EXTENSIONS TO THE METHOD OF MULTIPLICITIES, WITH APPLICATIONS TO KAKEYA SETS AND MERGERS[*]

ZEEV DVIR[†], SWASTIK KOPPARTY[‡], SHUBHANGI SARAF[‡], AND MADHU SUDAN[‡]

Abstract. We extend the "method of multiplicities" to get the following results, of interest in combinatorics and randomness extraction. (i) We show that every Kakeya set in $\mathbb{F}_q^n$, the n-dimensional vector space over the finite field on q elements, must be of size at least $q^n/2^n$. This bound is tight to within a $2 + o(1)$ factor for every n as $q \to \infty$. (ii) We give improved "randomness mergers": Mergers are seeded functions that take as input ℓ (possibly correlated) random variables in $\{0,1\}^N$ and a short random seed and output a single random variable in $\{0,1\}^N$ that is statistically close to having entropy $(1 - \delta) \cdot N$ when one of the ℓ input variables is distributed uniformly. The seed we require is only $(1/\delta) \cdot \log \ell$-bits long, which significantly improves upon previous construction of mergers. (iii) We give improved randomness extractors, based on our improved mergers. Specifically, we show how to construct randomness extractors that use logarithmic length seeds while extracting $1 - o(1)$ fraction of the min-entropy of the source. Previous results could extract only a constant fraction of the entropy while maintaining logarithmic seed length. The "method of multiplicitie" was used in prior work to analyze combinatorial parameters of "algebraically nice" subsets of vector spaces over finite fields. The method works by constructing somewhat low-degree interpolating polynomials that vanish on every point in the subset *with high multiplicity*. The typical use of this method involves using the "algebraic niceness" to show that the interpolating polynomial also vanishes on some points outside the subset. It then uses simple bounds on the number of zeroes of low-degree polynomials to bound the combinatorial parameter of interest. Our augmentation to this technique is that we prove, under appropriate conditions, that the interpolating polynomial vanishes *with high multiplicity* outside the set. This novelty leads to significantly tighter analyses. To develop the extended method of multiplicities, we provide a number of basic technical results about multiplicity of zeroes of polynomials that may be of general use. For instance, we strengthen the Schwartz–Zippel lemma to show that the expected multiplicity of zeroes of a nonzero degree d polynomial at a random point in S^n, for any finite subset S of the underlying field, is at most $d/|S|$ (a fact that does not seem to have been noticed in the CS literature before).

Key words. finite fields, list-decoding, polynomial method, randomness extraction

AMS subject classifications. 68Q01, 51E99

DOI. 10.1137/100783704

1. Introduction. The goal of this paper is to improve on an algebraic method that has lately been applied, quite effectively, to analyze combinatorial parameters of subsets of vector spaces that satisfy some given algebraic/geometric conditions. This technique, which we refer to as as the *polynomial method* (of combinatorics), proceeds in three steps: Given a subset K satisfying the algebraic conditions, one first constructs a nonzero low-degree polynomial that vanishes on K. Next, one uses the algebraic conditions on K to show that the polynomial vanishes at other points outside K as well. Finally, one uses the fact that the polynomial is zero too often to derive bounds on the combinatorial parameters of interest. The polynomial method has seen utility in the computer science literature in works on "list-decoding" starting with Sudan [17] and subsequent works. Recently the method has been applied to

analyze "extractors" by Guruswami, Umans, and Vadhan [7]. Most relevant to this current paper are its applications to lower bound the cardinality of "Kakeya sets" by Dvir [2], and the subsequent constructions of "mergers" and "extractors" by Dvir and Wigderson [4]. (We will elaborate on some of these results shortly.)

The *method of multiplicities*, as we term it, may be considered an extension of this method. In this extension one constructs polynomials that vanish with *high multiplicity* on the subset K. This requirement often forces one to use polynomials of higher degree than in the polynomial method, but it gains in the second step by using the high multiplicity of zeroes to conclude "more easily" that the polynomial is zero at other points. This typically leads to a tighter analysis of the combinatorial parameters of interest. This method has been applied widely in list-decoding starting with the work of Guruswami and Sudan [6] and continuing through many subsequent works, most significantly in the works of Parvaresh and Vardy [10] and Guruswami and Rudra [5], leading to rate-optimal list-decodable codes. Very recently this method was also applied to improve the lower bounds on the size of "Kakeya sets" by Saraf and Sudan [14].

