Low-Degree Polynomials Extract From Local Sources

Omar Alrabiah
EECS Department, University of California, Berkeley, CA, USA

Eshan Chattopadhyay
Computer Science Department, Cornell University, Ithaca, NY, USA

Jesse Goodman
Computer Science Department, Cornell University, Ithaca, NY, USA

Xin Li
Computer Science Department, Johns Hopkins University, Baltimore, MD, USA

João Ribeiro
Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, USA

Abstract

We continue a line of work on extracting random bits from weak sources that are generated by simple processes. We focus on the model of locally samplable sources, where each bit in the source depends on a small number of (hidden) uniformly random input bits. Also known as local sources, this model was introduced by De and Watson (TOCT 2012) and Viola (SICOMP 2014), and is closely related to sources generated by AC^0 circuits and bounded-width branching programs. In particular, extractors for local sources also work for sources generated by these classical computational models.

Despite being introduced a decade ago, little progress has been made on improving the entropy requirement for extracting from local sources. The current best explicit extractors require entropy $n^{1/2}$, and follow via a reduction to affine extractors. To start, we prove a barrier showing that one cannot hope to improve this entropy requirement via a black-box reduction of this form. In particular, new techniques are needed.

In our main result, we seek to answer whether low-degree polynomials (over $\mathbb{F}_2$) hold potential for breaking this barrier. We answer this question in the positive, and fully characterize the power of low-degree polynomials as extractors for local sources. More precisely, we show that a random degree r polynomial is a low-error extractor for n-bit local sources with min-entropy $\Omega(r(n \log n)^{1/r})$, and we show that this is tight.

Our result leverages several new ingredients, which may be of independent interest. Our existential result relies on a new reduction from local sources to a more structured family, known as local non-oblivious bit-fixing sources. To show its tightness, we prove a "local version" of a structural result by Cohen and Tal (RANDOM 2015), which relies on a new "low-weight" Chevalley-Warning theorem.

2012 ACM Subject Classification Theory of computation → Pseudorandomness and derandomization

Keywords and phrases Randomness extractors, local sources, samplable sources, AC^0 circuits, branching programs, low-degree polynomials, Chevalley-Warning

Digital Object Identifier 10.4230/LIPIcs.ICALP.2022.10

Category Track A: Algorithms, Complexity and Games

Related Version *Full Version*: `https://eccc.weizmann.ac.il/report/2022/082/`

Funding *Eshan Chattopadhyay*: Supported by NSF CAREER Award 2045576.
Jesse Goodman: Supported by NSF CAREER Award 2045576.
Xin Li: Supported by NSF CAREER Award CCF-1845349.
João Ribeiro: Research supported in part by the NSF grants CCF-1814603 and CCF-2107347 and by the following grants of Vipul Goyal: the NSF award 1916939, DARPA SIEVE program, a gift from Ripple, a DoE NETL award, a JP Morgan Faculty Fellowship, a PNC center for financial services innovation award, and a Cylab seed funding award.

49th International Colloquium on Automata, Languages, and Programming (ICALP 2022).
Editors: Mikołaj Bojańczyk, Emanuela Merelli, and David P. Woodruff;
Article No. 10; pp. 10:1–10:20

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

1 Introduction

Randomness is a fundamental resource in many areas in computer science, such as algorithm design and cryptography. However, such tasks often assume access to a source of independent and uniform bits, while real-world physical processes (e.g., electromagnetic noise, timings of user keystrokes) generate randomness that is far from perfect. This state of affairs motivates the problem of *randomness extraction*. The goal is to design a deterministic function, called an *extractor*, that can distill a (nearly) uniform bit from any source belonging to a certain family.

▶ **Definition 1.1** (Extractor). *A function* $\mathsf{Ext} : \{0,1\}^n \to \{0,1\}$ *is an* extractor *for a family of distributions* $\mathcal{X}$ *over* $\{0,1\}^n$ *with error* ε *if, for every* $\mathbf{X} \in \mathcal{X}$*,*

$$\left|\Pr[\mathsf{Ext}(\mathbf{X}) = 1] - \frac{1}{2}\right| \leq \varepsilon.$$

Besides their practical motivation, randomness extractors (and other related pseudorandom objects such as dispersers, condensers, and expander graphs) have deep connections to coding theory, combinatorics, and complexity theory.

In order to construct an extractor for a family $\mathcal{X}$ of sources, the most general assumption one can make about $\mathcal{X}$ is that each $\mathbf{X} \in \mathcal{X}$ has some "randomness." Here, it is typical to measure the randomness of a source $\mathbf{X}$ by its min-entropy $H_\infty(\mathbf{X}) := -\log \max_x \Pr[\mathbf{X} = x]$. However, even if we assume each $\mathbf{X} \in \mathcal{X}$ has a very high amount of this very strong notion of entropy, extraction is still impossible: indeed, one cannot hope to extract from $\mathcal{X}$ even if each source $\mathbf{X} \in \mathcal{X}$ is guaranteed to have min-entropy $k \geq n - 1$ [6]. To enable extraction, one must make additional assumptions on the structure of each $\mathbf{X} \in \mathcal{X}$.

Extractors for local sources, AC^0 sources, and small-space sources

In a seminal work, Trevisan and Vadhan [18] initiated the study of randomness extraction from sources that can be sampled by "simple" processes. In addition to the generality of such sources, it can be argued that they serve a reasonable model of randomness that might actually be found in nature.

More formally, Trevisan and Vadhan studied sources that can be sampled by polynomial size circuits that are given uniform bits as input. However, extracting randomness from this class of sources requires strong computational hardness assumptions. This motivated De and Watson [9] and Viola [21] to consider *unconditional* extraction from sources sampled by more restricted, but still natural, circuit families. To this end, they introduced the notion of *local sources*. Intuitively, a local source $\mathbf{X}$ is one that can be sampled by a low-depth circuit with bounded fan-in (a low-complexity process).

▶ **Definition 1.2** (Local source [9, 21]). *A distribution* $\mathbf{X} \sim \{0,1\}^n$ *is a* d-local source *if* $\mathbf{X} = g(\mathbf{U}_m)$*, where* $\mathbf{U}_m$ *is the uniform distribution over* m *bits (for some* m*), and* $g : \{0,1\}^m \to \{0,1\}^n$ *is a function where each output bit depends on at most* d *input bits.*

Local sources are closely connected to other models of sources sampled by simple processes. Viola [21] proved that every source generated by AC^0 circuits is (close to) a convex combination of local sources with small locality and slightly lower min-entropy. More recently, Chattopadhyay and Goodman [3] showed a similar result for sources generated by bounded-width branching programs [13]. Thus, extractors for local sources also work for sources generated by these classical computational models. In fact, the current state-of-the-art extractors for sources generated by AC^0 circuits and bounded-width branching programs are extractors for 1-local sources.

A barrier at $\sqrt{n}$ min-entropy

Despite the applications above and being introduced over a decade ago, little progress has been made on constructing extractors for local sources [9, 21, 16]. In particular, all known constructions require min-entropy at least $\sqrt{n}$, and follow via a reduction to extractors for affine sources (i.e., sources that are uniform over affine subspaces of $\mathbb{F}_2^n$). Thus, there appears to be a "barrier" at $\sqrt{n}$ min-entropy, at least when using affine extractors [23]. It is natural to ask how we might break this $\sqrt{n}$ barrier, which raises the question:

Can affine extractors be used to extract from local sources with min-entropy $k \ll \sqrt{n}$?

As motivation, we start by providing strong evidence that the answer to the above question is negative. In particular, we prove the following.

▶ **Theorem 0** (Barrier result). *It is not possible to extract randomness from 2-local sources with min-entropy $k \geq \sqrt{n}$ by applying an affine extractor in a black-box manner.*

Thus, if we would like to construct extractors for local sources with min-entropy significantly below $\sqrt{n}$, new techniques are needed.

Towards breaking the $\sqrt{n}$ barrier using low-degree polynomials

While explicit extractors that break the $\sqrt{n}$ min-entropy barrier for local sources are the end goal, these still seem beyond reach. We believe that the next best thing are non-explicit extractors that are of "low complexity". Our hope is that such extractors may help us eventually construct truly explicit extractors, as non-explicit extractors are more likely to be easier to derandomize if they belong to a low complexity class. At the same time, such non-explicit extractors may have applications in complexity theory (i.e., since the current state-of-the-art circuit lower bounds are against extractors [15]). There is a long line of work [20, 11, 5, 17, 19, 2, 5, 10, 12, 7] on the power of low-complexity computational models for extracting from various families of sources.

From the forefront of "low complexity" classes, we choose to study in this work the class of low-degree $\mathbb{F}_2$-polynomials. In particular, we ask whether (non-explicit) low-degree polynomials can help break the $\sqrt{n}$ barrier for extracting from local sources, and more generally we seek to answer the following question:

▶ **Question 1.** *How powerful are low-degree $\mathbb{F}_2$-polynomials as extractors for local sources?*

Beyond being a natural algebraic class, low-degree $\mathbb{F}_2$-polynomials have a natural combinatorial interpretation. We can represent a degree-2 $\mathbb{F}_2$-polynomial f as a graph G_f on n vertices, with edges representing monomials included in f. Then, f being a good extractor for local sources translates into a parity constraint on the number of edges in certain induced subgraphs of G_f. Likewise, a degree-3 $\mathbb{F}_2$-polynomial can be represented as a 3-hypergraph, and so on. Given the correspondence between low-degree polynomials and hypergraphs with small edge sizes, we hope that tools from combinatorics can be leveraged to make our constructions explicit and break the $\sqrt{n}$ min-entropy barrier for extracting from local sources (which would also give improved extractors for small-space sources).

