Electronic Colloquium on Computational Complexity, Report No. 79 (2019)

Average Bias and Polynomial Sources

Arnab Bhattacharyya[*] Philips George John[†] Suprovat Ghoshal[‡] Raghu Meka[§]

Abstract

We identify a new notion of pseudorandomness for randomness sources, which we call the *average bias*. Given a distribution Z over $\{0,1\}^n$, its average bias is:

$$b_{\text{av}}(Z) = 2^{-n} \sum_{c \in \{0,1\}^n} \left| \mathop{\mathbb{E}}_{z \sim Z} (-1)^{\langle c,z \rangle} \right|$$

A source with average bias at most 2^{-k} has min-entropy at least k, and so low average bias is a stronger condition than high min-entropy. We observe that the inner product function is an extractor for any source with average bias less than $2^{-n/2}$.

The notion of average bias especially makes sense for *polynomial sources*, i.e., distributions sampled by low-degree n-variate polynomials over $\mathbb{F}_2$. For the well-studied case of affine sources, it is easy to see that min-entropy k is exactly equivalent to average bias of 2^{-k}. We show that for quadratic sources, min-entropy k implies that the average bias is at most $2^{-\Omega(\sqrt{k})}$. We use this relation to design dispersers for *separable* quadratic sources with a min-entropy guarantee.

1 Introduction

Given a source (distribution) Z, the problem of designing an *extractor* for Z is that of constructing a function which when applied to a sample from Z makes the output be close to a random bit string. Explicitly constructing extractors is a fundamental object of study in theoretical computer science and has wide-ranging applications to a variety of areas, including complexity theory [Zuc96, Tre01], data structures [BMRV02, TS02], coding theory [TSZ04], hashing [GW97], cryptography [Vad04], graph theory [WZ99], and geometry [Ind07]. See the book by Vadhan [Vad12] and references therein for an overview of randomness extraction.

1.1 Average Bias

Formally, given a family $\mathcal{F}$ of sources over $\{0,1\}^n$, an *ε-extractor* for $\mathcal{F}$ is a function $E : \{0,1\}^n \to \{0,1\}^m$ such that for any source Z in $\mathcal{F}$, the random variable $E(Z)$ is ε-close in statistical distance to the uniform distribution over $\{0,1\}^m$. Throughout this paper, we will restrict ourselves to one-bit extractors, i.e., $m = 1$. It is well-known that in order to extract a bit, the source must[1] have

[*] National University of Singapore & Indian Institute of Science. E-mail: arnabb@nus.edu.sg

[†] IBM Research. Work partly completed while at IISc. E-mail: pgeorg04@in.ibm.com

[‡] Indian Institute of Science. E-mail: suprovat@iisc.ac.in

[§] University of California, Los Angeles. E-mail: raghum@cs.ucla.edu

[1] To be precise, the source must be *ε-close* to one having min-entropy ≥ 1 but let's ignore this technicality for now.

ISSN 1433-8092

min-entropy at least 1, where the min-entropy is defined as:

$$H_\infty(Z) = -\log_2 \max_{z\in\{0,1\}^n} \Pr[Z = z].$$

However, the min-entropy condition is far from sufficient. In fact, no extractor exists even for the family of sources of min-entropy $n-1$!

In this work, we identify another notion of pseudorandomness which is also necessary for extraction, namely that of average bias, which we define as follows.

Definition 1.1 (Average Bias) *Given a source Z over $\{0,1\}^n$, its* average bias *is:*

$$b_{\text{av}}(Z) = 2^{-n} \sum_{c\in\{0,1\}^n} \left| \mathop{\mathbb{E}}_{z\sim Z} (-1)^{\langle c,z\rangle} \right|.$$

It is clear that $b_{\text{av}}(Z) < 1$ in order for Z to admit an extractor, because otherwise if $b_{\text{av}}(Z) = 1$, Z must be constant. The name "average bias" derives from the well-studied notion of *bias* of a distribution:

$$\text{bias}(Z) = \left| \mathop{\mathbb{E}}_{z\sim Z} [(-1)^z] \right|.$$

To the best of our knowledge, the notion of average bias has not been systematically studied previously. Our motivation originated from additive combinatorics where bias has played an important role in linking analytic and algebraic properties of functions [KL08, GT09].

We begin by observing that low average bias is a stronger notion of pseudorandomness than high min-entropy. Precisely:

Lemma 1.1 *For a distribution Z on $\{0,1\}^n$, if $b_{\text{av}}(Z) \le 2^{-k}$, then $H_\infty(Z) \ge k$.*

While it is true, as mentioned above, that one cannot extract from all sources of min-entropy $n-1$, one can extract from all sources with a small average bias guarantee. This is our second observation (see Section 2 for definitions).

Lemma 1.2 *Any bent function on n bits is an ε-extractor for sources with average bias $< 2^{-(1+\varepsilon)n/2}$.*

We also observe that it is impossible to extract from arbitrary sources with larger average bias.

1.2 Polynomial Sources

Lemma 1.2 indicates that for general sources, small average bias can be a much stronger guarantee than large min-entropy. However, we now show that for a class of structured sources, average bias and min-entropy give qualitatively similar guarantees, and we make use of this result to extract randomness from such sources when there is only a min-entropy lower bound.

A well-studied class of sources studied previously is the set of affine sources. An *affine source* of min-entropy k is a random variable that is uniformly distributed on some k-dimensional affine subspace of $\mathbb{F}_2^n$. Another way to describe the distribution is by defining n affine functions on k variables, $a_1(t_1,\dots,t_k),\dots,a_n(t_1,\dots,t_k)$, that have rank k. Now, the affine source is interpreted as the output of $(a_1(t),\dots,a_n(t))$ on a uniformly chosen input $t\in\mathbb{F}_2^k$. The following is easy to check[2]:

[2] The bias of an affine function is 0 if it is not constant and 1 otherwise.

Fact 1.3 *An affine source of min-entropy k has average bias 2^{-k}.*

Hence, Lemma 1.1 is exactly tight for affine sources. Our main technical contribution is showing a qualitatively similar result for *quadratic sources*.

Definition 1.4 (Quadratic Sources) *A* quadratic source *over $\mathbb{F}_2^n$ is generated by a function $P = (P_1, \ldots, P_n)$ where each $P_i \in \mathbb{F}_2[x_1, \ldots, x_m]$ is a polynomial of degree ≤ 2 (for some integer $m \geq 1$). The source is defined as the random variable $P(X) = (P_1(X), \ldots, P_n(X))$ where X is uniformly chosen from $\mathbb{F}_2^m$. By slight abuse of notation, we let $H_\infty(P)$ and $b_{\mathrm{av}}(P)$ denote the min-entropy and average bias respectively of the source generated by P.*

To the best of our knowledge, there are no known explicit extractors for quadratic sources over $\mathbb{F}_2^n$ with any non-trivial min-entropy guarantee. Previous work on quadratic, and more generally, polynomial sources have been over $\mathbb{F}^n$ for large fields $\mathbb{F}$, as we discuss later in Section 1.3.

We show the following relation between average bias and min-entropy for quadratic sources:

Theorem 1.1 *Let $P = (P_1, \ldots, P_n)$ generate a quadratic source over $\mathbb{F}_2^n$. If $H_\infty(P) \geq d$, then $b_{\mathrm{av}}(P) \leq 2^{-\Omega(\sqrt{d})}$.*

In other words, if for a quadratic source generated by P, we have $b_{\mathrm{av}}(P) = 2^{-k}$, then: $k \leq H_\infty(P) \leq O(k^2)$. One quick application of this result yields the following structural information about quadratic sources that may be of independent interest:

Corollary 1.1 *Suppose $P = (P_1, \ldots, P_n)$ generates a quadratic source over $\mathbb{F}_2^n$ where each $P_i \in \mathbb{F}_2[x_1, \ldots, x_m]$, and suppose there exists $\beta \in \mathbb{F}_2^n$ with $\Pr_x[P(x) = \beta] \geq 2^{-k}$. Then:*

$$\Pr_{z \in \mathsf{Im}(P)} \left[\Pr_{x \in \mathbb{F}_2^m} [P(x) = z] \geq 2^{-O(k^2)} \right] \geq 2^{-O(k^2)}.$$

.

We also use Theorem 1.1 to design *dispersers* for a class of quadratic sources. A disperser for a family $\mathcal{F}$ of sources over $\{0,1\}^n$ is a function $D : \{0,1\}^n \to \{0,1\}$ such that for any source $X \in \mathcal{F}$, the random variable $D(X)$ is not constant. Dispersers are clearly weaker objects than extractors, but no explicit disperser constructions are known for quadratic sources over $\mathbb{F}_2^n$ with a min-entropy guarantee.

We consider the class of *r-separable quadratic sources*. Informally speaking, such a source is generated by quadratic polynomials $P_1, \ldots, P_n : \mathbb{F}_2^m \to \mathbb{F}_2$ so that the graph G_i on m vertices corresponding to each quadratic P_i is r-partite (with the same partition for each graph). A 1-separable quadratic sources is an affine source, while an m-separable quadratic source is a general quadratic source. Thus, r parameterizes in some sense how far from affine the quadratic source is. In this paper, we show that for a broad range of r, affine dispersers are also dispersers for r-separable quadratic sources.