The main contribution of this paper is an extension to this method that we call the *extended method of multiplicities*, which develops this method (hopefully) fully to derive even tighter bounds on the combinatorial parameters. In our extension, we start, as in the method of multiplicities, to construct a polynomial that vanishes with high multiplicity on every point of K. But then we extend the second step where we exploit the algebraic conditions to show that the polynomial vanishes with *high multiplicity* on some points outside K as well. Finally we extend the third step to show that this gives better bounds on the combinatorial parameters of interest.

By these extensions we derive nearly optimal lower bounds on the size of Kakeya sets and qualitatively improved analysis of mergers leading to new extractor constructions. We also rederive algebraically a known bound on the list size in the list-decoding of Reed–Solomon codes. We describe these contributions in detail next, before going on to describe some of the technical observations used to derive the extended method of multiplicities (which we believe are of independent interest).

1.1. Kakeya sets over finite fields. Let $\mathbb{F}_q$ denote the finite field of cardinality q. A set $K \subseteq \mathbb{F}_q^n$ is said to be a *Kakeya set* if it "contains a line in every direction." In other words, for every "direction" $\mathbf{b} \in \mathbb{F}_q^n$, there should exist an "offset" $\mathbf{a} \in \mathbb{F}_q^n$ such that the "line" through $\mathbf{a}$ in direction $\mathbf{b}$, i.e., the set $\{\mathbf{a}+t\mathbf{b}|t \in \mathbb{F}_q\}$, is contained in K. A question of interest in combinatorics/algebra/geometry, posed originally by Wolff [21], is "What is the size of the smallest Kakeya set, for a given choice of q and n?"

The trivial upper bound on the size of a Kakeya set is q^n, and this can be improved to roughly $\frac{1}{2^{n-1}}q^n$. (Precisely, the bound is $\frac{1}{2^{n-1}}q^n + O(q^{n-1})$; see [14] for a proof of this bound due to Dvir.) An almost trivial lower bound is $q^{n/2}$. (Every Kakeya set "contains" at least q^n lines, but there are at most $|K|^2$ lines that intersect K at least twice.) Until recently even the exponent of q was not known precisely (see [2] for details of work prior to 2008). This changed with the result of [2] (combined with an observation of Alon and Tao), which showed that for every n, $|K| \geq c_n q^n$, for some constant c_n depending only on n.

Subsequently the work [14] explored the growth of the constant c_n as a function of n. The result of [2] shows that $c_n \geq 1/n!$, and [14] improves this bound to show that $c_n \geq 1/(2.6)^n$. This still leaves a gap between the upper bound, and the lower bound, and we effectively close this gap.

THEOREM 1.1. *If K is a Kakeya set in $\mathbb{F}_q^n$, then $|K| \geq \frac{1}{2^n} q^n$.*

Note that our bound is tight to within a $2 + o(1)$ multiplicative factor as long as $q = \omega(2^n)$ and in particular when $n = O(1)$ and $q \to \infty$.

1.2. Randomness mergers and extractors. A general quest in the computational study of randomness is the search for simple primitives that manipulate random variables to convert their randomness into more useful forms. The exact notion of utility varies with applications. The most common notion is that of "extractors" that produce an output variable that is distributed statistically close to uniformly on the range. Other notions of interest include "condensers," "dispersers," etc. One such object of study (partly because it is useful to construct extractors) is a "randomness merger." A randomness merger takes as input Λ random variables $A_1, \ldots, A_\Lambda$, which are possibly correlated, along with a short uniformly random seed B, which is independent of $A_1, \ldots, A_\Lambda$, and "merges" the randomness of $A_1, \ldots, A_\Lambda$. Specifically the output of the merger should be statistically close to a high-entropy-rate source of randomness provided that at least one of the input variables $A_1, \ldots, A_\Lambda$ is uniform.