Our motivation to study low-degree polynomials as our "low complexity" model also comes from the work of Cohen and Tal [7], which studied the same question in the context of *affine* sources. In their work, they showed that there exist degree-r $\mathbb{F}_2$-polynomials that extract from affine sources with min-entropy $O(rn^{\frac{1}{r-1}})$, and that this is tight. To answer Question 1, we aim to provide a local source analogue of this result.

1.1 Summary of our results

In this paper, we fully characterize the power of low-degree polynomials as extractors for local sources, answering Question 1 and proving a local-source analogue of Cohen-Tal. Along the way, we rely on several new ingredients which may be of independent interest. We present these results in Section 1.1.1 and Section 1.1.2, respectively.

1.1.1 Main result

Our main result gives a tight characterization of the power of low-degree polynomials as extractors for local sources. We state it formally, below.

▶ **Theorem 1** (Main result). *For every $d, r \in \mathbb{N}$ there exist constants $C, c > 0$ such that the following holds. For every $n \in \mathbb{N}$ there exists a (not necessarily explicit) degree r polynomial $f \in \mathbb{F}_2[x_1, \ldots, x_n]$ that is an $2^{-\Omega(k)}$-extractor for d-local sources with min-entropy*

$$k \geq C(n \log n)^{1/r},$$

but for every degree r polynomial $g \in \mathbb{F}_2[x_1, \ldots, x_n]$ there exists a d-local source with min-entropy

$$k \geq c(n \log n)^{1/r}$$

on which it is constant.

Theorem 1 implies that degree-3 polynomials are already enough to extract from min-entropy $k = O((n \log n)^{1/3})$, which (non-explicitly) breaks the $\sqrt{n}$ min-entropy barrier on previous techniques (Theorem 0). Furthermore, given known reductions from AC^0 sources and small-space sources to local sources [21, 3], it also follows that low-degree polynomials can be used to break existing min-entropy barriers for extracting from these other models of weak sources. We refer the reader to the full version of this paper for a more in-depth discussion, where we also outline a different application of our result to sampling lower bounds against AC^0 (generalizing a result of Viola [22]).

While Theorem 1 is stated for constant locality d and constant degree r, we actually prove stronger results that hold for superconstant d, r. In particular, Theorem 1 follows immediately from the following two results, which provide upper and lower bounds on the entropy required to extract from d-local sources using a degree $\leq r$ polynomials (where d, r need not be constant).

▶ **Theorem 1.1** (Technical version of Theorem 1, Upper Bound). *There are universal constants $C, c > 0$ such that for all $n, d, r \in \mathbb{N}$, the following holds. With probability at least 0.99 over the choice of a random degree $\leq r$ polynomial $f \in \mathbb{F}_2[x_1, \ldots, x_n]$, it holds that f is an ε-extractor for d-local sources with min-entropy*

$$k = C2^d d^2 r \cdot (2^d n \log n)^{1/r},$$

where $\varepsilon = 2^{-\frac{ck}{r^3 2^d d^2}}$.

▶ Remark 1.3. If instead of extractors we aim to construct *dispersers*,[1] then we are able to improve the dependency on the locality d in Theorem 1.1 to hold for min-entropy $k = Cd^2 r \cdot (dn \log n)^{1/r}$.

[1] A function $\mathsf{Disp} : \{0,1\}^n \to \{0,1\}$ is a *disperser for a class of sources $\mathcal{C}$* if the support of $\mathsf{Disp}(\mathbf{X})$ is $\{0,1\}$ for all sources $\mathbf{X} \in \mathcal{C}$.

▶ **Theorem 1.2** (Technical version of Theorem 1, Lower Bound). *There are constants $C, c > 0$ such that for all $n, d, r \in \mathbb{N}$ with $r \leq c\log(n)$ and $d \leq 2^{\sqrt{\log n}}$, the following holds. For any degree $\leq r$ polynomial $f \in \mathbb{F}_2[x_1, \ldots, x_n]$, there is a d-local source $\mathbf{X} \sim \{0,1\}^n$ with min-entropy at least*

$$k \geq Cr(dn\log n)^{1/r}$$

such that $f(\mathbf{X})$ is constant.

1.1.2 Key new ingredients

Our main result follows from a collection of new ingredients, which may be of independent interest. In order to prove our upper bound on min-entropy (Theorem 1.1), we prove a new reduction from d-local sources to *d-local non-oblivious bit fixing (NOBF) sources*. Informally, a d-local NOBF source $\mathbf{X} \sim \{0,1\}^n$ of min-entropy k' is a source that has k' uniform independent bits, with all other bits depending on at most d of the k' bits.

▶ **Theorem 2** (Reduction from d-local sources to d-local NOBF sources). *There exists a universal constant $c > 0$ such that for any $n, k, d \in \mathbb{N}$, the following holds. Let $\mathbf{X} \sim \{0,1\}^n$ be a d-local source with min-entropy $\geq k$. Then $\mathbf{X}$ is ε-close to a convex combination of d-local NOBF sources with min-entropy $\geq k'$, where $\varepsilon = 2^{-ck'}$ and*

$$k' = \frac{ck}{2^d d^2}.$$

The family of d-local NOBF sources, introduced in [4], is a significant specialization of d-local sources. The above reduction shows that, at least for constant locality d, we can just focus on extracting from this simpler class - even in future explicit constructions.

To prove our lower bound on min-entropy (Theorem 1.2), we actually prove this lower bound for the special class of d-local sources known as *d-local affine sources*: such a source $\mathbf{X} \sim \mathbb{F}_2^n$ is uniform over a d-local affine subspace $X \subseteq \mathbb{F}_2^n$, which is a special type of affine subspace that admits a basis $v_1, \ldots, v_k \in \mathbb{F}_2^n$ where each coordinate $i \in [n]$ holds the value 1 in at most d of these vectors. For this special class of sources, our lower bound is actually tight, and can be viewed as a "local" version of a result by Cohen and Tal [7].

▶ **Theorem 3** (Local version of Cohen-Tal). *There exist universal constants $C, c > 0$ such that for every $n, r, d \in \mathbb{N}$ such that $r \leq c\log(n)$ and $d \leq 2^{\sqrt{\log n}}$, the following holds. For any degree r polynomial $f \in \mathbb{F}_2[x_1, \ldots, x_n]$, there exists a d-local affine subspace $X \subseteq \mathbb{F}_2^n$ of dimension*

$$k \geq Cr(dn\log n)^{1/r}$$

on which f is constant.

This is tight: there exists a degree r polynomial $g \in \mathbb{F}_2[x_1, \ldots, x_n]$ which is an extractor for d-local affine sources of dimension $k \geq Cr(dn\log n)^{1/r}$, which has error $\varepsilon = 2^{-ck/r}$.

We prove Theorem 3 by extending the techniques of Cohen and Tal [7], while leveraging a key new ingredient: a "low-weight" Chevalley-Warning theorem. This result, which may be of independent interest, shows that any small system of low-degree polynomials admits a (nontrivial) solution of low Hamming weight.

▶ **Theorem 4** (Low-weight Chevalley-Warning). *Let* $\{f_i\} \subseteq \mathbb{F}_2[x_1, \ldots, x_n]$ *be a set of polynomials with degree*[2] *at most* $D < n$ *and nonlinear degree*[3] *at most* Δ *such that* 0 *is a common solution. Then there is a common solution* $x \neq 0$ *of Hamming weight*

$$w \leq 24\Delta + 2D/\log(n/D).$$

1.2 Open problems

Our work leaves open several interesting avenues for future work. We highlight three of them:

- For any constant $r \geq 2$, does there exist an explicit $\mathbb{F}_2$-polynomial of degree r that extracts from 2-local sources of min-entropy $o(n)$?
- In Theorem 2, we showed a reduction from a d-local source of min-entropy k to a d-local non-oblivious bit fixing (NOBF) source of min-entropy $\Omega(k/2^d)$. It would be interesting to show a reduction to show a reduction from a d-local source of min-entropy k to a d-local NOBF source of min-entropy $\Omega(k/\text{poly}(d))$ (or show that such a reduction is impossible).
- Theorem 4 shows that if a collection of low-degree polynomials of total degree D and nonlinear degree Δ has the zero vector as a solution, then there exists a nonzero solution of weight at most $O(\Delta + D/\log(n/D))$. When $\Delta = 0$, this becomes asymptotically tight by the Hamming bound. Moreover, when $D = \Delta$, this will also be tight by picking the polynomial $f(x) = \sum_{1 \leq |S| \leq \Delta} \prod_{i \in S} x_i$. However, if we had $\Delta/2$ quadratic polynomials, will the upper bound of $O(\Delta)$ be tight?

2 Overview of our techniques

In this section, we provide an overview of the techniques that go into our three main results:

- An entropy *upper bound* for low-degree extraction from local sources (Theorem 1.1).
- An entropy *lower bound* for low-degree extraction from local sources (Theorem 1.2).
- A *barrier* for extracting from local sources using black-box affine extractors (Theorem 0).

Along the way, we will overview the several new key ingredients (Theorems 2, 3, and 4) that go into these main results.

2.1 Upper bounds

We begin by discussing our entropy upper bounds for low-degree extraction from local sources. By this, we mean that we upper bound the entropy *requirement* for extracting from d-local sources using degree $\leq r$ polynomials. In other words, we show that low-degree polynomials extract from local sources.