Theorem 1.2 *There exists a constant $C > 0$ such that the following holds. Let* $\textsc{Aff}$ *: ${}^n_2 \mapsto {}_2$ be a disperser for affine sources with min-entropy at least $k_{\min} = k_{\min}(n)$. Let X be a r-separable quadratic source with min-entropy $k := H_\infty(X)$ which satisfies $k \geq Cr^2k_{\min}^2$. Then,* $\textsc{Aff}$ *is a disperser for X.*

Li's work [Li16] gives an affine extractor (hence, a disperser) for sources with $\mathrm{poly}\log n$ min-entropy. Plugging this into Theorem 1.2 yields dispersers for $O(1)$-separable quadratic sources with min-entropy $\mathrm{poly}\log n$ and for $\tilde{O}(\sqrt{n})$-separable quadratic sources with min-entropy $\Omega(n)$.

1.3 Related Work

Bias. The bias of randomness sources has a long history of study in the context of pseudorandom generators. A distribution X on $\{0,1\}^n$ is said to be ε-*biased* if:

$$\max_{c \in \{0,1\}^n} \left| \mathop{\mathbb{E}}_{x \sim X} (-1)^{\langle c,x \rangle} \right| \leq \varepsilon.$$

In other words, X fools all $\mathbb{F}_2$-linear tests. Sums of small-bias sources are known to fool polynomials of bounded degree [BV10, Lov08, Vio09]. Beautiful works by Naor and Naor [NN93], Alon et al [AGHP92, ABN+92], and Ta-Shma [TS17] have yielded explicit ε-biased sources with nearly optimal support size. However, from the point of view of extractors, ε-biased sources are trivial to extract from, and small bias in the above sense is clearly not necessary for extraction.

Bias as a notion of pseudorandomness has also been extensively studied in the context of additive combinatorics. An influential result by Kaufman and Lovett [KL08] states that if a polynomial $P : \mathbb{F}_2^n \to \mathbb{F}_2$ has degree $\leq d$ and bias at least ε, then it has *rank* at most $r(d, \varepsilon)$ for a bounded function r, where $\operatorname{rank}(P)$ is defined as the number of $(d-1)$-degree polynomials P can be written as a function of. This result is a generalization of a fundamental fact about quadratic polynomials, known as *Dickson's Lemma* [Dic58], which relates the bias of a quadratic form to the rank of its associated matrix. For cubics, quartics, and quintics, Haramaty and Shpilka [HS10] and Hatami [Hat16] have obtained quantitatively stronger bounds on the rank in terms of the bias than what [KL08] yields.

Seedless Extractors. Previous work on deterministic extractors have considered various restrictions on the randomness source, such as partitioning the source into a few independent sources [CG88, BIW06, BKS+10, Rao07, BRSW06, Li15, CZ16, Coh16b, Coh16a, Li17], requiring the distribution to be the output of a computational process [vN51, Blu86, TV00, KRVZ06, DW12], making the distribution be fixed on some bits [KZ06, GRS06], and putting algebraic structure on the source [BKS+10, Bou07, Dvi12]. We focus on the last type of restriction.

Over $\mathbb{F}_2$, [Bou07] gave an affine extractor for dimension (min-entropy) $\Omega(n)$, which was improved slightly by Yehudayoff [Yeh11] and Li [Li11] to $\Omega(n/\sqrt{\log \log n})$. Rao [Rao09] gave extractors for $\operatorname{poly} \log n$ dimension when the affine subspace is guaranteed to have a basis of low-weight vectors, while Ben-Sasson and Kopparty [BSK12] and Shaltiel [Sha11] gave affine dispersers for $n^{o(1)}$ dimension. All these results were vastly improved recently by Li [Li16] who constructed affine extractors for polylogarithmic min-entropy.

Dvir, Gabizon and Wigderson [DGW09] introduced the problem of constructing extractors for polynomial sources as a natural extension of affine sources, albeit over large fields. Ben-Sasson and Gabizon [BSG12] considered polynomial sources over $\mathbb{F}_{p^\ell}^n$ and constructed extractors for sources with linear entropy rate with constant ℓ and p. However, their construction does not apply to the hardest case of $\mathbb{F}_2^n$ which we consider here. In fact, we do not even know of explicit dispersers in this regime. An explicit construction of dispersers or extractors for quadratic sources over $\mathbb{F}_2^n$ is known to imply worst case circuit size lower bounds [GKST16].

1.4 Proof Techniques

We sketch the proofs for Theorems 1.1 and 1.2 in this section.

For Theorem 1.1, assume that $b_{\mathrm{av}}(P) \geq \varepsilon$. Assuming that $P(0) = 0$, we show a lower bound on $\Pr_x[P(x) = \alpha]$ for some $\alpha \in \mathbb{F}_2^n$. This implies the theorem by a simple translation trick.

If we consider the set of quadratics $\{P_S = \sum_{i \in S} P_i \mid S \subseteq [n]\}$, then we know that at least $\varepsilon/2$ fraction of the P_S's have bias at last $\varepsilon/2$. Hence, by Dickson's theorem, at least $\varepsilon/2$ fraction of the P_S's have small rank. Call an S *bad* if P_S has small rank.

Since rank is sub-additive, we might expect that the bad S's form a subspace over $\mathbb{F}_2^n$. Indeed, we can make this intuition formal using the Bogolyubov-Chang Lemma from additive combinatorics. Our arguments here closely follow the analysis by Haramaty and Shpilka in [HS10], even though their context is completely different! Whereas we are looking at the space of linear combinations of quadratics $P_1, \ldots, P_n$, they were looking at the space of additive derivatives of a cubic polynomial. But we can use their techniques to find a linear space $V \leq \mathbb{F}_2^n$ of dimension $\geq n - O(\log \varepsilon^{-1})$ such that any P_S for $S \in V$ is a quadratic with rank at most $O(\log^2 \varepsilon^{-1})$. We also need and obtain an additional property: many of the quadratics in V have nonzero bias, not just low rank[3].

Following the line of arguments in [HS10] yields a subspace $N \leq \mathbb{F}_2^m$ of co-dimension $d = O(\log^2 \varepsilon^{-1})$ such that all the quadratics in V restricted to N become affine polynomials. That is, there exist linear forms $\ell_1, \ldots, \ell_d$ such that for every $P_S \in V$, we can write $P_S = \sum_{i=1}^d \ell_i \cdot \ell_i^{(S)} + \ell_0^{(S)}$. For any P_S whose bias is nonzero, it must be the case that $\ell_0^{(S)}$ in the span of $\{\ell_1^{(S)}, \ldots, \ell_d^{(S)}\}$, for otherwise, the bias would be 0 as the bias of any linear form is 0. Hence, for any P_S whose bias is nonzero, we can write $P_S = \sum_{i=1}^d (\ell_i + c_i^{(S)})\ell_i^{(S)}$. Since we also ensure that there are many $S \in V$ for which P_S has nonzero bias, an averaging argument shows that there exists $c_i^*, \ldots, c_d^*$ such that $2^{-O(\log^2 1/\varepsilon)}$ fraction of P_S's are in the ideal generated by $\langle \ell_1 + c_1^*, \ldots, \ell_d + c_d^* \rangle$. Thus, restricting to the affine subspace N^* given by $\ell_1 = c_1^*, \ldots, \ell_d = c_d^*$ makes all these polynomials vanish. It is then easy to see that in fact, the remaining $O(\log^2 1/\varepsilon)$ polynomials are constant on a $2^{-O(\log^2 1/\varepsilon)}$ fraction of N^*, concluding the proof of Theorem 1.1.

For Theorem 1.2, consider the case $r = 2$, so that the source is generated by polynomials $P_1, \ldots, P_n$ where each $P_i(x, y) = x^T A_i y + b_i^T y + c_i^T x + r_i$. Let M_x be the matrix where the ith row is given by $A_i^T x + b_i$ and C be the matrix whose i'th row is given by c_i. Consider the following two cases:

- Suppose $\mathrm{rank}(M_x) > d$ for some x. Then, $P_1, \ldots, P_n$ take values over a d-dimensional affine subspace, for that fixed x and ranging over all y.
- Suppose $\mathrm{rank}(M_x) \leq d$ for all x but $\mathrm{rank}(C) > d'$. Then if we fix a y such that $M_x y = 0$, then $P_1, \ldots P_n$ take values over a d' dimensional affine subspace as we range over x.

We prove that these are the only two cases possible. This is so, since if $\mathrm{rank}(M_x)$ is small for all x and $\mathrm{rank}(C)$ is also small, then we can show a lower bound on the average bias which contradicts the guaranteed min-entropy bound. Thus, if we use an affine disperser for dimension $\min(d, d')$, it is non-constant on our source. For larger r, we can argue similarly using induction.

1.5 Future Directions

Many questions are raised by the work we describe in this paper. While it is clear that the average bias of a source must be < 1 to admit a one-bit extractor, we believe a stronger statement that holds for extracting multiple bits.

Conjecture 1.5 *Let X be a source for which there exists an extractor which can extract m-uniform bits. Then we must have $b_{\mathrm{av}}(X) \leq \exp(-m)$.*

[3] The quadratic $(x_1 + \cdots + x_m)(y_1 + \cdots + y_m) + z$, for instance, has rank 1 but bias 0.

Proving this conjecture would establish our hypothesis that low average bias is a tighter condition for "extractability" than high min-entropy.

Turning to polynomial sources, we showed that for quadratic sources X, $H_\infty(X) \geq k$ implies $b_{\mathrm{av}}(X) \leq 2^{-\Omega(\sqrt{k})}$. We hope such a result is true for any bounded degree source.