Mergers were first introduced by Ta-Shma [18] in the context of explicit constructions of extractors. A general framework was given in [18] that reduces the problem of constructing good extractors into that of constructing good mergers. Subsequently, in [9], mergers were used in a more complicated manner to create extractors which were optimal to within constant factors. The mergers of [9] had a very simple algebraic structure: the output of the merger was a random linear combination of the blocks over a finite vector space. The [9] merger analysis was improved in [3] by using the connection to the finite field Kakeya problem and the (then) state of the art results on Kakeya sets.

The new technique in [2] inspired Dvir and Wigderson [4] to give a very simple, algebraic, construction of a merger which can be viewed as a derandomized version of the [9] merger. They associate the domain of each random variable A_i with a vector space $\mathbb{F}_q^n$. With the Λ-tuple of random variables $A_1, \ldots, A_\Lambda$, they associate a curve $C : \mathbb{F}_q \to \mathbb{F}_q^n$ of degree $\leq \Lambda$ which "passes" through all the points $A_1, \ldots, A_\Lambda$ (that is, the image of C contains these points). They then select a random point $u \in \mathbb{F}_q$ and output $C(u)$ as the "merged" output. They show that if $q \geq \text{poly}(\Lambda \cdot n)$, then the output of the merger is statistically close to a distribution of entropy rate arbitrarily close to 1 on $\mathbb{F}_q^n$.

While the polynomial (or at least linear) dependence of q on Λ is essential to the construction above, the requirement $q \geq \text{poly}(n)$ appears only in the analysis. In our work we remove this restriction to show the following theorem.

INFORMAL THEOREM (merger). *For every Λ, q the output of the Dvir–Wigderson merger is close to a source of entropy rate $1 - \log_q \Lambda$. In particular, there exists an explicit merger for Λ sources (of arbitrary length) that outputs a source with entropy rate $1 - \delta$ and has seed length $(1/\delta) \cdot \log(\Lambda/\epsilon)$ for any error ϵ.*

The above theorem (in its more formal form given in Theorem 5.3) allows us to merge Λ sources using seed length which is only logarithmic in the number of sources and has no dependence on the length of each source. Earlier constructions of mergers required the seed to depend either linearly on the number of blocks [9, 22] or to depend also on the *length* of each block [4].[1]

One consequence of our improved merger construction is an improved construc-

---

[1] The result we refer to in [22, Theorem 5.1] is actually a condenser (which is stronger than a merger).

tion of extractors. Recall that a (k, ϵ)-extractor $E : \{0, 1\}^n \times \{0, 1\}^d \to \{0, 1\}^m$ is a deterministic function that takes any random variable X with min-entropy at least k over $\{0, 1\}^n$ and an independent uniformly distributed seed $\mathsf{Y} \in \{0, 1\}^d$ and converts it to the random variable $E(\mathsf{X}, \mathsf{Y})$ that is ϵ-close in statistical distance to a uniformly distributed random variable over $\{0, 1\}^m$. Such an extractor is efficient if E is polynomial time computable.

A diverse collection of efficient extractors are known in the literature (see the survey [15] and the more recent [7, 4] for references), and many applications have been found for explicit extractors in various research areas spanning theoretical computer science. Yet all previous constructions lost a linear fraction of the min-entropy of the source (i.e., achieved $m = (1 - \epsilon)k$ for some constant $\epsilon > 0$) or used superlogarithmic seed length ($d = \omega(\log n)$). We show that our merger construction yields, by combining with several of the prior tools in the arsenal of extractor constructions, an extractor which extracts a $1 - \frac{1}{\text{polylog(n)}}$ fraction of the min-entropy of the source while still using $O(\log n)$-length seeds. We now state our extractor result in an informal way (see Theorem 6.3 for the formal statement).

INFORMAL THEOREM (extractor). *There exists an explicit (k, ϵ)-extractor for all min-entropies k with $O(\log n)$ seed, entropy loss $O(k/\text{polylog}(n))$, and error $\epsilon = 1/\text{polylog}(n)$, where the powers in the $\text{polylog}(n)$ can be arbitrarily high constants.*

As seen in the above theorem, our extractors can handle error which is only polylogarithmic. For smaller values of ϵ there are better constructions of extractors (e.g., those of [7]).