2.1.1 Low-degree extractors for local sources

We start by sketching the techniques behind our main upper bound (Theorem 1.1), which shows that most degree $\leq r$ polynomials are low-error extractors for d-local sources with min-entropy at least

$$k = O(2^d d^2 r \cdot (2^d n \log n)^{1/r}).$$

[2] The *degree* D is the sum of the degrees of the f_i's.
[3] The *nonlinear degree* is the sum of the degrees of the f_i's which have degree at least 2.

A strawman application of the probabilistic method

A natural first attempt at proving our result would use a standard application of the probabilistic method, which looks something like the following. First, let $f \in \mathbb{F}_2[x_1, \ldots, x_n]$ be a uniformly random polynomial of degree $\leq r$, meaning that each monomial of size $\leq r$ is included in f with probability $1/2$. Then, we let $\mathcal{X}$ be the family of d-local sources over $\{0,1\}^n$, each with min-entropy at least k. To prove that most of these polynomials are low-error extractors for this family, a standard application of the probabilistic method would suggest that we:

1. Prove that f is an extractor for a single $\mathbf{X} \in \mathcal{X}$ with extremely high probability.
2. Show that the family $\mathcal{X}$ does not contain too many sources.
3. Conclude, via the union bound, that f is an extractor for *every* $\mathbf{X} \in \mathcal{X}$ with high probability.

It is not too hard to complete Steps 2 and 3 in the above framework, but Step 1 turns out to be much more challenging. To see why, let us consider an arbitrary d-local source $\mathbf{X} \sim \{0,1\}^n$ with min-entropy at least k. By definition of d-local source, there exists some $m \in \mathbb{N}$ and functions $g_1, \ldots, g_n : \{0,1\}^m \to \{0,1\}$ such that each g_i depends on just d of its inputs, and such that given a uniform $\mathbf{Y} \sim \{0,1\}^m$, we have

$$\mathbf{X} = (g_1(\mathbf{Y}), g_2(\mathbf{Y}), \ldots, g_n(\mathbf{Y})).$$

Now, we want to argue that a random degree $\leq r$ polynomial $f \in \mathbb{F}_2[x_1, \ldots, x_n]$ is a low-error extractor for $\mathbf{X}$. To do so, consider the function $F : \{0,1\}^m \to \{0,1\}$ defined as

$$F(y_1, \ldots, y_m) := (f \circ g)(y_1, \ldots, y_m) = f(g_1(y_1, \ldots, y_m), \ldots, g_n(y_1, \ldots, y_m)).$$

Notice that by the definition of $\mathbf{X}$ and by Definition 1.1 of extractor, we know that F is an extractor for $\mathbf{X}$ with error ε if

$$|\mathsf{bias}(F)| := \left| \Pr_{y \sim \mathbf{U}_m} [F(y) = 1] - \Pr[F(y) = 0] \right| \leq 2\varepsilon.$$

Thus, to argue that a random degree $\leq r$ polynomial $f \in \mathbb{F}_2[x_1, \ldots, x_n]$ is a low-error extractor for $\mathbf{X}$, it suffices to argue that the function $F = f \circ g$ has low bias (with high probability over the selection of f).

Of course, the question now becomes: how can we ensure that F has low bias? We can start by noticing some properties of F. First, we know $F = f \circ g$, where f is a random degree $\leq r$ polynomial and g is a fixed function where each output bit depends on $\leq d$ input bits. Thus, it is not hard to argue that F will have degree $\leq rd$. Furthermore, since f is random and g is fixed, one may hope to argue that F is a *uniformly random* polynomial of degree $\leq rd$: in this case we would be done, since it is well-known that uniformly random low-degree polynomials have extremely low bias (with extremely high probability) [1].

Unfortunately, it is too much to hope that F is a uniformly random low-degree polynomial. Indeed, it is not hard to see that the distribution of F over degree $\leq rd$ polynomials depends heavily on the exact selection of g. Furthermore, for most selections of g, the random function F is *not* uniformly distributed over degree $\leq t$ polynomials for *any* t.

Thus, there is no obvious way to apply [1] in order to argue that F will have low bias. To proceed, it seems like we will somehow need to argue that the distribution of F over low-degree polynomials is guaranteed to have some specific *structure*, and then somehow argue that a random polynomial from any such structured distribution is guaranteed to have low-bias. Each of these steps seems quite challenging.

Reductions to the rescue

As it turns out, there is a simple trick we can use to greatly simplify the above approach. The key idea is to *reduce* local sources to a simpler class of sources. Given two familes $\mathcal{X}, \mathcal{Y}$ of distributions over $\{0,1\}^n$, we say that $\mathcal{X}$ *reduces* to $\mathcal{Y}$ if each $\mathbf{X} \in \mathcal{X}$ is (close to) a convex combination of $\mathbf{Y} \in \mathcal{Y}$.[4] Reductions are extremely useful, because of the following well-known fact: if $\mathcal{X}$ reduces to $\mathcal{Y}$, and $f : \{0,1\}^n \to \{0,1\}$ is an extractor for $\mathcal{Y}$, then f *is also an extractor for* $\mathcal{X}$.

Thus, in order to show that low-degree polynomials extract from d-local sources, a key new ingredient we use is a reduction from d-local sources to a simpler class of sources called *d-local non-oblivious bit-fixing (NOBF) sources* [4]. Using the above discussion, it then suffices to show that low-degree polynomials extract from d-local NOBF sources. Thus, we proceed by:

1. Defining local NOBF sources, and showing how we can appropriately tailor our previous attempt at the probabilistic method so that it works for local NOBF sources.
2. Providing a new reduction from local sources to local NOBF sources.

Low-degree extractors for local NOBF sources

A *d-local NOBF source* $\mathbf{X} \sim \{0,1\}^n$ is a natural specialization of a d-local source where the entropic bits of the source must show up "in plain sight" somewhere in the source.[5] More formally, a d-local NOBF source with min-entropy k is a random variable $\mathbf{X} \sim \{0,1\}^n$ for which there exist functions $g_1, \ldots, g_n : \{0,1\}^k \to \{0,1\}$ such that the following holds: each $g_i, i \in [n]$ depends on $\leq d$ input bits; for every $i \in [k]$ there is some $i' \in [n]$ such that $g_{i'}(y) = y_i$; and for uniform $\mathbf{Y} \sim \{0,1\}^k$ we have

$$\mathbf{X} = (g_1(\mathbf{Y}), g_2(\mathbf{Y}), \ldots, g_n(\mathbf{Y})).$$

In other words, some k "good" bits in $\mathbf{X}$ are uniform, and the remaining $n-k$ "bad" bits are d-local functions of the good bits.

We must now show that a random degree $\leq r$ polynomial $f \in \mathbb{F}_2[x_1, \ldots, x_n]$ extracts from d-local NOBF sources with entropy k. As in our strawman application of the probabilistic method, consider an arbitary d-local NOBF source $\mathbf{X} = (g_1(\mathbf{Y}), \ldots, g_n(\mathbf{Y}))$ and let $f \in \mathbb{F}_2[x_1, \ldots, x_n]$ be a uniformly random degree $\leq r$ polynomial. To show that f extracts from all d-local NOBF sources, recall that we just need to show that the function $F : \{0,1\}^k \to \{0,1\}$ defined as

$$F(y) := (f \circ g)(y) = f(g_1(y), \ldots, g_n(y))$$

has extremely low bias with extremely high probability. Furthermore, recall that if we can show that F itself is a uniformly random low-degree polynomial, then we know via [1] that this is true.

It is still too much to hope that F is a uniform low-degree polynomial, but F is now "close enough in structure" to one so that we can make this work. To see why, we can first assume without loss of generality (by definition of local NOBF source) that $g_1(y) = y_1, \ldots, g_k(y) = y_k$. Thus, we can define $c_S \sim \{0,1\}$ as an independent uniform bit (for each $S \subseteq [n]$ of size $\leq r$) and write

[4] By this we mean that each $\mathbf{X} \in \mathcal{X}$ can be written in the form $\mathbf{X} = \sum_i p_i \mathbf{Y}_i$, where each $\mathbf{Y}_i \in \mathcal{Y}$, $\sum_i p_i = 1$, and $\mathbf{X}$ samples from $\mathbf{Y}_i$ with probability p_i.

[5] The relationship between local sources and local *NOBF* sources is not dissimilar to the relationship between error-correcting codes and *systematic* error-correcting codes.

$$F(y) = \sum_{S\subseteq[k]:|S|\leq r} c_S \prod_{i\in S} g_i(y) + \sum_{S\subseteq[n]:|S|\leq r, S\not\subseteq[k]} c_S \prod_{i\in S} g_i(y) = A(y) + B(y),$$

where $A \in \mathbb{F}_2[y_1, \ldots, y_k]$ is a uniformly random polynomial of degree $\leq r$, and $B \in \mathbb{F}_2[y_1, \ldots, y_k]$ is a random polynomial whose selection of monomials is *not uniformly random*, but is nevertheless *independent* of the selections made by A.

Thus, to show that F has extremely low bias, it suffices to show that $A+B$ has extremely low bias. And to show that $A+B$ has extremely low bias, it suffices to show that $A+B'$ has low bias for any fixed polynomial B' induced by fixing the random monomials selected by B.

To conclude, we actually show something stronger: recalling that $A \in \mathbb{F}_2[y_1, \ldots, y_k]$ is a uniformly random polynomial of degree $\leq r$, we show that: for *any* fixed function $B^* : \{0,1\}^k \to \{0,1\}$, it holds that $A + B^*$ has low bias. This does not follow immediately from the result [1] that a random low-degree polynomial has low bias: indeed, it is more general, since [1] is the special case where $B^* = 0$. However, it *does* follow immediately from known upper bounds on the list size of Reed-Muller code [14].