Conjecture 1.6 *Let X be a source generated by degree-d polynomials over $\mathbb{F}_2^n$. Then ,if $H_\infty(X) \geq k$, then $b_{\mathrm{av}}(X) \leq 2^{-f_d(k)}$ for some function f_d that depends only on d.*

We identified a broad sub-class of quadratic sources (namely, separable sources), for which affine extractors also act as dispersers. We believe that a stronger claim can also be established, which we state as a conjecture:

Conjecture 1.7 *For every $k > 0$, there exists a k' such that the following holds. An affine extractor for min-entropy at least k' is a disperser for quadratic sources of min-entropy at least k.*

1.6 Organization

In Section 2, we establish some preliminaries and prove Lemmas 1.1 and 1.2. We show Theorem 1.1 and Corollary 1.1 in Section 3 and Theorem 1.2 in Section 4.

2 Preliminaries

Definition 2.1 (Randomness Extractor) *Suppose that $\mathcal{C}$ is a class of sources, i.e. distributions over $\mathbb{F}_2^n$. Then a function* $\mathsf{Ext} : \mathbb{F}_2^n \to \mathbb{F}_2^t$ *is a* ε-extractor *for $\mathcal{C}$-sources, if for all $X \in \mathcal{C}$, we have*

$$\delta_{TV}(\mathsf{Ext}(X), \mathsf{Unif}(\mathbb{F}^t)) \leq \varepsilon$$

where $\delta_{TV}(\cdot, \cdot)$ denotes the statistical (total variation) distance between two distributions.

Definition 2.2 (Dispersers) *Suppose that $\mathcal{C}$ is a class of sources, i.e. distributions over $\mathbb{F}_2^n$. Then a function* $\mathsf{Disp} : \mathbb{F}_2^n \to \mathbb{F}_2$ *is a* disperser *for $\mathcal{C}$-sources, if for all $X \in \mathcal{C}$, we have that* $\mathsf{Disp}(X)$ *is non-constant.*

In our proof of Theorem 1.1 for quadratic polynomial sources, we make heavy use of *Dickson's lemma*. Dickson's lemma provides a canonical representation for quadratic polynomials over finite fields, as well as a very useful notion of rank (which we refer to as *Dickson's lemma rank*) which is tightly connected to the bias of the quadratic when working over $\mathbb{F}_2$.

Lemma 2.3 (Dickson's Lemma (Chapter 15, Theorem 4 of [MS77])) *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$, $f(x) = x^\top Ax + b^\top x + c$ be a quadratic polynomial mapping. Then, there exists a non-singular linear transformation $T : \mathbb{F}_2^n \to \mathbb{F}_2^n$ such that, with the transformation of variables $y = T(x)$, we can write*

$$f \circ T^{-1}(y) = \sum_{i=1}^{r} y_{2i-1} \cdot y_{2i} + b_1^\top y + c$$

where $r = \mathrm{rank}(A + A^\top)/2$. Furthermore, if $b_1^\top y$ is linearly dependent on $y_1, \ldots, y_{2r}$, then $\mathrm{bias}(f) = 2^{-r}$. Otherwise, $\mathrm{bias}(f) = 0$.

Definition 2.4 (Dickson's Lemma Rank) *For a quadratic polynomial mapping $f = x^\top Ax + b^\top x + c$, we define its Dickson's lemma rank as $\mathrm{rank}_2(f) = \mathrm{rank}(A + A^\top)/2$, i.e. the number of quadratic terms in its Dickson's lemma representation.*

We use standard notions of Fourier analysis over $\mathbb{F}_2^n$. For a function $f : \mathbb{F}_2^n \to \mathbb{R}$, we denote its Fourier coefficients by $\hat{f}(c) = \mathbb{E}_x[f(x) \cdot (-1)^{\langle c,x \rangle}]$. We refer to [O'D14] for useful facts about Fourier coefficients of functions over $\mathbb{F}_2^n$.

2.1 General Facts about Average Bias

Given a distribution Z on $\mathbb{F}_2^n$, let $p_Z : \mathbb{F}_2^n \to [0,1]$ be its probability density function. Then, observe that:

$$b_{\text{av}}(Z) = \frac{1}{2^n} \sum_{c \in \mathbb{F}_2^n} \left| \mathop{\mathbb{E}}_{x \sim Z} (-1)^{\langle c,x \rangle} \right| = \frac{1}{2^n} \sum_{c \in \mathbb{F}_2^n} \left| \sum_{x \in \mathbb{F}_2^n} p_Z(x) \cdot (-1)^{\langle c,x \rangle} \right| = \sum_{c \in \mathbb{F}_2^n} |\hat{p}_Z(c)|, \tag{1}$$

which is also known as the *spectral norm* of p_Z.

Small Average Bias implies Large Min-Entropy

Lemma 1.1 *For a distribution Z on $\{0,1\}^n$, if $b_{\text{av}}(Z) \leq 2^{-k}$, then $H_\infty(Z) \geq k$.*

Proof. As above, let p_Z be the probability density function of Z. Then, for any $x \in \mathbb{F}_2^n$:

$$p_Z(x) = \left| \sum_{c \in \mathbb{F}_2^n} \hat{p}_Z(c)(-1)^{\langle c,x \rangle} \right| \leq \sum_{c \in \mathbb{F}_2^n} |\hat{p}_Z(c)| = b_{\text{av}}(Z)$$

due to the Fourier inversion formula, triangle inequality and (1) respectively. □

A one-bit extractor for low average bias sources

As a first application of average bias as a measure of pseudorandomness, we show that any bent function (Ch. 6, [O'D14]) is an extractor for sources with small average bias, as shown in the following lemma.

Lemma 2.5 (Lemma 1.2 restated) *Let $G : \mathbb{F}_2^n \to \mathbb{F}_2$ be a bent function. If X is a source over $\mathbb{F}_2^n$ with $b_{\text{av}}(X) \leq 2^{-k}$ for $k \geq (1+\alpha)(n/2)$ and $\alpha > 0$, then G is a ε-extractor for X with $\varepsilon = O(2^{-\alpha n}) = o(1)$. In particular, if $k > n/2$, then G is a disperser.*

Proof. Let $g : \mathbb{F}_2^n \to \{\pm 1\}$ be defined as $g := (-1)^G$.

$$\left| \mathop{\mathbb{E}}_{x \sim X} g(x_1, \ldots, x_n) \right| = \left| \sum_{x \in \mathbb{F}_2^n} g(x) \cdot p_X(x) \right| = 2^n \cdot \left| \sum_{c \in \mathbb{F}_2^n} \hat{g}(c) \cdot \hat{p}_X(c) \right| \leq 2^n \cdot 2^{-n/2} \cdot \sum_{c \in \mathbb{F}_2^n} |\hat{p}_X(c)|$$

where the last equality is Plancherel's identity and the last inequality is from the definition of a bent function. Applying (1), we see that if $b_{\text{av}}(X) \leq 2^{-(k-n/2)}$, $\mathbb{E}_{x \sim X}\, g(x) \leq 2^{-(k-n/2)}$ (which is < 1, if $k > n/2$).

Applying Vazirani's XOR lemma [Gol95] in this case, we get $\|G(X) - U_1\|_1 \leq \sqrt{2} \cdot 2^{-(k-n/2)}$. So if we have $k \geq (1+\alpha)(n/2)$ for some $\alpha > 0$, we get that the bent function G ε-extractor for sources X such that $b_{\text{av}}(X) \leq 2^{-(1+\alpha)(n/2)}$, with $\varepsilon = \sqrt{2} \cdot 2^{-\alpha n} = o(1)$. □

We reproduce an argument of Andrej Bogdanov that it is in general impossible to extract from sources of larger average bias.

Lemma 2.6 *For any function $G : \mathbb{F}_2^n \to \{0,1\}$, there exists a source X over $\mathbb{F}_2^n$ with $b_{av}(X) \leq 2^{-(n-1)/2}$ such that $G(X)$ is constant.*

Proof. Suppose without loss of generality that $|G^{-1}(1)| \geq 2^{n-1}$. Let X be the distribution that is uniform over $G^{-1}(1)$. Then:

$$b_{\mathrm{av}}(X) = \sum_c |\hat{p}_X(c)| \leq 2^{n/2} \cdot \sqrt{\sum_c \hat{p}_X^2(c)} = 2^{n/2}\sqrt{\mathop{\mathbb{E}}_x p_X^2(x)} \leq 2^{n/2}\sqrt{\frac{1}{2^n} \cdot \frac{1}{2^{n-1}}} = 2^{-(n-1)/2}$$

using (1), Cauchy-Schwarz, Parseval's identity, and the definition of X respectively. □

3 Average Bias vs. Min-entropy for Quadratic Sources

We have already seen the relationship between average bias and min-entropy in one direction, namely that a source with average bias $\leq 2^{-k}$ has min-entropy $\geq k$ (Lemma 1.1). In this section, we derive a relationship between min-entropy and average bias in the other direction, for the special case of quadratic polynomial sources.

Theorem 1.1 *Let $P = (P_1, \ldots, P_n)$ generate a quadratic source over $\mathbb{F}_2^n$. If $H_\infty(P) \geq d$, then $b_{\mathrm{av}}(P) \leq 2^{-\Omega(\sqrt{d})}$.*

In the following subsection, we prove the above theorem.