1.3. List-decoding of Reed–Solomon codes. The problem of list-decoding Reed–Solomon codes is the following: Given a sequence of points

$$(\alpha_1, \beta_1), \ldots, (\alpha_n, \beta_n) \in \mathbb{F}_q \times \mathbb{F}_q,$$

and parameters k and t, find the list of all polynomials $p_1, \ldots, p_L$ of degree at most k that agree with the given set of points on t locations; i.e., for every $j \in \{1, \ldots, L\}$ the set $\{i | p_j(\alpha_i) = \beta_i\}$ has at least t elements. The associated combinatorial problem is "How large can the list size, L, be for a given choice of k, t, n, q (when maximized over all possible sets of distinct input points)?"

A somewhat nonstandard, yet reasonable, interpretation of the list-decoding algorithms of [17, 6] is that they give algebraic proofs, by the polynomial method and the method of multiplicities, of known combinatorial upper bounds on the list size, when $t > \sqrt{kn}$. Their proofs also *happen* to be algorithmic and so lead to algorithms to find a list of all such polynomials.

However, the bound given on the list size in the above works does not match the best known combinatorial bound. The best known bound to date seems to be that of Cassuto and Bruck [1], who show that, letting $R = k/n$ and $\gamma = t/n$, if $\gamma^2 > R$, then the list size L is bounded by $O(\frac{\gamma}{\gamma^2 - R})$ (in contrast, the Johnson bound and the analysis of [6] gives a list size bound of $O(\frac{1}{\gamma^2 - R})$, which is asymptotically worse for, say, $\gamma = (1 + O(1))\sqrt{R}$ and R tending to 0). In Theorem 7.2 we recover the bound of [1] using our extended method of multiplicities.

1.4. Technique: Extended method of multiplicities. The common insight to all the above improvements is that the extended method of multiplicities can be applied to each problem to improve the parameters. Here we attempt to describe the technical novelties in the development of the extended method of multiplicities.

For concreteness, let us take the case of the Kakeya set problem. Given a set $K \subseteq \mathbb{F}_q^n$, the method first finds a nonzero polynomial $P \in \mathbb{F}_q[X_1, \ldots, X_n]$ that vanishes with high multiplicity m on each point of K. The next step is to prove that P vanishes with fairly high multiplicity ℓ at every point in $\mathbb{F}_q^n$ as well. This step turns out to be somewhat subtle (and is evidenced by the fact that the exact relationship between m and ℓ is not simple). Our analysis here crucially uses the fact that the (Hasse) derivatives of the polynomial P, which are central to the notion of multiplicity of roots, are themselves polynomials, and also vanish with high multiplicity at points in K. This fact does not seem to have been needed/used in prior works and is central to ours.

A second important technical novelty arises in the final step of the method of multiplicities, where we need to conclude that if the degree of P is "small," then P must be identically zero. Unfortunately, in our application, the degree of P may be much larger than q (or nq, or even q^n). To prove that it is identically zero, we need to use the fact that P vanishes *with high multiplicity* at every point in $\mathbb{F}_q^n$, and this requires some multiplicity enhanced version of the standard Schwartz–Zippel lemma. We prove such a strengthening, showing that the expected multiplicity of zeroes of a degree d polynomial (even when $d \gg q$) at a random point in $\mathbb{F}_q^n$ is at most d/q (see Lemma 2.7). Using this lemma, we are able to derive much better benefits from the "polynomial method." Indeed, we feel that this allows us to go beyond the limitations of the function space mapping $\mathbb{F}_q^n$ to $\mathbb{F}_q$ and capitalize on the full power of the polynomial ring $\mathbb{F}_q[\mathbf{X}]$.

Putting these ingredients together, the analysis of the Kakeya sets follows easily. The analysis of the mergers follows a similar path and may be viewed as a "statistical" extension of the Kakeya set analysis to "curve" based sets; i.e., here we consider sets S that have the property that, for a noticeable fraction of points $\mathbf{x} \in \mathbb{F}_q^n$, there exists a low-degree curve passing through $\mathbf{x}$ that has a noticeable fraction of its points in S. We prove that such sets must also be large, and this leads to the analysis of the Dvir–Wigderson merger.