In the language we are using here, an upper bound on the list size of a Reed-Muller code is equivalent to saying that for any fixed function $D \in \mathbb{F}_2[x_1, \ldots, x_k]$, A will differ from D on many inputs, with very high probability. Thus, such bounds tell us that A differs from B^* on many inputs with very high probability, and A differs from $1 + B^*$ (or rather, *equals* B^*) on many inputs with very high probability. In other words, A is completely uncorrelated with B^*, meaning that $\mathsf{bias}(A + B^*)$ is extremely small, as desired.

Thus a random low-degree polynomial f extracts from the d-local NOBF source $\mathbf{X}$ with min-entropy k with very high probability. In other words, it fails to do so with some very small probability $\delta = \delta(k)$ which decreases rapidly as k grows. By applying the union bound, we get that f extracts from the entire family $\mathcal{X}$ of d-local NOBF sources, provided $\delta(k) \cdot |\mathcal{X}| \ll 1$. All that remains is to upper bound the size of $\mathcal{X}$, which can easily be done using the d-locality of the sources.

A reduction to local NOBF sources

Above, we saw that random low-degree polynomials extract from *local NOBF sources*. To complete the proof that they also extract from more general *local sources*, recall that we need to provide a reduction from local sources to local NOBF sources. In other words, we need to show that every d-local source with min-entropy k is (close to) a convex combination of d-local NOBF sources with min-entropy $k' \approx k$. This is the main key ingredient in our result that low-degree polynomials extract from local sources (Theorem 1.1).

Our reduction works as follows. First, pick an arbitrary d-local source $\mathbf{X} \sim \{0,1\}^n$ with min-entropy k. Let k' be a parameter which is slightly smaller than k, which will be picked later. We start by arguing that $\mathbf{X}$ is (close to) a convex combination of d-local NOBF sources where there are k' good bits, but the good bits may be biased (but not constant).

Towards this end, recall that $\mathbf{X} = (g_1(\mathbf{Y}), \ldots, g_n(\mathbf{Y}))$ for some d-local functions $g_1, \ldots, g_n : \{0,1\}^m \to \{0,1\}$ and uniform $\mathbf{Y} \sim \{0,1\}^m$. The key idea is to consider the *largest possible set* $T \subseteq [n]$ of "good bits," i.e., such that $\{\mathbf{X}_i\}_{i\in T}$ are independent (and none are constants). Then, we let $T' \subseteq [m]$ be the bits of $\mathbf{Y}$ on which $\{\mathbf{X}_i\}_{i\in T}$ depend. The key observation is that *every* bit in $\mathbf{X}$ depends on *some* bit in $\{\mathbf{Y}_i\}_{i\in T'}$, by the maximality of T. Using this observation, there are *two possible cases*, over which we perform a *win-win analysis*.

First, it is possible that T contains $\geq k'$ bits. In this case, we consider fixing all bits $\{\mathbf{Y}_i\}_{i \notin T'}$. It is then not too hard to show that $\mathbf{X}$ becomes a source which contains $\geq k'$ good bits (which are mutually independent and not constants), and the remaining bad bits in $\mathbf{X}$ are deterministic d-local functions of these good bits.[6] Thus in this case, we get that $\mathbf{X}$ is a convex combination of NOBF sources of the desired type.

Second, it is possible that T contains $< k'$ bits. In this case, we consider fixing all bits $\{\mathbf{Y}_i\}_{i \in T'}$. But since all bits in $\mathbf{X}$ depend on *some* bit in this set, this fixing *decrements* the locality $d \to d-1$. And furthermore, since this fixes $|T'| \leq d|T| < dk'$ bits, the entropy only decreases from $k \to k - dk'$ by the entropy chain rule. We then recurse until we hit the first case, or until we hit $d = 1$. If we eventually hit the first case, we already know that $\mathbf{X}$ is a convex combination of NOBF sources of the desired type. On the other hand, it is easy to show that a 1-local source is actually a 1-local NOBF source! Thus we will always arrive at a (biased) d'-local NOBF source with $d' \leq d$, proving that $\mathbf{X}$ is always convex combination of NOBF sources of the desired type. Depending on when this recursion stops, we will arrive at an NOBF source with the number of good bits equal to at least

$$\min\{k', k - dk', k - d(d-1)k', \ldots, k - k' \prod_{i \in [d]} i\} \geq \min\{k', k - d^2 k'\},$$

which is always at least k' provided $k' \leq \frac{k}{2d^2}$.

Thus we see that any d-local source with min-entropy k can be written as a convex combination of d-local NOBF source with $\Omega(k/d^2)$ good bits, where the good bits are mutually independent (and nonconstant), but they may be heavily biased. So all that remains is to show that such biased d-local NOBF sources can be written as a convex combination of *unbiased* d-local NOBF sources (as they were originally defined). This step is not difficult, by applying a standard Chernoff bound. However, since each good bit depends on up to d bits, each such good bit $\mathbf{X}_i$ may have $|\mathsf{bias}(\mathbf{X}_i)| = 1 - 2 \cdot 2^{-d}$. As a result, we end up with $\Omega(\frac{k}{d^2 2^d})$ unbiased good bits.

This completes the reduction from local to local NOBF sources. Given our earlier proof sketch that low-degree polynomials extract from local NOBF sources, we finally get that low-degree polynomials also extract from local sources, as desired.

2.1.2 Low-degree dispersers for local sources

We now proceed to sketch the proof of Remark 1.3, which shows that most degree $\leq r$ polynomials are dispersers for d-local sources with min-entropy at least

$$k = O(d^2 r \cdot (dn \log n)^{1/r}).$$

This improves our min-entropy requirement for extractors (which was $k = O(2^d d^2 r \cdot (2^d n \log n)^{1/r})$) by removing two terms of the form 2^d. We use a different key idea to remove each 2^d term. While the outer exponential term 2^d is the more dramatic one to remove, it turns out that it is also the easier one. To do this, we simply note that for dispersers, we can forego the last step in our local to local NOBF reduction, which incurs a factor of 2^d by making the biased local NOBF source into an unbiased one. This improves the entropy requirement for dispersers from

$$k = O(2^d d^2 r \cdot (2^d n \log n)^{1/r}) \to k = O(d^2 r \cdot (2^d n \log n)^{1/r}).$$

[6] Technically we need to fix a little more randomness to make this happen, but this can be done without much trouble by invoking some standard tricks from the extractor literature.

Removing the inner exponential term

Next, we focus on improving the entropy requirement for dispersers from

$$k = O(d^2 r \cdot (2^d n \log n)^{1/r}) \to k = O(d^2 r \cdot (dn \log n)^{1/r}),$$

turning the inner exponential term 2^d into d. This improvement is more challenging: while our first improvement relied on improving the local to local-NOBF reduction, this improvement relies on improving the entropy requirement for dispersing from local NOBF sources.

In order to show that low-degree polynomials extract from local NOBF sources, recall that we: (1) showed that a random low-degree polynomial extracts from an arbitrary local NOBF source with extremely high probability; and (2) used a union bound over the family $\mathcal{X}$ of local NOBF sources to conclude that it extracts from *all* local NOBF sources with high probability. To get our second improvement on the min-entropy requirement for dispersers, we get improved upper bounds on the size of $\mathcal{X}$, so that our union bound is over fewer terms.

Towards this end, the key new idea is to show that in order to disperse from the family of d-local NOBF sources $\mathcal{X}$, it actually suffices to disperse from the much smaller family $\mathcal{X}'$ of so-called *d-local, degree $\leq r$ NOBF sources.*[7] This source family is the exact same as d-local NOBF sources, except it has the added restriction that the bad bits (which still depend on $\leq d$ good bits each) can each be written as a degree $\leq r$ polynomials. However, this family is significantly smaller: natural estimates on the sizes of $\mathcal{X}, \mathcal{X}'$ give

$$|\mathcal{X}| \leq \binom{n}{k} \cdot \left(\binom{k}{d} \cdot 2^{2^d} \right)^{n-k},$$
$$|\mathcal{X}'| \leq \binom{n}{k} \cdot \left(\binom{k}{d} \cdot 2^{\binom{d}{\leq r}} \right)^{n-k}.$$

After plugging in these improved size bounds, it is straightforward calculation to see that the inner 2^d term from the entropy requirement drops out. So all that remains is to show the above claim that a disperser for d-local, degree $\leq r$ NOBF sources automatically works for the more general family of d-local NOBF sources.

The key ingredient that goes into this claim is a simple lemma on polynomial decomposition. We show the following: for any function $f : \{0,1\}^n \to \{0,1\}$, any degree $\leq r$ polynomials $a_1, \ldots, a_n : \{0,1\}^k \to \{0,1\}$, and any polynomials $b_1, \ldots, b_n : \{0,1\}^k \to \{0,1\}$ that have no monomials of size $\leq r$, the following holds. There exists a polynomial $h : \{0,1\}^k \to \{0,1\}$ with no monomials of size $\leq r$ such that

$$f(a_1(y) + b_1(y), \ldots, a_n(y) + b_n(y)) = f(a_1(y), \ldots, a_n(y)) + h(y).$$

Then, given an arbitrary d-local NOBF source $\mathbf{X} \sim \{0,1\}^n$, the idea is to write it in the form

$$\mathbf{X} = (a_1(\mathbf{Y}) + b_1(\mathbf{Y}), \ldots, a_n(\mathbf{Y}) + b_n(\mathbf{Y})),$$

where $\mathbf{Y} \sim \{0,1\}^k$ is uniform and a_i, b_i are as before. Using our polynomial decomposition lemma, it then (roughly) holds that f is a disperser for $\mathbf{X}$ if f is a disperser for the simpler class of d-local, degree $\leq r$ NOBF sources. More precisely, we actually require from f a property that is ever-so-slightly stronger than being a disperser: we require that for any d-local, degree $\leq r$ NOBF source $\mathbf{X}' = (a_1(\mathbf{Y}), \ldots, a_n(\mathbf{Y}))$, it holds that the polynomial

[7] Technically, we require something slightly stronger than dispersion from such sources, but this does not make a huge difference. We will go into more details below.