3.1 Structure of quadratic forms with High Average Bias

Throughout this subsection, we will consider $P_1, \ldots, P_n \in \mathbb{F}_2[x_1, x_2, \ldots, x_m]$ to be quadratic polynomials which satisfy $b_{\mathrm{av}}(P) \geq \varepsilon$ for $\varepsilon = 2^{-o(n)}$. Additionally we shall assume that $P(0) = 0$. The main result of this subsection is the following lemma:

Lemma 3.1 *Let $P = (P_1, P_2, \ldots, P_n)$ be quadratic polynomials in $\mathbb{F}_2[x_1, \ldots, x_m]$ satisfying $P(0) = 0$. Let $\varepsilon := b_{\mathrm{av}}(P)$ denote the average bias of P. Then there exists a subspace $U \subseteq \mathbb{F}_2^n$ of co-dimension $O(\log^2(1/\varepsilon))$ and an affine subspace $N^* \subseteq \mathbb{F}_2^m$ of co-dimension $O(\log^2(1/\varepsilon))$ satisfying the following condition: for every set $S \in U$, the corresponding quadratic $P_S = \sum_{i \in S} P_i$ vanishes on N^* i.e., $P_S|_{N^*} = 0$.*

The proof of the above lemma consists of 3 steps.

1. Finding a large subspace V of low rank quadratics: Since $b_{\mathrm{av}}(P) = \mathbb{E}_S \operatorname{bias}(P_S) = \varepsilon$, by an averaging argument we have that for at least $\varepsilon/2$ fraction of choices of S, the corresponding bias term satisfies $\operatorname{bias}(P_S) \geq \varepsilon/2$. Let $A := \{S \in \mathbb{F}_2^n : \operatorname{bias}(P_S) \geq \varepsilon/2\}$ denote the set of S's with large bias, and let $\mu_0 = |A|/2^n$ satisfying $\mu_0 \geq \varepsilon/2$. Let $\operatorname{rank}_2(P)$ of a quadratic form P be the number of quadratic terms in its Dickson's lemma representation $\sum_{i=1}^r x_{2i-1}x_{2i} + \ell(x)$ (see Definition 2.4).

Lemma 3.1 *There exists a subspace $V \subseteq \mathbb{F}_2^n$ of co-dimension $O(\log(1/\varepsilon))$ such that the following properties hold:*

1. *For all $S \in V$, $\operatorname{rank}_2(P_S) \leq O(\log^2(1/\varepsilon))$*
2. *The density of A in V satisfies $\mu = \dfrac{|A \cap V|}{|V|} \geq \mu_0 \geq \varepsilon/2$.*

Proof. We show the existence of the subspace V by applying the Bogolyubov-Chang Lemma to the set A. In particular, we invoke the proof of the lemma in [HS10], Lemma 2.3. We only sketch the construction and observe that part (2) of the Lemma also holds.

The claim is that there exists a subspace $V \subseteq \mathbb{F}_2^n$ of co-dimension $O(\log(1/2\mu_0)) = O(\log(1/\varepsilon))$ contained in $kA - kA$, for $k \leq \max\left(1, \left\lceil \frac{1}{2}\left(\log_{\frac{4}{3}}(2/\mu_0) + 2\right)\right\rceil\right) = O(\log(1/\varepsilon))$. The way V is constructed in [HS10], Lemma 2.3, is by producing a sequence of subspaces $\mathbb{F}_2^n = W_0, W_1, \ldots, W_t$ such that $|A \cap W_i|/|W_i| \geq 1.5 \cdot |A \cap W_{i-1}|/|W_{i-1}|$ and $t \leq \log_{3/2}(1/2\mu_0)$. Then, if $V = W_t$, it is shown that $V = k(A \cap V) - k(A \cap V)$ for k as above.

Part (1) of the Lemma follows because for any $S \in A$, $\mathrm{rank}_2(P_S) \leq \log(2/\varepsilon)$ (since the bias of a Dickson's lemma representation with r quadratic terms is 2^{-r} or 0) and so for any $S \in V$, $\mathrm{rank}_2(P_S) \leq k \cdot \log(2/\varepsilon) = O(\log^2 1/\varepsilon)$ since rank_2 is subadditive.

Part (2) of the Lemma also holds because the density $|A \cap W_i|/|W_i|$ strictly increases with i by construction. Hence, we have that the density of A in V is $\geq \mu_0$ (and in fact, is $> \mu_0$ if $t > 0$). □

2. Finding a large subspace N on which the quadratics are linear. Let $V \subseteq \mathbb{F}_2^n$ be a subspace as guaranteed by Lemma 3.1. We now have that there exists a small set of linear forms such that the quadratic component (in the Dickson's Lemma representation) of every $P \in V$ can be expressed as linear combinations of these linear forms.

Lemma 3.2 *There exist linearly independent linear forms $\ell_1, \ell_2, \ldots, \ell_d$, with $d = O(\log^2(1/\varepsilon))$ for which the following holds. For all $S \in V$, we can write:*

$$P_S = \sum_{i=1}^{d} \ell_i \ell_i^{(S)} + \ell_0^{(S)}$$

for some linear forms $\ell_0^{(S)}, \ell_1^{(S)}, \ldots, \ell_d^{(S)}$. Furthermore, without loss of generality, the ℓ_i's are linearly independent, and the $\ell_0^{(S)}$s are linearly independent of the ℓ_i's.

Proof. Follows directly from Lemma 3.7 of [HS10]. The last line is because if the ℓ_i's were linearly dependent, then d can be decreased, and if $\ell_0^{(S)}$ were linearly dependent on the ℓ_i's, then the $\ell_i^{(S)}$'s can be changed to make $\ell_0^{(S)}$ equal to 0. □

Let $N \subseteq \mathbb{F}_2^m$ be the subspace $\{x : \ell_1(x) = \cdots = \ell_d(x) = 0\}$ with co-dimension $d = O(\log^2(1/\varepsilon))$, where $\ell_1, \ldots, \ell_d$ are as in Lemma 3.2. For every choice of $c = (c_1, \ldots, c_d) \in \mathbb{F}_2^d$, let $N[c]$ denote the coset of N parametrized by $N[c] := \{x : \ell_i(x) = c_i \ \ \forall i \in [d]\}$. Note that by construction, there are exactly 2^d such cosets.

The following proposition states some useful structural properties of the linear forms $\ell_i^{(S)}$.

Proposition 3.3 *For all $S \in V$, the linear forms $\ell_i^{(S)}$ have an additive structure inherited from the polynomials $\{P_S\}$ themselves i.e,. for any choice of sets $S_1, S_2 \in V$ with $S = S_1 + S_2$, we have $\ell_i^{(S)} = \ell_i^{(S_1)} + \ell_i^{(S_2)}$ for $i = 0, \ldots, d$. Moreover, for any choice of $c \in \mathbb{F}_2^d$ and sets $S_1, S_2 \in V$, if $\ell_0^{(S_1)} = \sum_i c_i \ell_i^{(S_1)}$ and $\ell_0^{(S_2)} = \sum_i c_i \ell_i^{(S_2)}$, then $\ell_0^{(S_1+S_2)} = \sum_i c_i \ell_i^{(S_1+S_2)}$*

Proof. For the first part, fix any pair of sets $S_1, S_2 \in V$. Since $P_{S_1+S_2} = P_{S_1} + P_{S_2}$, with $S = S_1 + S_2$, we have $P_S = \sum_{i=1}^{d} \ell_i \cdot (\ell_i^{(S_1)} + \ell_i^{(S_2)}) + (\ell_0^{(S_1)} + \ell_0^{(S_2)})$ and the observation follows.

The second part can be argued as follows. We have $\ell_0^{(S)} = \sum_{i \in [d]} c_i \ell_i^{(S)}$ if and only if $P_S \mid_{N[c]} = 0$. Since $P_S = P_{S_1} + P_{S_2}$, $P_{S_1} \mid_{N[c]} = 0$ and $P_{S_2} \mid_{N[c]} = 0 \implies P_S \mid_{N[c]} = 0$. The observation then follows. □

We shall also need the following observation, which says that for any quadratic $P_S : S \in V$ with nonzero bias, the linear component $\ell_0^{(S)}$ must be in the span of the linear forms $\ell_1^S, \ldots, \ell_d^{(S)}$.

Proposition 3.4 *For each $S \in V$ such that* $\mathrm{bias}(P_S) \neq 0$*, there exists $c^{(S)} \in \mathbb{F}_2^d$ such that $\ell_0^{(S)} = \sum_i c_i^{(S)} \ell_i^{(S)}$*

Proof of Proposition 3.4. We fix a set $S \in V$. Recall that from Lemma 3.2, the linear forms $\ell_1, \ell_2, \ldots, \ell_d$ are linearly independent. Therefore, we can assume $\ell_i = x_i \ \forall i \in [d]$ by a non-singular linear renaming of variables. Let $W := \mathsf{Span}_{\mathbb{F}_2}\left(\ell_1^{(S)}, \ldots, \ell_d^{(S)}\right)$. Let $\mathcal{L}(\mathbb{F}_2^m, \mathbb{F}_2)$ denote the vector space of linear forms on $\mathbb{F}_2^m$. Then we can decompose $\mathcal{L}(\mathbb{F}_2^m, \mathbb{F}_2) = W \oplus W^\perp$ (non-unique direct sum complement).