Organization of this paper. In section 2 we define the notion of the multiplicity of the roots of a polynomial, using the notion of the Hasse derivative. We present some basic facts about multiplicities and Hasse derivatives, and we also present the multiplicity-based version of the Schwartz–Zippel lemma. In section 3 we present our lower bounds for Kakeya sets. In section 4 we extend this analysis for "curves" and for "statistical" versions of the Kakeya property. This leads to our analysis of the Dvir–Wigderson merger in section 5. We then show how to use our mergers to construct the novel extractors in section 6. Finally, in section 7, we include the algebraic proof of the list size bounds for the list-decoding of Reed–Solomon codes.

2. Preliminaries. In this section we formally define the notion of "multiplicity of zeroes" along with the companion notion of the "Hasse derivative." We also describe basic properties of these notions, concluding with the proof of the "multiplicity enhanced version" of the Schwartz–Zippel lemma.

2.1. Basic definitions. We start with some notation. We use $[n]$ to denote the set $\{1, \ldots, n\}$. For a vector $\mathbf{i} = \langle i_1, \ldots, i_n \rangle$ of nonnegative integers, its *weight*, denoted $\mathrm{wt}(\mathbf{i})$, equals $\sum_{j=1}^{n} i_j$.

Let $\mathbb{F}$ be any field, and let $\mathbb{F}_q$ denote the finite field of q elements. For $\mathbf{X} = \langle X_1, \ldots, X_n \rangle$, let $\mathbb{F}[\mathbf{X}]$ be the ring of polynomials in $X_1, \ldots, X_n$ with coefficients in $\mathbb{F}$. For a polynomial $P(\mathbf{X})$, we let $H_P(\mathbf{X})$ denote the homogeneous part of $P(\mathbf{X})$ of highest total degree. Specifically, if $P(\mathbf{X})$ is a polynomial of degree d, then $H_P(\mathbf{X})$

is the homogeneous degree d polynomial such that $P(\mathbf{X}) - H_P(\mathbf{X})$ has degree less than d.

For a vector of nonnegative integers $\mathbf{i} = \langle i_1, \ldots, i_n \rangle$, let $\mathbf{X^i}$ denote the monomial $\prod_{j=1}^{n} X_j^{i_j} \in \mathbb{F}[\mathbf{X}]$. Note that the (total) degree of this monomial equals wt($\mathbf{i}$). For n-tuples of nonnegative integers $\mathbf{i}$ and $\mathbf{j}$, we use the notation

$$\binom{\mathbf{i}}{\mathbf{j}} = \prod_{k=1}^{n} \binom{i_k}{j_k}.$$

Note that the coefficient of $\mathbf{Z^i W^{r-i}}$ in the expansion of $(\mathbf{Z} + \mathbf{W})^{\mathbf{r}}$ equals $\binom{\mathbf{r}}{\mathbf{i}}$.

DEFINITION 2.1 ((Hasse) derivative). *For $P(\mathbf{X}) \in \mathbb{F}[\mathbf{X}]$ and nonnegative vector $\mathbf{i}$, the $\mathbf{i}$th (Hasse) derivative of P, denoted $P^{(\mathbf{i})}(\mathbf{X})$, is the coefficient of $\mathbf{Z^i}$ in the polynomial $\tilde{P}(\mathbf{X}, \mathbf{Z}) \overset{\text{def}}{=} P(\mathbf{X} + \mathbf{Z}) \in \mathbb{F}[\mathbf{X}, \mathbf{Z}]$.*

Thus,

$$(2.1) \qquad\qquad P(\mathbf{X} + \mathbf{Z}) = \sum_{\mathbf{i}} P^{(\mathbf{i})}(\mathbf{X})\mathbf{Z^i}.$$

We are now ready to define the notion of the (zero-)multiplicity of a polynomial at any given point.