$f(a_1, \ldots, a_n)$ has a monomial of degree $\leq r$. Our polynomial decomposition lemma then guarantees that f will also have this property for the more general d-local source, since the polynomial h in our decomposition lemma does not have any monomials of degree $\leq r$. Intuitively, h is not able to "destroy" the monomial of degree $\leq r$ guaranteed to pop out of $f(a_1(y), \ldots, a_n(y))$.

Thus, it suffices to "disperse" from d-local, degree $\leq r$ NOBF sources in order to disperse from more general d-local NOBF sources, meaning that we can leverage our improved bound on the size of $\mathcal{X}'$ to get our claimed improvement on the disperser's entropy requirement.

2.2 Lower bounds

We now discuss our entropy lower bounds for low-degree extraction from local sources. By this, we mean that for every degree $\leq r$ polynomial $f \in \mathbb{F}_2[x_1, \ldots, x_n]$ we can find a d-local source $\mathbf{X} \sim \{0,1\}^n$ with relatively high min-entropy k on which f is constant. In other words, we show that in order to disperse (and thus extract) from d-local sources, they must have min-entropy exceeding this value k.

We show that every degree $r \leq \Omega(\log n)$ polynomial $f \in \mathbb{F}_2[x_1, \ldots, x_n]$ must admit a d-local source $\mathbf{X} \sim \{0,1\}^n$ of min-entropy at least

$$k = \Omega(r(dn \log n)^{1/r})$$

on which it is constant. That is, we sketch the proof of our lower bound (Theorem 1.2).

In order to prove this result, we actually prove a slightly stronger result: we show that we can find a d-local source $\mathbf{X} \sim \{0,1\}^n$ with the above parameters such that it is also *affine*.

Our starting point is a tight result of Cohen and Tal [7], which shows that any degree $\leq r$ polynomial $f : \mathbb{F}_2^n \to \mathbb{F}_2$ admits a subspace $V \subseteq \mathbb{F}_2^n$ of dimension $\Omega(rn^{1/(r-1)})$ on which it is constant. Here, we obtain a (tight) granular version of their result, and show that any degree $\leq r$ polynomial f admits a d-local subspace $X \subseteq \mathbb{F}_2^n$ of dimension $k = \Omega(r(dn \log n)^{1/r})$ on which it is constant. Here, we say that V is *d-local* if V has a basis $v_1, \ldots, v_k \in \mathbb{F}_2^n$ such that for any index $i \in [n]$, at most d of these basis vectors equal 1 at this index. It is straightforward to verify that the uniform distribution $\mathbf{X}$ over V is a d-local source with min-entropy k, so we focus now on proving the existence of such a V.

At a high level, the proof of Cohen and Tal proceeds by iteratively growing a subspace V on which f is constant. At each phase, they define a set $A \subseteq \mathbb{F}_2^n$ such that f is constant over $\mathsf{span}(V, x)$ for every $x \in A$. They note that if $|A|$ has size $> 2^{\mathsf{dim}(V)}$, then of course there is some $x \in A \setminus V$ and furthermore we already know that f is constant on $\mathsf{span}(V, x)$. Thus, they can grow their monochromatic subspace by one dimension.

In order to get a lower bound on $|A|$, they note that this set can be defined as the common solutions to a small collection of low-degree polynomials. A classical result known as the *Chevalley-Warning theorem* (Theorem 5.1) then shows that $|A| \geq 2^{n-t}$, where t is the sum of degrees across the collection of polynomials. To complete their proof, they grow their subspace V until they are no longer able to show $|A| > 2^{\mathsf{dim}(V)}$.

In our lower bound, we show that f is monochromatic on a *d-local subspace*. To prove this, we start with the same approach as Cohen and Tal. However, at each phase, we add extra constraints to A which guarantee the following: if we take any $x \in A$ and add it to our current subspace V (with basis, say, $v_1, \ldots, v_{|\dim(V)|}$), then the location of the 1s appearing in vectors $x, v_1, \ldots, v_{|\dim(V)|}$ satisfy the d-locality constraint defined above. Again, as long as A is large enough, we can find some $x \in A$ that grows the dimension of our d-local subspace.

In order to ensure that A remains large for as many iterations as possible, we would like to minimize the impact of the new "locality" constraints that we have added to A. Given the description of these constraints above, we observe that these constraints are minimized if we grow V by carefully selecting vectors that have the lowest possible Hamming weight. However, we now need an upper bound on the Hamming weight of the lightest (nontrivial) common solution to a system of polynomial equations. Thus, our key new ingredient will be a result of this type, which we call a "low-weight Chevalley-Warning theorem."

A low-weight Chevalley-Warning theorem

Above, we saw how the classical Chevalley-Warning theorem is critical in lower bounding the size of A, thereby showing that there is some (nontrivial) vector $v \in A$ by which we can grow our monochromatic subspace. Now, we need an additional guarantee that there is such a $v \in A$ that also has low Hamming weight. We prove such a result, and call it a *low-weight Chevalley-Warning theorem*. Our theorem roughly says the following. Given a collection of polynomials that have cumulative degree D (and a common solution 0), if most of these polynomials have degree ≤ 1 then they admit a nontrivial solution of Hamming weight

$$w \leq O(D/\log(n/D)).$$

In order to prove our result, we start by using the CLP lemma [8] (a result which was instrumental in the recent resolution of the cap set conjecture) in order to show that for any large enough set $A \subseteq \mathbb{F}_2^n$ of common solutions to a system of polynomial equations, it holds that $A + A$ also contains a (nontrivial) common solution to this system. We then combine this result with an argument that is similar in flavor to the classical proof of the Hamming bound, and thereby obtain our low-weight Chevalley-Warning theorem. Equipped with this new ingredient, our entropy lower bound for low-degree extraction from d-local affine spaces follows immediately via the proof sketch described above.

2.3 A barrier

To conclude our overview, we provide a proof sketch of our barrier result (Theorem 0), which shows that affine extractors (applied in a black-box manner) cannot extract from local sources with min-entropy $k = \Omega(\sqrt{n})$, even if the locality is 2. More formally, to show that an affine extractor also extracts from a different family $\mathcal{Q}$ of distributions with min-entropy k, the standard technique is to show that each source $\mathbf{Q} \in \mathcal{Q}$ with min-entropy k is (close to) a convex combination of affine sources with min-entropy slightly less than k. Here, we show that this is simply not possible for local sources with min-entropy $\sqrt{n}$. In particular, we show that there is a very simple 2-local source $\mathbf{Q} \sim \{0,1\}^n$ with min-entropy $\Omega(\sqrt{n})$ that has statistical distance *exponentially close to* 1 from any convex combination of affine sources with min-entropy k'.

In more detail, we consider the 2-local "clique" source $\mathbf{Q} \sim \{0,1\}^n$ defined as follows: first, pick any $\ell \in \mathbb{N}$ and set $n := \ell + \binom{\ell}{2}$. Then, pick uniform and independent bits $\mathbf{q}_1, \ldots, \mathbf{q}_\ell \sim \{0,1\}$ and set $\mathbf{Q}$ to be the concatenation of all $\mathbf{q}_i$ over $1 \leq i \leq n$ and $\mathbf{q}_i \cdot \mathbf{q}_j$ over $1 \leq i < j \leq \ell$. Now, let $\mathbf{X} \sim \{0,1\}^n$ be a convex combination of affine sources, each with min-entropy ℓ'. We argue that $|\mathbf{Q} - \mathbf{X}| \geq 1 - 2^{-\Omega(\ell')}$ by showing that for any ℓ'-dimensional affine $\mathbb{F}_2$-subspace $S \subseteq \{0,1\}^n$, it holds that $|\mathsf{supp}(\mathbf{Q}) \cap S|/|S| \leq 2^{-\Omega(\ell')}$. That is, we wish to show that $Q = \mathsf{supp}(\mathbf{Q})$ is *subspace-evasive*.

To show that cliques are subspace-evasive, we use the following key observation: For any nonempty set $Q' \subseteq Q$ of cliques, the set $Q' + Q' := \{u + v : u \in Q', v \in Q', u \neq v\}$ (where the sum is over $\mathbb{F}_2^n$) has a "Sidon property:" each element x in $Q' + Q'$ has a unique

pair $u, v \in Q'$ such that $x = u + v$. This observation is proven by noticing that by making another copy of each coordinate of the form $\mathbf{q}_i \cdot \mathbf{q}_j$, the set Q will correspond precisely to the symmetric rank-1 matrices of $\mathbb{F}_2^{\ell \times \ell}$. Thus all the elements in $Q' + Q'$ would correspond to symmetric $\mathbb{F}_2^{\ell \times \ell}$ matrices of rank at most 2. Hence by looking at the row space of $x \in Q' + Q'$, we can precisely find its symmetric rank-1 decomposition. That is, we can find $u, v \in Q'$ such that $x = u + v$. Now, pick $Q' = Q \cap S$. Since addition is closed in S, we see that $Q' + Q' \subseteq S$. Thus $|S| \geq |Q' + Q'| \geq \binom{|Q'|}{2} \geq \Omega(|Q \cap S|^2)$. Hence we find that $|Q \cap S|/|S| \leq O(\sqrt{|S|}/|S|) = O(1/\sqrt{|S|})$, which is $2^{-\Omega(\ell')}$ as $|S| = 2^{\ell'}$.