We shall prove our claim by contradiction i.e., we assume that $\ell_0^{(S)} \notin W$. Then we can write $\ell_0^{(S)} = \ell_{00}^{(S)} + \ell_{01}^{(S)}$ such that $\ell_{00}^{(S)} \in W$, and $\ell_{01}^{(S)} \in W^\perp$. In particular, since $\ell_0^{(S)} \notin W$, we have $\ell_{01}^{(S)} \neq 0$. Furthermore, from Lemma 3.2 we know that $\ell_0^{(S)}$ (and consequently $\ell_{01}^{(S)}$) must be linearly independent of the linear forms $\ell_1, \ell_2, \ldots, \ell_d$.

Let $w_1, \ldots, w_r$ be a basis of W and $w_{r+1}, \ldots, w_n$ to be a basis of $W^\perp$, which together forms a basis of $\mathcal{L}(\mathbb{F}_2^m, \mathbb{F}_2)$. Then, using this basis, we can perform another non-singular linear renaming $w_i \mapsto z_i$ so that $\ell_{01}^{(S)}$ is variable disjoint in terms of $\{z_i\}$ from $\{\ell_1^{(S)}, \ldots, \ell_d^{(S)}, \ell_{00}^{(S)}, \}$ and $\{\ell_1, \ell_2, \ldots, \ell_d\}$. Then without loss of generality, we can assume that $\ell_{01}^{(S)} = z_1$, and the rest of the linear forms can be expressed in terms of $z_2, z_3, \ldots, z_n$. But then we have

$$\begin{aligned}
\mathrm{bias}(P_S) = \left| \mathop{\mathbb{E}}_{x_1,\ldots,x_m} \left[(-1)^{P_S(x)} \right] \right| &= \left| \mathop{\mathbb{E}}_{x_1,\ldots,x_m} \left[(-1)^{\sum_{i\in[d]} \ell_i(x)\ell_i^{(S)}(x) + \ell_{00}^{(S)}(x) + \ell_{01}^{(S)}(x)} \right] \right| \\
&= \left| \mathop{\mathbb{E}}_{z_1,\ldots,z_n} \left[(-1)^{\sum_{i\in[d]} \ell_i(z_{\geq 2})\ell_i^{(S)}(z_{\geq 2}) + \ell_{00}^{(S)}(z_{\geq 2}) + z_1} \right] \right| \\
&= \left| \mathop{\mathbb{E}}_{z_2,\ldots,z_n} \left[(-1)^{\sum_{i\in[d]} \ell_i(z_{\geq 2})\ell^{(S)}(z_{\geq 2}) + \ell_{00}^{(S)}(z_{\geq 2})} \mathop{\mathbb{E}}_{z_1} (-1)^{z_1} \right] \right| \\
&= 0
\end{aligned}$$

□

3. Finding a large subspace U of quadratics vanishing on a large subspace N^*. Finally, we show that there exists a large subspace $U \subseteq \mathbb{F}_2^n$ and a coset of N such that every quadratic from U vanishes on the coset.

Lemma 3.5 *There exists a subspace $U \subseteq \mathbb{F}_2^n$ with $\mathsf{dim}(U) \geq n - O(\log^2(1/\varepsilon))$, and $c \in \mathbb{F}_2^d$ such that the following holds: For all $S \in U$, $P_S \mid_{N[c]} = 0$, or equivalently, $\ell_0^{(S)} = \sum_{i\in[d]} c_i \ell_i^{(S)}$.*

Proof. By Lemma 3.1, there exists a large subset $\widehat{V} \subseteq V$ of density $\geq \varepsilon/2$ such that for all $S \in \widehat{V}$, $\mathrm{bias}(P_S) \neq 0$. By Proposition 3.4, for all $S \in \widehat{V}$, there exists at least one $c^{(S)}$ such that $P_S \mid_{N[c^{(S)}]} = 0$. So for each S, choose one such representative $c^{(S)}$ and let $\mathcal{C} = \{c^{(S)} : S \in \widehat{V}\}$. Since every coset of N can be identified by a d-dimensional vector over $\mathbb{F}_2$, we must have $|\mathcal{C}| \leq 2^d = 2^{O(\log^2(1/\varepsilon))}$.

Now recall that from Lemma 3.1, our choice of $\hat{V}$ satisfies

$$|\widehat{V}| \geq \frac{\varepsilon}{2}|V| = \frac{1}{2^{O(\log(1/\varepsilon))+1}} \cdot 2^{n-O(\log(1/\varepsilon))} = 2^{n-O(\log(1/\varepsilon))}$$

Therefore, by an averaging argument, there exists $c \in \mathbb{F}_2^d$ such that for at least $|\widehat{V}|/2^d \geq 2^{n-O(\log^2(1/\varepsilon))}$ choices of sets $S \in V$, we have $c^{(S)} = c$. Let $U = \left\{S \in V : c^{(S)} = c\right\}$ be the collection of all such sets from V.

Finally, we claim that U is closed under addition over $\mathbb{F}_2$, and therefore, is a subspace. To see this, fix any pair of sets $S_1, S_2 \in \Gamma$. Then $P_{S_1}|_{N[c]} = P_{S_2}|_{N[c]} = 0$, which using Proposition 3.3 implies that $P_{S_1+S_2}|_{N[c]} = 0$, or equivalently $S_1 + S_2 \in U$. Since $|U| \geq 2^{n-O(\log^2(1/\varepsilon))}$ it follows that the dimension of U is at least $\log |U| \geq n - O(\log^2(1/\varepsilon))$. □

Since the above lemma directly gives us the a large dimensional subspace U of quadratics, all of which vanish on a large dimensional affine subspace $N^* := N[c]$, this concludes the proof of Lemma 3.1.

3.2 Finishing the proof of Theorem 1.1

We first handle the case when $P(0) = 0$.

Lemma 3.6 *Let $P = (P_1, P_2, \dots, P_n)$ be quadratic polynomials in $\mathbb{F}_2[x_1, \dots, x_m]$ satisfying $P(0) = 0$, and let $\varepsilon = b_{\mathrm{av}}(P)$. Then $\Pr_x(P(x) = \alpha) \geq 2^{-O(\log^2(1/\varepsilon))}$, for some $\alpha \in \mathbb{F}_2^n$.*

Proof. By a linear transformation of the polynomials, we can assume without loss of generality that the subspace U obtained from Lemma 3.1 is spanned by $P_1, \dots, P_t$ where $t = n - O(\log^2(1/\varepsilon))$.

Let N^* be the affine subspace from Lemma 3.1. By averaging, there exists some $\alpha' \in \mathbb{F}_2^t$ such that

$$\Pr_{x \in N^*}[(P_{t+1}(x), \dots, P_n(x)) = \alpha'] \geq \frac{1}{2^{n-t}}.$$

Hence:

$$\Pr_x[(P_1(x), \dots, P_n(x)) = (0^t, \alpha')] \geq \frac{|N^*|}{2^n} \cdot \frac{1}{2^{n-t}} \geq 2^{-O(\log^2(1/\varepsilon))}.$$

□

We are now ready to prove our main theorem

Proof of Theorem 1.1. Consider the vector of polynomials $P' = P - P(0)$. Then, the average bias of P' is equal to the average bias of P, and $P'(x) = 0 \iff P(x) = P(0)$. Applying Lemma 3.6 to P' gives that for some $\alpha \in \mathbb{F}_2^n$, $\Pr_x(P(x) = P(0) + \alpha) \geq 2^{-c \cdot \log^2(1/\varepsilon)}$, and so $H_\infty(P) \leq O(\log^2 1/\varepsilon)$. □

3.3 The distribution of a quadratic polynomial source

An affine source has the probability mass distributed uniformly in its support. Using the result above, we can say something similar for quadratic sources.

Corollary 1.1 *Suppose $P = (P_1, \dots, P_n)$ generates a quadratic source over $\mathbb{F}_2^n$ where each $P_i \in \mathbb{F}_2[x_1, \dots, x_m]$, and suppose there exists $\beta \in \mathbb{F}_2^n$ with $\Pr_x[P(x) = \beta] \geq 2^{-k}$. Then:*

$$\Pr_{z \in \mathsf{Im}(P)}\left[\Pr_{x \in \mathbb{F}_2^m}[P(x) = z] \geq 2^{-O(k^2)}\right] \geq 2^{-O(k^2)}.$$

.

Proof. Since there exists a β in $\mathsf{Im}(P)$ with probability mass $\geq 2^{-k}$, we have $H_\infty(P) \leq k$. This implies that $b_{\mathrm{av}}(P) \geq 2^{-k}$, by Lemma 1.1.

Consider any y in $\mathsf{Im}(P)$, and let x_0 be such that $y = P(x_0)$. Now, define P' as $P'(x) = P(x + x_0) + y$. It is easy to see that P' also generates a quadratic source with the same average bias as P (since x and $x + x_0$ are distributionally identical for a fixed x_0). From Lemma 3.6, there exists $\alpha \in \mathbb{F}_2^n$ such that:

$$\Pr_x[P'(x) = \alpha] = \Pr_x[P(x) = y + \alpha] \geq 2^{-O(k^2)}$$

for some constant $c > 0$. Let $z_y = y + \alpha$. (Note that α may well depend on y.)

From the proof of Lemma 3.6, α is nonzero only in $O(k^2)$ coordinates[4]. This means that for any $z \in \mathsf{Im}(P)$, there are at most $2^{O(k^2)}$ choices of $y \in \mathsf{Im}(P)$ such that $z = z_y$. Thus, there are $\mathsf{Im}(P)/2^{O(k^2)}$ many distinct z's such that $\Pr[P(x) = z] \geq 2^{-O(k^2)}$, establishing the claim. □

4 Separable Quadratic Sources

In this section, we introduce the notion of a r-separable quadratic source and show that affine dispersers are also dispersers for separable quadratic sources with high enough min-entropy. We begin by defining separable quadratic sources.

