to \mathbb{F}_2$ 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 $\mathsf{dim}(U) = k'$ such that $\mathrm{bias}(f|_{u_0+U}) > \delta$. Let $u_1,\dots,u_{k'}$ be a basis for U and let $g : \mathbb{F}_2^{k'} \to \mathbb{F}_2$ be the function defined by $g(y_1,\dots,y_{k'}) = f(u_0 + \sum_{i=1}^{k'} u_i \cdot y_i)$. Then, g is a biased polynomial of degree $\leq d$. Applying Theorem 6.1 to g, we can write it as $G(g_1,\dots,g_c)$, where the g_i's are of degree at most $d-1$, and $c = c(d,\delta)$ as defined in Theorem 6.1.

By Theorem 3.7, there is an affine subspace W of $\mathbb{F}_2^{k'}$ with dimension $\alpha \cdot (k'/c)^{1/(d-2)}$ for which all the g_i's are constant, for some constant $\alpha > 0$. In particular $g|_W$ is constant, which implies that there exists a subspace of $\mathbb{F}_2^n$, with the same dimension, on which the original function f is constant. Taking $k' = k^{d-2} \cdot \frac{c(d,\delta)}{\alpha^{d-2}}$ completes the proof. □

For degree 3 and 4, we rely on stronger results from [HS10]. Although degree 3 was treated in [BSK12], we present it here for completeness.

Theorem 6.2. *Let $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$ 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 1.5, it is enough to show that if g is a degree 3 or 4 polynomial over $\mathbb{F}_2$ 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 [HS10], 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 Claim 3.4, 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 [Dic01], 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 $\mathsf{dim}(W')$ is at least k.

Quartic, ($\deg(g) = 4$). Theorem 4 in [HS10] 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 3.7, 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 Claim 3.4, 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 $\mathrm{bias}(g) \geq \delta$, by averaging there is a coset on which $\mathrm{bias}(g|_{w+W}) \geq \delta$. 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 $k' = k \cdot \mathrm{poly}(1/\delta)$ ensures that the dimension of W' is at least k. □

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 3.1.

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

The next lemma, following Håstad's work [Hås86], appears in [BS90].

Lemma 7.1 ([BS90], Corollary 3.7, restated)**.** *Let $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$ 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 7.1 readily implies the following corollary.

Corollary 7.2. *Let $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$ 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 1.7, one only needs to prove Lemma 1.6.

Proof of Lemma 1.6. 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 $\mathsf{OR}, \mathsf{AND}$ gates to have fan-in smaller than $2d$ 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 $2d$ after the restriction. Let m be the fan-in of the AND gate before the restriction, and m' its fan-in afterwards. We have

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

Since $2p$ is smaller than 1, the right hand side of the above inequality is at most $(1/2)^m \cdot 2^m \cdot (2p)^{2d} = (2p)^{2d}$. Thus, $\mathbf{Pr}[m' \geq 2d] \leq (2p)^{2d}$. By our choice of parameter p, this is at most $1/(4n)^d$. By union bound over all $\leq n^d$ AND and OR gates, with probability at least $1 - 1/4^d \geq 3/4$ over the random restrictions, the fan-in of all AND and OR gates, under the restriction, is smaller than $2d$. 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 $2d$. Expressing the resulting circuit as a polynomial over

$\mathbb{F}_2$ we get a polynomial on $\sqrt{n}/5$ variables with degree at most $2d$ 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 $s = n^d$. However, under the assumption that $k < n/10 - \log(s)$ this is a contradiction to Corollary 7.2.

□

Acknowledgement

We wish to thank our advisor Ran Raz for many helpful discussions and for his encouragement. We thank Elad Haramaty and Amir Shpilka for helpful discussions regarding this work. We thank the user goes by the name david from stack exchange for pointing out [BEHL09].

References

[BEHL09] 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.

[BIW06] B. Barak, R. Impagliazzo, and A. Wigderson. Extracting randomness using few independent sources. *SIAM Journal on Computing*, 36(4):1095–1118, 2006.

[Bou07] J. Bourgain. On the construction of affine extractors. *GAFA Geometric And Functional Analysis*, 17(1):33–57, 2007.

[BS90] 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.

[BSG12] E. Ben-Sasson and A. Gabizon. Extractors for polynomials sources over constant-size fields of small characteristic. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 399–410. Springer, 2012.

[BSK12] E. Ben-Sasson and S. Kopparty. Affine dispersers from subspace polynomials. *SIAM Journal on Computing*, 41(4):880–914, 2012.

[BSZ11] 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.

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

[Dic01] 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.

[Dvi12] Z. Dvir. Extractors for varieties. *computational complexity*, 21(4):515–572, 2012.

[Hås86] J. Håstad. Almost optimal lower bounds for small depth circuits. In *STOC*, pages 6–20, 1986.

[Hås98] J. Håstad. The shrinkage exponent of de Morgan formulas is 2. *SIAM Journal on Computing*, 27(1):48–64, 1998.

[HS10] 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.

[Juk12] S. Jukna. *Boolean function complexity: advances and frontiers*, volume 27. Springerverlag Berlin Heidelberg, 2012.

[KL08] 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.

[Li11] 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.

[Raz87] 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.

[Raz88] 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.

[RR94] 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.

[Sav95] P. Savickỳ. Improved Boolean formulas for the Ramsey graphs. *Random Structures & Algorithms*, 6(4):407–415, 1995.

[Sha11] 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.

[Smo87] 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.

[Tre06] L. Trevisan, 2006. http://in-theory.blogspot.co.il/2006/06/polynomials-and-subspaces.html.

[TWXZ13] H. Y. Tsang, C. H. Wong, N. Xie, and S. Zhang. Fourier sparsity, spectral norm, and the log-rank conjecture. *arXiv preprint arXiv:1304.1245*, 2013.

[Vio09] E. Viola. Guest column: correlation bounds for polynomials over {0,1}. *ACM SIGACT News*, 40(1):27–44, 2009.

[Yeh11] A. Yehudayoff. Affine extractors over prime fields. *Combinatorica*, 31(2):245–256, 2011.

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

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

Claim A.1. *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 Claim A.1 makes use of Hoeffding bound.

Theorem A.2 (Hoeffding Bound)**.** *Let $X_1, \dots, X_n$ 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,*

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

Proof of Claim A.1. 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 A.2),

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

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 A.6, we introduce the following notion.

Definition A.3. *An (n,k,d)* linear injector *with size m is a family of $d \times n$ matrices $\{A_1, \dots, A_m\}$ 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 A.4. *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 $\mathbf{Pr}[Au = 0] = 2^{-d}$. By taking the union bound over all elements in $U \setminus \{0\}$, we get that

$$\mathbf{Pr}[\ker(A) \cap U \neq \{0\}] \leq 2^{k-d}.$$

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

$$\mathbf{Pr}[\forall i \in [m] \;\; \ker(A_i) \cap U \neq \{0\}] \leq 2^{m(k-d)}.$$

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 A.5. *Let n_0, c be the constants from Claim A.1. 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*

$$f(x) = \bigoplus_{i=1}^{m} f_i(A_i x) \tag{A.1}$$

is an affine extractor for dimension k with bias ε.

Proof. Recall that in the proof of Claim A.1, 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

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

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 A.6. *Let f be the function from Equation (A.1), where $\{A_1, \ldots, A_m\}$ is the (n, k, d) linear injector from Lemma A.4 (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 $d = k+1$, $\mathsf{size}(C) = O((n/\varepsilon)^2 \cdot \log^3(n/\varepsilon))$ 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. □