For more details, we refer the reader to the full version of this paper.

3 Preliminaries

We briefly outline some basic notation, definitions, and facts that will be used throughout the paper. We first discuss some notation. We use log to denote the base-2 logarithm, we define $[n] := \{1, 2, \ldots, n\}$, and we write $\binom{n}{\leq r} := \sum_{i=0}^{r} \binom{n}{i}$. Given a random variable $\mathbf{X}$, we let $\mathsf{supp}(\mathbf{X})$ denote its support and write $\mathbf{X} \sim S$ to denote that $\mathsf{supp}(\mathbf{X}) \subseteq S$. Finally, we use $\mathbf{U}_m$ to denote the uniform distribution over $\{0,1\}^m$.

We now discuss some basic notions from probability. First, the *statistical distance* between two random variables $\mathbf{X}, \mathbf{Y} \sim S$ is denoted by $\Delta(\mathbf{X}, \mathbf{Y})$ and defined as $\Delta(\mathbf{X}, \mathbf{Y}) := \max_{T \subseteq S} |\Pr[\mathbf{X} \in T] - \Pr[\mathbf{Y} \in T]| = \frac{1}{2} \sum_{s \in S} |\Pr[\mathbf{X} = s] - \Pr[\mathbf{Y} = s]|$. Moreover, we say $\mathbf{X}$ and $\mathbf{Y}$ are *ε-close*, denoted $\mathbf{X} \approx_\varepsilon \mathbf{Y}$, if $\Delta(\mathbf{X}, \mathbf{Y}) \leq \varepsilon$. Finally, recalling the definition of min-entropy from the introduction, we will use the following simple fact about this quantity. The proof is straightforward, and can be found in the full version.

▶ **Lemma 3.1.** *Suppose $\mathbf{X}$ and $\mathbf{Y}$ are arbitrary random variables such that $\mathbf{Y}$ is uniformly distributed over its support. Then, for every $y \in \mathsf{supp}(\mathbf{Y})$, it holds that*

$$H_\infty(\mathbf{X}|\mathbf{Y} = y) \geq H_\infty(\mathbf{X}) - \log |\mathsf{supp}(\mathbf{Y})|.$$

4 Entropy upper bounds

In this section, we obtain upper bounds on the entropy required to extract from d-local sources using degree $\leq r$ polynomials (Theorem 1.1). To do so, we combine two key ingredients. Our first key ingredient reduces d-local sources to d-local NOBF sources:

▶ **Theorem 4.1** (Theorem 2, restated)**.** *There exists a universal constant $c > 0$ such that for any $n, k, d \in \mathbb{N}$, the following holds. Let $\mathbf{X} \sim \{0,1\}^n$ be a d-local source with min-entropy $\geq k$. Then $\mathbf{X}$ is ε-close to a convex combination of d-local NOBF sources with min-entropy $\geq k'$, where $\varepsilon = 2^{-ck'}$ and*

$$k' = \frac{ck}{2^d d^2}.$$

Our second key ingredient gives upper bounds on the entropy required to extract from d-local NOBF sources using degree $\leq r$ polynomials.

▶ **Theorem 4.2** (Low-degree polynomials extract from d-local NOBF sources)**.** *There are universal constants $C, c > 0$ such that for all $n, d, r \in \mathbb{N}$, the following holds. With probability at least 0.99 over the choice of a random degree $\leq r$ polynomial $f \in \mathbb{F}_2[x_1, \ldots, x_n]$, it holds that f is an ε-extractor for d-local NOBF sources of min-entropy*

$$k = Cr \cdot (2^d \cdot n \log n)^{1/r}$$

with error $\varepsilon = 2^{-ck/r^3}$.

By combining the above two theorems, we immediately get Theorem 1.1. Moreover, if we only care about *dispersing* (instead of *extracting*), it turns out that we can streamline our arguments and remove the 2^d terms in Theorems 4.1 and 4.2, yielding Remark 1.3.

In the remainder of this section, we focus on proving our reduction, Theorem 4.1. But before doing so, we briefly note that to prove Theorem 4.2, the main idea is to show that a random low-degree polynomial has low correlation with any fixed function. This observation follows quite readily from known bounds [14] on the list-size of Reed-Muller codes, and Theorem 4.2 is not too hard to show given this observation. For a formal proof of Theorem 4.2, we refer the reader to Section 2 and the full version of this paper, where one can also find more details regarding Remark 1.3.

4.1 A reduction from d-local sources to d-local NOBF sources

We now turn towards proving our reduction, Theorem 4.1. In order to reduce d-local sources to d-local NOBF sources, we use an intermediate model called a *biased d-local NOBF source*.

▶ **Definition 4.3** (Biased local NOBF sources). *A random variable $\mathbf{X} \sim \{0,1\}^n$ is a (δ, k)-*biased d-local NOBF source *if there exists a set $S \subseteq [n]$ of size k such that both of the following hold:*

- *The bits in $\mathbf{X}_S$ are mutually independent (but need not be identically distributed), and each $\mathbf{X}_i, i \in S$ has bias $|\Pr[\mathbf{X}_i = 1] - \Pr[\mathbf{X}_i = 0]| \leq \delta$.*
- *Every other bit $\mathbf{X}_j, j \notin S$ is a deterministic function of at most d bits in $\mathbf{X}_S$.*

▶ Remark 4.4. A d-local NOBF source with entropy k is a $(0, k)$-biased d-local NOBF source.

Given this intermediate model, we prove Theorem 4.1 by combining two lemmas. The first lemma reduces d-local sources to biased d-local NOBF sources:

▶ **Lemma 4.5.** *Let $\mathbf{X} \sim \{0,1\}^n$ be a d-local source with min-entropy $\geq k$. Then $\mathbf{X}$ is a convex combination of (δ, k')-biased d-local NOBF sources, where $\delta \leq 1 - 2^{-d}$ and $k' \geq k/(2d^2)$.*

The second lemma reduces biased d-local NOBF sources to (unbiased) d-local NOBF sources.

▶ **Lemma 4.6.** *Let $\mathbf{X} \sim \{0,1\}^n$ be a (δ, k')-biased d-local NOBF source. Then $\mathbf{X}$ is ε-close to a convex combination of $(0, k'')$-biased d-local NOBF sources, where $k'' \geq (1-\delta)k'/4$ and $\varepsilon = 2^{-k''/4}$.*

By combining these two lemmas, Theorem 4.1 follows immediately. Lemma 4.6 is not too difficult to prove by simulating each biased bit with two consecutive independent coin flips (one with bias $|1 - 2\delta|$ and one with bias 0) and applying a standard Chernoff bound over the result of the first coin flip. We refer to the full version for more details, and conclude this section by proving Lemma 4.5.

Proof of Lemma 4.5. Let $\mathbf{X} \sim \{0,1\}^n$ be a d-local source with min-entropy $\geq k$. The key observation that we will prove is that for any t, one of the following *must* hold: either

- $\mathbf{X}$ is a convex combination of (δ, t)-biased d-local NOBF sources, for $\delta \leq 1 - 2^{-d}$; or
- $\mathbf{X}$ is a convex combination of $(d-1)$-local sources with min-entropy $> k - td$.

Before we prove this key observation, let us see how we can use it to prove the desired result. First, recall that convex combinations "stack" in the following sense: if a source $\mathbf{X}$ is a convex combination of convex combinations of sources from a family $\mathcal{X}$, then $\mathbf{X}$ is just a convex combination of sources from $\mathcal{X}$. Thus, by repeatedly applying the key observation until either the first item becomes true or we arrive at a 1-local source (the "base case"), we see that $\mathbf{X}$ is a convex combination of sources $\{\mathbf{Z}_i\}$, where each $\mathbf{Z}_i$ is either:

- A (δ, t)-biased d-local NOBF source, for $\delta \leq 1 - 2^{-d}$; or
- A 1-local source with min-entropy $> k - t \cdot (d + (d-1) + \cdots + 2) = k - t \cdot (d^2 + d - 2)$.

However, it is clear from the definitions that a 1-local source with min-entropy k' is a 1-local NOBF source with min-entropy k'. Furthermore, it is easy to see that a 1-local NOBF source with min-entropy $\geq k'$ is a convex combination of 1-local NOBF sources with min-entropy exactly k', by fixing any additional random "good" bits. Thus, for any $t \leq k'$, we know that a 1-local source with min-entropy $\geq k'$ is a convex combination of (δ, t)-biased d-local NOBF sources, for $\delta \leq 1 - 2^{-d}$.

By the above discussion, we see that for any $t \leq k - t \cdot (d^2 + d - 2)$, $\mathbf{X}$ is a convex combination of (δ, t)-biased d-local NOBF sources, where $\delta \leq 1 - 2^{-d}$. Setting $t = \frac{k}{2d^2}$ yields the result.