Definition 4.1 (r-Separable Quadratic Source) *An* r-separable quadratic source *is a quadratic source generated by* $P = (P_1, \ldots, P_n)$ *where each* $P_i \in \mathbb{F}_2[x_1^{(1)}, \ldots, x_{m_1}^{(1)}, x_1^{(2)}, \ldots, x_{m_2}^{(2)}, \ldots, x_1^{(r)}, \ldots, x_{m_r}^{(r)}]$ *is a quadratic polynomial containing no monomial of the form* $x_{j_1}^{(i)} \cdot x_{j_2}^{(i)}$ *for any* $i \in [r]$, $j_1, j_2 \in [m_i]$.

In particular, a 2-separable quadratic source *is generated by polynomials of the form* $P_i(x, y) = x^\top A_i y + b_i^\top y + c_i^\top x + e_i$, *where* $A_i \in \mathbb{F}_2^{m_1 \times m_2}$, $b_i \in \mathbb{F}_2^{m_2}$, $c_i \in \mathbb{F}_2^{m_1}$, $e_i \in \mathbb{F}_2$.

Intuitively, r-separable quadratic sources with small r should behave close to affine sources. We establishing this intuition by showing that affine dispersers are also dispersers for r-separable quadratic sources with large min-entropy, as stated formally in the following theorem:

Theorem 1.2 *There exists a constant* $C > 0$ *such that the following holds. Let* $\textsc{Aff} : {}^n_2 \mapsto {}_2$ *be a disperser for affine sources with min-entropy at least* $k_{\min} = k_{\min}(n)$. *Let* X *be a* r*-separable quadratic source with min-entropy* $k := H_\infty(X)$ *which satisfies* $k \geq Cr^2 k_{\min}^2$. *Then,* $\textsc{Aff}$ *is a disperser for* X.

Instantiating the above theorem with Li's extractor from [Li16] directly gives us the following corollary.

Corollary 4.2 *Let* $\textsc{Aff} : {}^n_2 \mapsto$ *be the affine disperser (actually an extractor) from [Li16], and let* Z *be a* r*-separable quadratic source with min-entropy at least* $\Omega(r^2 \log^{2C}(n))$ *where* C *is as in Theorem 1.8 in [Li16]. Then* $\textsc{Aff}$ *is a disperser for* Z.

In particular, one can set r to be as large as $\tilde{O}(\sqrt{n})$ (where the $\tilde{O}(\cdot)$ hides polylogarithmic factors in n), for which Li's extractor [Li16] will be a disperser for r-separable quadratic source with polylogarithmic min-entropy.

The proof of Theorem 1.2 crucially uses average bias as an intermediate measure of psuedorandomness to go back and forth between the min-entropy of the separable source and the ranks of the coefficient matrices of the separable source. For simplicity of exposition, we first prove the

[4] Assume without loss of generality that the linear transformation of the polynomials indicated in the first paragraph of the proof of Lemma 3.6 has been applied on P'.

theorem here for the case $r = 2$, which will highlight the key steps for the proving the general case, and later extend these techniques to prove the full theorem.

4.1 The case $r = 2$

We state the guarantees for the case $r = 2$ in the following theorem:

Theorem 4.3 *There exists a constant $\gamma > 0$ such that the following holds. Let* $\textsc{Aff}$: ${}_2^n \mapsto$ *be a disperser for affine sources with min-entropy at least $k_{\min} = k_{\min}(n)$. Let Z be a 2-separable quadratic source such that $H_\infty(Z) \geq \gamma k_{\min}^2$. Then* $\textsc{Aff}$ *is a disperser for Z.*

Before we proceed with the proof of the above theorem, we setup some additional notation that will be used in the rest of the proof. As in Definition 4.1, let the quadratic source Z be generated by polynomials $P_i(x,y) = x^\top A_i y + b_i^\top y + c_i^\top x + e_i$, where $A_i \in \mathbb{F}_2^{m_1 \times m_2}$, $b_i \in \mathbb{F}_2^{m_2}$, $c_i \in \mathbb{F}_2^{m_1}$, $e_i \in \mathbb{F}_2$. For ease of notation, we can write the vector function $P = (P_1, \ldots, P_n)$ as $P(x,y) = M_x y + Cx + e$, where the i^{th} row of M_x is $(M_x)_i = (A_i)^\top x + b_i$, $C = [c_1\, c_2, \cdots, c_n]^\top$ and $e = [e_1\, e_2, \cdots, e_n]^\top$. The vector source is then $Z = P(U,V)$ where $U \sim \mathsf{Unif}(\mathbb{F}_2^{m_1})$, $V \sim \mathsf{Unif}(\mathbb{F}_2^{m_2})$.

We shall need the following lemma, which says that, assuming the source Z has large min-entropy (and hence, small average bias), if $\operatorname{rank}(M_x)$ is small for all x, the min-entropy must result from the quantity Cx, and therefore, C must have large rank.

Lemma 4.4 *Suppose that we have* $\operatorname{rank}(M_x) \leq d$ *with for all $x \in \mathbb{F}_2^{m_1}$. Then $b_{\text{av}}(P) \geq 2^{-(d+\operatorname{rank}(C))}$.*

Proof. We ignore the constant term e in the calculations, since average bias is translation invariant.

$$
\begin{aligned}
b_{\text{av}}(P) := \mathop{\mathbb{E}}_{\lambda} \left| \mathop{\mathbb{E}}_{x,y} \left[(-1)^{\lambda^\top M_x y + \lambda^\top Cx} \right] \right| &= \mathop{\mathbb{E}}_{\lambda} \left| \mathop{\mathbb{E}}_{x} \left[(-1)^{\lambda^\top Cx} \mathop{\mathbb{E}}_{y} (-1)^{\lambda^\top M_x y} \right] \right| \\
&= \mathop{\mathbb{E}}_{\lambda} \left| \mathop{\mathbb{E}}_{x} \left[(-1)^{\lambda^\top Cx} \quad {}_{\{\lambda^\top M_x = 0\}} \right] \right| \\
&\geq \mathop{\mathbb{E}}_{\lambda} \mathop{\mathbb{E}}_{x} \left[(-1)^{\lambda^\top Cx} \quad {}_{\{\lambda^\top M_x = 0\}} \right] \\
&= \mathop{\mathbb{E}}_{x} \left[\Pr_{\lambda} \left[\lambda^\top M_x = 0 \right] \mathop{\mathbb{E}}_{\lambda \sim \mathrm{Null}(M_x^\top)} \left[(-1)^{\lambda^\top Cx} \right] \right] \\
&\geq 2^{-d} \mathop{\mathbb{E}}_{x} \mathop{\mathbb{E}}_{\lambda \sim \mathrm{Null}(M_x^\top)} \left[(-1)^{\lambda^\top Cx} \right] \\
&\overset{1}{\geq} 2^{-d} \mathop{\mathbb{E}}_{x} \mathop{\mathbb{E}}_{\lambda} \left[(-1)^{\lambda^\top Cx} \right] \\
&= 2^{-d} \mathop{\mathbb{E}}_{x} \left[\mathbb{1}_{\{Cx = 0\}} \right] \\
&= 2^{-(d+\operatorname{rank}(C))}
\end{aligned}
$$

Step 1 uses the following proposition:

Proposition 4.5 *For any subspace $V \leq \mathbb{F}_2^t$ and vector $a \in \mathbb{F}_2^t$:*

$$
0 \leq \mathop{\mathbb{E}}_{\lambda \sim \mathbb{F}_2^t} [(-1)^{\lambda^\top a}] \leq \mathop{\mathbb{E}}_{\lambda \sim V} [(-1)^{\lambda^\top a}].
$$

Proof. Note that for any vector space W, $\mathbb{E}_{\lambda \sim W}[(-1)^{\lambda^\top a}]$ is either 0 or 1. So, the only case we need to rule out is that $\mathbb{E}_{\lambda \sim \mathbb{F}_2^t}[(-1)^{\lambda^\top a}] = 1$ but $\mathbb{E}_{\lambda \sim V}[(-1)^{\lambda^\top a}] = 0$. But this is impossible since the first equality implies $a = 0$. □

□

Using the above lemma we now proceed to prove Theorem 4.3. Let $\textsc{Aff} : \mathbb{F}_2^n \mapsto \mathbb{F}_2$ be an affine disperser for min-entropy at least $k_{\min}$.

Let X be a separable quadratic source with min-entropy $\geq k$. Since $k \geq \gamma k_{\min}^2$, using Theorem 1.1, this implies that its average bias is $\leq 2^{-\gamma_1 \sqrt{k}} \leq 2^{-\gamma_1 k_{\min}}$ for some constant γ_1, which grows with γ. Furthermore, we set $d := \gamma_1 k_{\min}/2$. Note that we can choose γ to be a large enough constant, such that $\gamma_1 > 2$ and $d > k_{\min}$. We now divide the analysis into two cases:

- **Case (i)**: Suppose $\operatorname{rank}(M_x) > d$ for at least one $x = x^* \in \mathbb{F}_2^{m_1}$. Then, restricted to $x = x^*$ and y uniformly chosen from $\mathbb{F}_2^{m_2}$, $P(x^*, y) = M_{x^*} y + Cx^* + e$ which is uniform over an affine subspace of dimension $> d$. By definition, $\textsc{Aff}$ is non-constant over any affine subspace of dimension $\geq k_{\min}$.
- **Case (ii)**: Suppose $\operatorname{rank}(M_x) \leq d$ for all x. Then using Lemma 4.4 we have $2^{-\operatorname{rank}(C)} \leq 2^d\big(b_{\text{av}}(f)\big) \leq 2^d\big(2^{-\gamma_1 k_{\min}}\big) = 2^{-d}$ or equivalently, $\operatorname{rank}(C) \geq d$.

  Let $p := \Pr_{x,y}\left[y \in \operatorname{Null}(M_x)\right) \geq 2^{-d}$. Then, we can upper-bound the bias of the disperser as follows.

  $$\begin{aligned} \Big| \mathop{\mathbb{E}}_x \mathop{\mathbb{E}}_y (-1)^{\textsc{Aff}(M_x y + Cx + e)} \Big| &\leq p \Big| \mathop{\mathbb{E}}_x \mathop{\mathbb{E}}_{y \sim \operatorname{Null}(M_x)} (-1)^{\textsc{Aff}(M_x y + Cx + e)} \Big| + 1 - p \\ &= p \Big| \mathop{\mathbb{E}}_x \mathop{\mathbb{E}}_{y \sim \operatorname{Null}(M_x)} (-1)^{\textsc{Aff}(Cx + e)} \Big| + 1 - p \\ &= p \Big| \mathop{\mathbb{E}}_x (-1)^{\textsc{Aff}(Cx + e)} \Big| + 1 - p < 1 \end{aligned}$$

  where in the last step, we use the fact that $\textsc{Aff}$ is non-constant on an affine subspace of dimension $\geq d$.

Since in both cases (i) and (ii), we have established that $\Big| \mathbb{E}_{x,y}(-1)^{\textsc{Aff}(M_x y + Cx + e)} \Big| < 1$, it follows that $\textsc{Aff}$ is a disperser for X. This concludes the proof of Theorem 4.3.

4.2 Proof for general r

An r-separable quadratic source is generated by polynomials of the form $P_i(x_1, \ldots, x_r) = \sum_{1 \leq j_1 < j_2 \leq r} x_{j_1}^\top A_{j_1 j_2}^{(i)} x_{j_2} + \sum_{j \in [r]} (b_j^{(i)})^\top x_j + c^{(i)}$ for some $A_{j_s, j_t}^{(i)} \in \mathbb{F}_2^{m_{j_1} \times m_{j_2}}$, $b_j^{(i)} \in \mathbb{F}_2^{m_j}$ and $c^{(i)} \in \mathbb{F}_2$ Succinctly, we can also write the vector function P as $P = \sum_{i=1}^r M_{\leq i} x_i + c$ where the matrix $M_{\leq i}$ depends on $x_1, x_2, \ldots, x_{i-1}$.

The proof of Theorem 1.2 goes through the following generalization of Lemma 4.4.

Lemma 4.6 *For P as above, suppose $b_{\text{av}}(P) \leq 2^{-k}$. Let $0 < C < 1$ be a constant. Suppose that for all $1 < i \leq r$, we have $\operatorname{rank}\big(M_{\leq i}\big) \leq d$ with probability 1, where d satisfies $rd \leq (1 - C) \cdot k$. Then, we have $\operatorname{rank}(M_{\leq 1}) \geq C \cdot k$.*

Proof. The proof follows by applying the arguments from Lemma 4.4 to iteratively remove the variables, starting with x_r. By definition, we can express the average bias $b_{\text{av}}(P)$ as

$$
\begin{aligned}
\mathop{\mathbb{E}}_{\lambda}\left|\mathop{\mathbb{E}}_{x_1,\ldots,x_r}\left[(-1)^{\lambda^\top\left(\sum_{i\in[r]} M_{\le i}x_i\right)}\right]\right| &= \mathop{\mathbb{E}}_{\lambda}\left|\mathop{\mathbb{E}}_{x_1,\ldots,x_{r-1}}\left[(-1)^{\lambda^\top\left(\sum_{i\in[r-1]} M_{\le i}x_i\right)}\mathop{\mathbb{E}}_{x_r}(-1)^{\lambda^\top M_{\le r}x_r}\right]\right| \\
&\ge \mathop{\mathbb{E}}_{\lambda}\mathop{\mathbb{E}}_{x_1,\ldots,x_{r-1}}\left[(-1)^{\lambda^\top\left(\sum_{i\in[r-1]} M_{\le i}x_i\right)}\mathop{\mathbb{E}}_{x_r}(-1)^{\lambda^\top M_{\le r}x_r}\right] \\
&= \mathop{\mathbb{E}}_{\lambda}\mathop{\mathbb{E}}_{x_1,\ldots,x_{r-1}}\left[(-1)^{\lambda^\top\left(\sum_{i\in[r-1]} M_{\le i}x_i\right)}\ _{\{\lambda^\top M_{\le r}=0\}}\right] \\
&= \mathop{\mathbb{E}}_{x_1,\ldots,x_{r-1}}\Pr\left[\lambda\in\mathrm{Null}(M_{\le r}^\top)\right]\cdot\mathop{\mathbb{E}}_{\lambda\sim\mathrm{Null}(M_{\le r}^\top)}\left[(-1)^{\lambda^\top\left(\sum_{i\in[r-1]} M_{\le i}x_i\right)}\right] \\
&\overset{1}{\ge} 2^{-d}\mathop{\mathbb{E}}_{x_1,\ldots,x_{r-1}}\mathop{\mathbb{E}}_{\lambda\sim\mathrm{Null}(M_{\le r}^\top)}\left[(-1)^{\lambda^\top\left(\sum_{i\in[r-1]} M_{\le i}x_i\right)}\right] \\
&\overset{2}{\ge} 2^{-d}\mathop{\mathbb{E}}_{\lambda}\mathop{\mathbb{E}}_{x_1,\ldots,x_{r-1}}\left[(-1)^{\lambda^\top\left(\sum_{i\in[r-1]} M_{\le i}x_i\right)}\right]
\end{aligned}
$$

Here, in step 1 we use the observation that each of the probability terms is at least 2^{-d} by the upper bound on the rank. Step 2 uses Proposition 4.5. Rearranging, we get that

$$
\mathop{\mathbb{E}}_{\lambda}\mathop{\mathbb{E}}_{x_1,\ldots,x_{r-1}}\left[(-1)^{\lambda^\top\left(\sum_{i\in[r-1]} M_{\le i}x_i\right)}\right] \le 2^d b_{\mathrm{av}}(X_1, X_2, \ldots, X_r) \le 2^{d-k}
$$

Applying the above inequality iteratively we get that

$$
2^{-\mathrm{rank}(M_{\le 1})} \le 2^{rd-k}
$$

Now we have by assumption that $rd - k \le -C\cdot k$, to get the claim.

□

4.3 Proof of Theorem 1.2

Since $H_\infty(X) \ge Cr^2k_{\min}^2$, using Theorem 1.1 we have $b_{\mathrm{av}}(f) = 2^{-k}$ where $k \ge C'rk_{\min}$ for some constant C' that grows with C. Furthermore, we set $d = 2k_{\min}$. C is chosen such that $k \ge 2rd$. As before, we break our analysis into two cases:

- **Case (i)**: First, suppose there exists an $1 < i \le r$ such that $\Pr_{x_1,x_2,\ldots,x_{i-1}}\left[\mathrm{rank}(M_{\le i}) \ge d\right] > 0$, and let i^* be the largest such i. By the maximality of i^*, we have $\Pr_x(\mathrm{rank}(M_{\le j}) \le d) = 1$ for all $j > i^*$. Now we can upper bound the bias of the disperser as

$$
\begin{aligned}
&\left| \mathop{\mathbb{E}}_{x_1,x_2,\ldots,x_r}\left[(-1)^{\textsc{Aff}\left(\sum_{l\in[r]} M_{\le l}x_l + c\right)} \right] \right| \\
&\le \Pr_{x_1,\ldots,x_r}\left[\bigwedge_{j>i^*} M_{\le j}x_j = 0 \right] \cdot \left| \mathop{\mathbb{E}}_{x_1,\ldots,x_r}\left[(-1)^{\textsc{Aff}\left(\sum_{\ell\in[r]} M_{\le \ell}x_\ell + c\right)} \;\middle|\; \bigwedge_{j>i^*} M_{\le j}x_j = 0 \right] \right| \\
&\qquad\qquad + \Pr_{x_1,\ldots,x_r}\left[\bigvee_{j>i^*} M_{\le j}x_j \neq 0 \right] \\
&= \Pr_{x_1,\ldots,x_r}\left[\bigwedge_{j>i^*} M_{\le j}x_j = 0 \right] \cdot \left| \mathop{\mathbb{E}}_{x_1,\ldots,x_r}\left[(-1)^{\textsc{Aff}\left(\sum_{\ell\in[i^*]} M_{\le \ell}x_\ell + c\right)} \right] \right| + \Pr_{x_1,\ldots,x_r}\left[\bigvee_{j>i^*} M_{\le j}x_j \neq 0 \right]
\end{aligned}
$$

Note that $\Pr_{x_1,\ldots,x_r}\left[\bigwedge_{j>i^*} M_{\le j}x_j = 0\right] > 0$ by definition of i^*. So, we will be done if we can show that:

$$
\left| \mathop{\mathbb{E}}_{x_1,\ldots,x_r}\left[(-1)^{\textsc{Aff}\left(\sum_{\ell\in[i^*]} M_{\le \ell}x_\ell + c\right)} \right] \right| < 1.
$$

This can be argued as follows. Since $\Pr_{x_1,x_2,\ldots,x_{i^*-1}}\left[\mathrm{rank}(M_{\le i^*}) \ge d\right] > 0$, there exists an assignment $x_1 = a_1, \ldots, x_{i^*-1} = a_{i^*-1}$ such that $M_{\le i^*}$ has rank $\ge d$. Restricted to this assignment, $\sum_{\ell\in[i^*]} M_{\le \ell}x_\ell + c$ generates an affine source of dimension $\ge d > k_{\min}$. So, the above bias is strictly less than 1 because $\textsc{Aff}$ is an affine disperser for dimensions $\ge k_{\min}$.