Thus, all that remains is to prove the key observation stated at the beginning of the proof. Towards this end, let $\mathbf{X} \sim \{0,1\}^n$ be a d-local source with min-entropy $\geq k$. By definition of d-local source, there exists some ℓ and $f : \{0,1\}^\ell \to \{0,1\}^n$ such that $\mathbf{X} = f(\mathbf{Y})$ for uniform $\mathbf{Y} \sim \mathbf{U}_\ell$, such that each bit $\mathbf{X}_i$ is a deterministic function of at most d bits in $\mathbf{Y}$. In other words, there exist sets $S_1, \ldots, S_n \subseteq [\ell]$ of size d and functions $f_1, \ldots, f_n : \{0,1\}^d \to \{0,1\}^n$ such that

$$\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n) = (f_1(\mathbf{Y}_{S_1}), f_2(\mathbf{Y}_{S_2}), \ldots, f_n(\mathbf{Y}_{S_n})).$$

Now, let $T \subseteq [n]$ be any set of coordinates of *maximal size* such that:

- $H_\infty(\mathbf{X}_i) > 0$ for all $i \in T$; and
- $S_i \cap S_j = \emptyset$ for any distinct $i, j \in T$.

Suppose T has size τ. Without loss of generality, assume $T = [\tau]$. We conclude with two cases.

Case (i): $\tau < t$. In this case, we fix the random variable $\mathbf{Y}_{S_1}, \ldots, \mathbf{Y}_{S_\tau}$. We know that with probability 1 over this fixing, all bits $\mathbf{X}_i, i \in [n]$ become deterministic functions of at most $d-1$ unfixed variables in $\mathbf{Y}$, by the maximality of T and its intersection property. In other words, $\mathbf{X}$ becomes a $(d-1)$-local source. Furthermore, by Lemma 3.1, we know that with probability 1 over this fixing, $\mathbf{X}$ loses $\sum_{i \in [\tau]} |S_i| = d\tau < dt$ bits of min-entropy. Thus in this case, $\mathbf{X}$ is a convex combination of $(d-1)$-local sources of min-entropy $> k - dt$.

Case (ii): $\tau \geq t$. In this case, define $\overline{S} := [n] - (\bigcup_{i \in [\tau]} S_i)$ and notice that $S_1, S_2, \ldots, S_\tau, \overline{S}$ partition the coordinates of $\mathbf{Y}$. Next, define the random variables $\mathbf{Z}_i := \mathbf{Y}_{S_i}$ for each $i \in [\tau]$, and define $\overline{\mathbf{Z}} := \mathbf{Y}_{\overline{S}}$. Notice that $\mathbf{X}_i = f_i(\mathbf{Z}_i)$ for each $i \in [\tau]$. Furthermore, it is straightforward to verify that for all $j > \tau$, there exists a set $Q_j \subseteq [\tau]$ of size at most d and a deterministic function f'_j such that $\mathbf{X}_j = f'_j(\mathbf{Z}_{Q_j}, \overline{\mathbf{Z}})$. In other words, we can rewrite $\mathbf{X}$ as

$$\begin{aligned}\mathbf{X} &= (\mathbf{X}_1, \ldots, \mathbf{X}_\tau, \mathbf{X}_{\tau+1}, \ldots, \mathbf{X}_n) \\ &= (f_1(\mathbf{Z}_1), \ldots, f_\tau(\mathbf{Z}_\tau), f'_{\tau+1}(\mathbf{Z}_{Q_{\tau+1}}, \overline{\mathbf{Z}}), \ldots, f'_n(\mathbf{Z}_{Q_n}, \overline{\mathbf{Z}})).\end{aligned}$$

Now, for each $i \in [\tau]$, define $\mathbf{A}_i := f_i(\mathbf{Z}_i)$. Furthermore, it is straightforward to show that we can define a new random variable $\mathbf{B}$ independent of $\mathbf{Y}$, and for each $i \in [\tau]$ a deterministic function g_i such that $g_i(\mathbf{A}_i, \mathbf{B}) = \mathbf{Z}_i$ for all $i \in [\tau]$. Thus, for any subset $Q \subseteq [\tau]$ we have $\mathbf{Z}_Q = g'_Q(\mathbf{A}_Q, \mathbf{B})$ for some deterministic function g'_Q. And finally, for each $j > \tau$ there must be some deterministic function ψ_j such that

$$f'_j(\mathbf{Z}_{Q_j}, \overline{\mathbf{Z}}) = \psi_j(\mathbf{A}_{Q_j}, \mathbf{B}, \overline{\mathbf{Z}}).$$

Thus we can rewrite $\mathbf{X}$ as:

$$\mathbf{X} = (\mathbf{A}_1, \ldots, \mathbf{A}_\tau, \psi_{\tau+1}(\mathbf{A}_{Q_{\tau+1}}, \mathbf{B}, \overline{\mathbf{Z}}), \ldots, \psi_n(\mathbf{A}_{Q_n}, \mathbf{B}, \overline{\mathbf{Z}})).$$

Notice that the collection $\{\mathbf{A}_i\}_{i\in[\tau]}$ are mutually independent, and each has bias at most $1 - 2^{-d}$ since it is a non-constant deterministic function of d uniform bits. Thus no matter how $\mathbf{B}, \overline{\mathbf{Z}}$ are fixed, $\mathbf{X}$ becomes a (δ, t)-biased d-local NOBF source, for $\delta \leq 1 - 2^{-d}$. ◀

5 Entropy lower bounds

In this section, we obtain lower bounds on the entropy required to extract from d-local sources using degree $\leq r$ polynomials (Theorem 1.2). To do so, we actually prove a stronger theorem, which can be viewed as a *local* version of a result by Cohen and Tal [7]. In particular, Cohen and Tal show that any low degree polynomial admits a large subspace on which it is constant. We show that this holds even for this special subclass of *local* subspaces.

▶ **Theorem 3** (Theorem 3, restated). *There exist universal constants $C, c > 0$ such that for every $n, r, d \in \mathbb{N}$ such that $r \leq c\log(n)$ and $d \leq 2^{\sqrt{\log n}}$, the following holds. For any degree r polynomial $f \in \mathbb{F}_2[x_1, \ldots, x_n]$, there exists a d-local subspace $X \subseteq \mathbb{F}_2^n$ of dimension*

$$k \geq Cr(dn\log n)^{1/r}$$

on which f is constant.

This is tight: there exists a degree r polynomial $g \in \mathbb{F}_2[x_1, \ldots, x_n]$ which is an extractor for d-local affine sources of dimension $k \geq Cr(dn\log n)^{1/r}$, which has error $\varepsilon = 2^{-ck/r}$.

Notice that this immediately implies Theorem 1.2, since (the uniform distribution over) a d-local subspace is not only an affine source, but it is also a d-local source.

The tightness claim in Theorem 3 is not too difficult to prove. To do so, one can simply use Gaussian elimination to observe that a d-local affine source is actually a d'-local *NOBF* source (for some d' that is not guaranteed to be equal to d). Then, the tightness claim follows via a standard application of the probabilistic method, using known bounds [1] on the bias of a random degree $\leq r$ polynomial.

The main part of Theorem 3 (preceding the tightness claim) is much more challenging to prove. The key new ingredient we rely on is a so-called "low-weight Chevalley-Warning theorem," which may be of independent interest. We present and prove this theorem in the following subsection. As discussed in Section 2, it is then not too difficult to use our low-weight Chevalley-Warning theorem to obtain Theorem 3, and we refer the reader to the full version for more details.

5.1 A low-weight Chevalley-Warning theorem

The classical *Chevalley-Warning theorem* guarantees that a small set of low-degree polynomials admits a common nontrivial solution:

▶ **Theorem 5.1** (Chevalley-Warning theorem [24]). *Let $\{f_i\} \subseteq \mathbb{F}_2[x_1, \ldots, x_n]$ be a set of polynomials with degree at most D such that 0 is a common solution. Then there are at least 2^{n-D} common solutions to $\{f_i\}$. In particular, if $D < n$, then there must be a nontrivial common solution.*

In this subsection, we prove a "low-weight" version of this theorem. In particular, it is natural to ask not only if $\{f_i\}$ contains a nontrivial common solution, but if $\{f_i\}$ contains a nontrivial common solution of *low Hamming weight* w. It is straightforward to use Theorem 5.1 to show that $w \leq D+1$, and we remark that this is tight in general. However, we show that if most of the polynomials in $\{f_i\}$ are linear, then we can improve this bound to roughly $w \leq O(D/\log(n/D))$.

▶ **Theorem 5.2.** *Let $\{f_i\} \subseteq \mathbb{F}_2[x_1, \ldots, x_n]$ be a set of polynomials with degree at most $D < n$ and nonlinear degree at most Δ such that 0 is a common solution. Then for any w satisfying*

$$\binom{n}{\leq \lfloor w/2 \rfloor} > 2^{D+1} \cdot \binom{n}{\leq \lfloor \Delta/2 \rfloor}, \tag{1}$$

there exists a nontrivial common solution with Hamming weight at most w.

It is straightforward to show that $w = 24\Delta + 2D/\log(n/D)$ satisfies inequality 1, yielding Theorem 4. The main ingredient that goes into the proof of Theorem 5.2 is the following lemma, which says that for any big enough set $A \subseteq \mathbb{F}_2^n$ of common solutions to a system of low-degree polynomials, it holds that $A + A$ also contains a (nontrivial) common solution.

▶ **Lemma 5.3.** *Let $\{f_i\} \subseteq \mathbb{F}_2[x_1, \ldots, x_n]$ be a set of polynomials with nonlinear degree at most Δ such that 0 is a common solution. Then for any set $A \subseteq \mathbb{F}_2^n$ of common solutions of size*

$$|A| > 2\binom{n}{\leq \lfloor \Delta/2 \rfloor}$$

it holds that $A + A$ contains a nontrivial common solution.