- **Case (ii)**: In this case, for every $1 < i \le r$, we have $\Pr\left[\mathrm{rank}(M_{\le i}) \le d\right] = 1$. Then using Lemma 4.6, we get that $\mathrm{rank}(M_{\le 1}) \ge k/2 \ge 2k_{\min}$. Note that again in this case, we want to show that

  $$
  \left| \mathop{\mathbb{E}}_{x_1,x_2,\ldots,x_r}\left[(-1)^{\textsc{Aff}\left(\sum_{i\in[r]} M_{\le i}x_i + c\right)} \right] \right| < 1
  $$

  Using arguments identical to the previous case, we can iteratively remove variables $x_r, x_{r-1}, \ldots, x_2$ and reduce this to the task of arguing

  $$
  \left| \mathop{\mathbb{E}}_{x_1}\left[(-1)^{\textsc{Aff}\left(M_{\le 1}x_1 + c\right)} \right] \right| < 1
  $$

  But then, $M_{\le 1}x_1 + c$ is an affine source in x_1 of min-entropy at least $\mathrm{rank}(M_{\ge 1}) \ge 2k_{\min}$, and so, $\textsc{Aff}$ is non-constant on it. This concludes the analysis for case (ii).

Acknowledgements

We would like to thank Bharatram Rangarajan who participated in the initial discussions and without whom, this work would not have started. We would also like to thank Andrej Bogdanov and Luca Trevisan for helpful discussions. In particular, we are grateful to Andrej for allowing us to include his proof of Lemma 2.6. Finally, this work was conducted as part of the activities for the Indo-US Joint Center for research on pseudorandomness in computer science. We would like to acknowledge the Indo-US Science and Technology Forum for the support.

References

[ABN$^+$92] Noga Alon, Jehoshua Bruck, Joseph Naor, Moni Naor, and Ron M Roth. Construction of asymptotically good low-rate error-correcting codes through pseudo-random graphs. *IEEE Transactions on information theory*, 38(2):509–516, 1992.

[AGHP92] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures & Algorithms*, 3(3):289–304, 1992.

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

[BKS$^+$10] Boaz Barak, Guy Kindler, Ronen Shaltiel, Benny Sudakov, and Avi Wigderson. Simulating independence: New constructions of condensers, Ramsey graphs, dispersers, and extractors. *Journal of the ACM (JACM)*, 57(4):20, 2010.

[Blu86] Manuel Blum. Independent unbiased coin flips from a correlated biased source—a finite state markov chain. *Combinatorica*, 6(2):97–108, 1986.

[BMRV02] Harry Buhrman, Peter Bro Miltersen, Jaikumar Radhakrishnan, and Srinivasan Venkatesh. Are bitvectors optimal? *SIAM Journal on Computing*, 31(6):1723–1744, 2002.

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

[BRSW06] Boaz Barak, Anup Rao, Ronen Shaltiel, and Avi Wigderson. 2-source dispersers for sub-polynomial entropy and ramsey graphs beating the frankl-wilson construction. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 671–680. ACM, 2006.

[BSG12] Eli Ben-Sasson and Ariel 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] Eli Ben-Sasson and Swastik Kopparty. Affine dispersers from subspace polynomials. *SIAM Journal on Computing*, 41(4):880–914, 2012.

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

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

[Coh16a] Gil Cohen. Local correlation breakers and applications to three-source extractors and mergers. *SIAM Journal on Computing*, 45(4):1297–1338, 2016.

[Coh16b] Gil Cohen. Two-source dispersers for polylogarithmic entropy and improved Ramsey graphs. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 278–284. ACM, 2016.

[CZ16] Eshan Chattopadhyay and David Zuckerman. Explicit two-source extractors and resilient functions. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 670–683. ACM, 2016.

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

[Dic58] Leonard Eugene Dickson. *Linear groups: With an exposition of the Galois field theory*. Dover, 1958.

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

[DW12] Anindya De and Thomas Watson. Extractors and lower bounds for locally samplable sources. *ACM Transactions on Computation Theory (TOCT)*, 4(1):3, 2012.

[GKST16] Alexander Golovnev, Alexander S Kulikov, Alexander V Smal, and Suguru Tamaki. Circuit size lower bounds and #SAT upper bounds through a general framework. In *41st International Symposium on Mathematical Foundations of Computer Science (MFCS 2016)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2016.

[Gol95] Oded Goldreich. Three XOR-Lemmas – an exposition, 1995.

[GRS06] Ariel Gabizon, Ran Raz, and Ronen Shaltiel. Deterministic extractors for bit-fixing sources by obtaining an independent seed. *SIAM Journal on Computing*, 36(4):1072–1094, 2006.

[GT09] Ben Green and Terence Tao. The distribution of polynomials over finite fields, with applications to the gowers norms. *Contributions to Discrete Mathematics*, 4(2), 2009.

[GW97] Oded Goldreich and Avi Wigderson. Tiny families of functions with random properties: A quality-size trade-off for hashing. *Random Structures & Algorithms*, 11(4):315–343, 1997.

[Hat16] Pooya Hatami. On the structure of quintic polynomials. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2016)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2016.

[HS10] Elad Haramaty and Amir Shpilka. On the structure of cubic and quartic polynomials. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 331–340. ACM, 2010.

[Ind07] Piotr Indyk. Uncertainty principles, extractors, and explicit embeddings of l2 into l1. In *STOC*, volume 7, pages 615–620, 2007.

[KL08] Tali Kaufman and Shachar Lovett. Worst case to average case reductions for polynomials. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 166–175. IEEE, 2008.

[KRVZ06] Jesse Kamp, Anup Rao, Salil Vadhan, and David Zuckerman. Deterministic extractors for small-space sources. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 691–700. ACM, 2006.

[KZ06] Jesse Kamp and David Zuckerman. Deterministic extractors for bit-fixing sources and exposure-resilient cryptography. *SIAM Journal on Computing*, 36(5):1231–1247, 2006.

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

[Li15] Xin Li. Three-source extractors for polylogarithmic min-entropy. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 863–882. IEEE, 2015.

[Li16] Xin Li. Improved two-source extractors, and affine extractors for polylogarithmic entropy. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 168–177. IEEE, 2016.

[Li17] Xin Li. Improved non-malleable extractors, non-malleable codes and independent source extractors. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1144–1156. ACM, 2017.

[Lov08] Shachar Lovett. Unconditional pseudorandom generators for low degree polynomials. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 557–562. ACM, 2008.

[MS77] F.J. MacWilliams and N.J.A. Sloane. *The Theory of Error-correcting Codes*. North-Holland Publishing Company, 1977.

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM Journal on Computing*, 22(4):838–856, 1993.

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

[Rao07] Anup Rao. An exposition of Bourgain's 2-source extractor. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 14, 2007.

[Rao09] Anup Rao. Extractors for low-weight affine sources. In *2009 24th Annual IEEE Conference on Computational Complexity*, pages 95–101. IEEE, 2009.

[Sha11] Ronen Shaltiel. Dispersers for affine sources with sub-polynomial entropy. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*, pages 247–256. IEEE, 2011.

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

[TS02] Amnon Ta-Shma. Storing information with extractors. *Information Processing Letters*, 83(5):267–274, 2002.

[TS17] Amnon Ta-Shma. Explicit, almost optimal, epsilon-balanced codes. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 238–251. ACM, 2017.

[TSZ04] Amnon Ta-Shma and David Zuckerman. Extractor codes. *IEEE Transactions on Information Theory*, 50(12):3015–3025, 2004.

[TV00] Luca Trevisan and Salil Vadhan. Extracting randomness from samplable distributions. In *Proceedings 41st Annual Symposium on Foundations of Computer Science*, pages 32–42. IEEE, 2000.

[Vad04] Salil P Vadhan. Constructing locally computable extractors and cryptosystems in the bounded-storage model. *Journal of Cryptology*, 17(1):43–77, 2004.

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

[Vio09] Emanuele Viola. The sum of d small-bias generators fools polynomials of degree d. *Computational Complexity*, 18(2):209–217, 2009.

[vN51] John von Neumann. Various techniques used in connection with random digits. *Applied Math Series*, 12:36–58, 1951.

[WZ99] Avi Wigderson and David Zuckerman. Expanders that beat the eigenvalue bound: Explicit construction and applications. *Combinatorica*, 19(1):125–138, 1999.

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

[Zuc96] David Zuckerman. Simulating bpp using a general weak random source. *Algorithmica*, 16(4-5):367–391, 1996.