In order to prove this, we will make use of the CLP lemma, which was instrumental in the recent resolution of the cap set conjecture.

▶ **Lemma 5.4** (CLP lemma [8])**.** *Let $f \in \mathbb{F}_2[x_1, \ldots, x_n]$ be a polynomial of degree at most r, and let M denote the $2^n \times 2^n$ matrix with entries $M_{x,y} = f(x+y)$ for $x, y \in \mathbb{F}_2^n$. Then*

$$\mathsf{rank}(M) \leq 2\binom{n}{\leq \lfloor r/2 \rfloor}.$$

Next, we show how to prove Lemma 5.3 using Lemma 5.4. Then, we conclude this section by showing how to obtain Theorem 5.2 from Lemma 5.3.

Proof of Lemma 5.3. First, let $\{g_i\} \subseteq \{f_i\}$ be the set of polynomials in $\{f_i\}$ that have degree > 1. Notice that if $A + A$ contains a nontrivial common solution to the system $\{g_i\}$, then it also contains a nontrivial common solution to $\{f_i\}$: this follows from the linearity of the polynomials of $\{f_i\} - \{g_i\}$ and the fact that every $a \in A$ is a common solution (by definition of A). Thus, it suffices to show the result for the set $\{g_i\}$.

Next, consider the polynomial $g \in \mathbb{F}_2[x_1, \ldots, x_n]$ defined as

$$g(x) := \prod_i (1 + g_i(x)).$$

It is straightforward to verify that g has degree at most Δ, and that $g(x) = 1$ if and only if x is a common solution to $\{g_i\}$. Now, suppose for contradiction that $A + A$ contains no nontrivial common solution to $\{g_i\}$: that is, for every distinct $x, y \in A$ it holds that $g(x+y) = 0$. Then, consider the $2^n \times 2^n$ matrix M with entries $M_{x,y} = g(x+y)$ for every

$x, y \in \mathbb{F}_2^n$. Define $k := |A|$, and let $M[A, A]$ denote the $k \times k$ submatrix of M obtained by taking the rows and columns of M indexed by A. Since 0 is a common solution to $\{g_i\}$, we get that $M[A, A] = I_k$ and thus

$$\mathsf{rank}(M) \geq \mathsf{rank}(M[A, A]) = \mathsf{rank}(I_k) = k > 2\binom{n}{\leq \lfloor \Delta/2 \rfloor}$$

which directly contradicts Lemma 5.4. ◀

Finally, we conclude this subsection with our proof of Theorem 5.2.

Proof of Theorem 5.2. Let $A \subseteq \mathbb{F}_2^n$ be the collection of common solutions to $\{f_i\}$. Fix any $w \in [1, n]$ such that all nonzero common solutions to $\{f_i\}$ have weight $> w$. Then for any Hamming ball $\mathcal{B} \subseteq \mathbb{F}_2^n$ of radius $\lfloor w/2 \rfloor$, it must hold that

$$|\mathcal{B} \cap A| \leq 2\binom{n}{\leq \lfloor \Delta/2 \rfloor}, \tag{2}$$

because otherwise Lemma 5.3 tells us there exist distinct $x, y \in \mathcal{B} \cap A$ such that $x + y \in A$, and $x + y$ is a nonzero vector with weight at most $2\lfloor w/2 \rfloor \leq w$ (by the triangle inequality).

Now, given any vector $v \in \mathbb{F}_2^n$, let $\mathcal{B}(v)$ denote the Hamming ball around v of radius $\lfloor w/2 \rfloor$. We consider the quantity $\sum_{v \in A} |\mathcal{B}(v)|$, and seek to sandwich it between two inequalities. By definition of Hamming ball we know that each $|\mathcal{B}(v)| = \binom{n}{\leq \lfloor w/2 \rfloor}$, and by Theorem 5.1 we know that $|A| \geq 2^{n-D}$. Combining these observations with inequality 2, we get that

$$2^{n-D}\binom{n}{\leq \lfloor w/2 \rfloor} \leq \sum_{v \in A} |\mathcal{B}(v)| \leq 2^n \cdot 2\binom{n}{\leq \lfloor \Delta/2 \rfloor},$$

since each $u \in \mathbb{F}_2^n$ is contained by at most $2\binom{n}{\leq \lfloor \Delta/2 \rfloor}$ balls (with radius $\lfloor w/2 \rfloor$) centered at a common solution $a \in A$, by inequality 2 (i.e., consider the set $\mathcal{B}(u) \cap A$). Thus

$$\binom{n}{\leq \lfloor w/2 \rfloor} \leq 2^{D+1} \cdot \binom{n}{\leq \lfloor \Delta/2 \rfloor}.$$

In summary, we have shown that any $w \in [1, n]$ for which all nonzero common solutions to $\{f_i\}$ have weight $> w$ must satisfy the above inequality. The result follows. ◀

References

1. Ido Ben-Eliezer, Rani Hod, and Shachar Lovett. Random low-degree polynomials are hard to approximate. *Comput. Complex.*, 21(1):63–81, 2012. doi:10.1007/s00037-011-0020-6.
2. Andrej Bogdanov and Siyao Guo. Sparse extractor families for all the entropy. In *Proceedings of the 4th Conference on Innovations in Theoretical Computer Science*, ITCS '13, pages 553–560. Association for Computing Machinery, 2013. doi:10.1145/2422436.2422496.
3. Eshan Chattopadhyay and Jesse Goodman. Improved extractors for small-space sources. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 610–621, 2021. doi:10.1109/FOCS52979.2021.00066.
4. Eshan Chattopadhyay, Jesse Goodman, Vipul Goyal, and Xin Li. Extractors for adversarial sources via extremal hypergraphs. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2020, pages 1184–1197, 2020.
5. Kuan Cheng and Xin Li. Randomness extraction in AC^0 and with small locality. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018)*, pages 37:1–37:20, 2018. doi:10.4230/LIPIcs.APPROX-RANDOM.2018.37.

6 Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing*, 17(2):230–261, 1988. doi:10.1137/0217015.

7 Gil Cohen and Avishay Tal. Two structural results for low degree polynomials and applications. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2015)*, pages 680–709, 2015. doi:10.4230/LIPIcs.APPROX-RANDOM.2015.680.

8 Ernie Croot, Vsevolod F Lev, and Péter Pál Pach. Progression-free sets in are exponentially small. *Annals of Mathematics*, pages 331–337, 2017.

9 Anindya De and Thomas Watson. Extractors and lower bounds for locally samplable sources. *ACM Trans. Comput. Theory*, 4(1), March 2012. doi:10.1145/2141938.2141941.

10 Yevgeniy Dodis and Kevin Yeo. Doubly-affine extractors, and their applications. In *2nd Conference on Information-Theoretic Cryptography (ITC 2021)*, pages 13:1–13:23, 2021. doi:10.4230/LIPIcs.ITC.2021.13.

11 Oded Goldreich, Emanuele Viola, and Avi Wigderson. On randomness extraction in AC^0. In *30th Conference on Computational Complexity (CCC 2015)*, pages 601–668, 2015. doi:10.4230/LIPIcs.CCC.2015.601.

12 Xuangui Huang, Peter Ivanov, and Emanuele Viola. Affine extractors and AC^0-parity. *Electron. Colloquium Comput. Complex.*, page 137, 2021. URL: https://eccc.weizmann.ac.il/report/2021/137.

13 Jesse Kamp, Anup Rao, Salil Vadhan, and David Zuckerman. Deterministic extractors for small-space sources. *J. Comput. Syst. Sci.*, 77(1):191–220, 2011. doi:10.1016/j.jcss.2010.06.014.

14 Tali Kaufman, Shachar Lovett, and Ely Porat. Weight distribution and list-decoding size of Reed–Muller codes. *IEEE Transactions on Information Theory*, 58(5):2689–2696, 2012. doi:10.1109/TIT.2012.2184841.

15 Jiatu Li and Tianqi Yang. $3.1n - o(n)$ circuit lower bounds for explicit functions. In *Electron. Colloquium Comput. Complex.*, volume 28, page 23, 2021.

16 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, 2016. doi:10.1109/FOCS.2016.26.

17 Chi-Jen Lu. Encryption against storage-bounded adversaries from on-line strong extractors. *J. Cryptol.*, 17(1):27–42, 2004. doi:10.1007/s00145-003-0217-1.

18 Luca Trevisan and Salil Vadhan. Extracting randomness from samplable distributions. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, pages 32–42, 2000. doi:10.1109/SFCS.2000.892063.

19 Salil P. Vadhan. Constructing locally computable extractors and cryptosystems in the bounded-storage model. *J. Cryptol.*, 17(1):43–77, 2004. doi:10.1007/s00145-003-0237-x.

20 Emanuele Viola. The complexity of constructing pseudorandom generators from hard functions. *Comput. Complex.*, 13(3-4):147–188, 2005. doi:10.1007/s00037-004-0187-1.

21 Emanuele Viola. Extractors for circuit sources. *SIAM Journal on Computing*, 43(2):655–672, 2014. doi:10.1137/11085983X.

22 Emanuele Viola. Quadratic maps are hard to sample. *ACM Trans. Comput. Theory*, 8(4), June 2016. doi:10.1145/2934308.

23 Emanuele Viola. Sampling lower bounds: Boolean average-case and permutations. *SIAM Journal on Computing*, 49(1):119–137, 2020. doi:10.1137/18M1198405.

24 Ewald Warning. Bemerkung zur vorstehenden arbeit von Herrn Chevalley. *Abh. Math. Sem. Univ. Hamburg*, 11:76–83, 1936.