DEFINITION 2.2 (multiplicity). *For $P(\mathbf{X}) \in \mathbb{F}[\mathbf{X}]$ and $\mathbf{a} \in \mathbb{F}^n$, the multiplicity of P at $\mathbf{a} \in \mathbb{F}^n$, denoted mult$(P, \mathbf{a})$, is the largest integer M such that for every nonnegative vector $\mathbf{i}$ with wt($\mathbf{i}$) $< M$, we have $P^{(\mathbf{i})}(\mathbf{a}) = 0$ (if M may be taken arbitrarily large, we set mult$(P, \mathbf{a}) = \infty$).*

Note that mult$(P, \mathbf{a}) \geq 0$ for every $\mathbf{a}$. Also, $P(\mathbf{a}) = 0$ if and only if mult$(P, \mathbf{a}) \geq 1$. We also note that the condition $P^{(\mathbf{i})}(\mathbf{a}) = 0$ imposes a homogeneous linear constraint on the coefficients of $P(\mathbf{X})$ (for every $\mathbf{a} \in \mathbb{F}^n$). A consequence that we will use often later is the condition that mult$(P, \mathbf{a}) \geq M$ imposes at most $\binom{M+n-1}{n}$ homogeneous linear constraints on the coefficients of $P(\mathbf{X})$, one such constraint corresponding to each vector $\mathbf{i}$ of weight less than M.

The above notation and definitions also extend naturally to a tuple $\mathbf{P}(\mathbf{X}) = \langle P_1(\mathbf{X}), \ldots, P_m(\mathbf{X}) \rangle$ of polynomials with $P^{(\mathbf{i})} \in \mathbb{F}[\mathbf{X}]^m$ denoting the vector

$$\langle (P_1)^{(\mathbf{i})}, \ldots, (P_m)^{(\mathbf{i})} \rangle.$$

In particular, we define mult$(\mathbf{P}, \mathbf{a}) = \min_{j \in [m]} \{\text{mult}(P_j, \mathbf{a})\}$.

The definition of multiplicity above is similar to the standard (analytic) definition of multiplicity with the difference that the standard partial derivative has been replaced by the Hasse derivative. The Hasse derivative is also a reasonably well-studied quantity (see, for example, [8, pp. 144–155]) and seems to have first appeared in the CS literature (without being explicitly referred to by this name) by way of the work of Guruswami and Sudan [6]. It typically behaves like the standard derivative, but with some key differences that make it more useful/informative over finite fields. For completeness we review basic properties of the Hasse derivative and multiplicity in the following subsections.

2.2. Properties of Hasse derivatives. The following proposition lists basic properties of the Hasse derivatives. Parts (1)–(3) below are the same as for the usual derivative, while part (4) is not! Part (4) considers the derivatives of the derivatives of a polynomial and shows a different relationship than is standard for the analytic derivative. However it does show that the $\mathbf{j}$th derivative of the $\mathbf{i}$th derivative is zero

if (though not necessarily only if) the $(\mathbf{i} + \mathbf{j})$th derivative is zero, and this is critical for our purposes.

PROPOSITION 2.3 (basic properties of derivatives). *Let* $P(\mathbf{X}), Q(\mathbf{X}) \in \mathbb{F}[\mathbf{X}]$ *and let* $\mathbf{i}$, $\mathbf{j}$ *be vectors of nonnegative integers. Then,*

1. $P^{(\mathbf{i})}(\mathbf{X}) + Q^{(\mathbf{i})}(\mathbf{X}) = (P + Q)^{(\mathbf{i})}(\mathbf{X})$.
2. *If* $P(\mathbf{X})$ *is homogeneous of degree* d, *then either* $P^{(\mathbf{i})}(\mathbf{X})$ *is homogeneous of degree* $d - \mathrm{wt}(i)$ *or* $P^{(\mathbf{i})}(\mathbf{X}) = 0$.
3. *Either* $(H_P)^{(\mathbf{i})}(\mathbf{X}) = H_{P^{(\mathbf{i})}}(\mathbf{X})$ *or* $(H_P)^{(\mathbf{i})}(\mathbf{X}) = 0$.
4. $\left(P^{(\mathbf{i})}\right)^{(\mathbf{j})}(\mathbf{X}) = \binom{\mathbf{i}+\mathbf{j}}{\mathbf{i}} P^{(\mathbf{i}+\mathbf{j})}(\mathbf{X})$.

Proof. Item 1 follows immediately from (2.1).

For item 2, observe that if $P(\mathbf{X})$ is homogeneous of degree d, then so is $P(\mathbf{X}+\mathbf{Z})$. Thus by (2.1), $P^{(\mathbf{i})}(\mathbf{X})\mathbf{Z}^{\mathbf{i}}$ must be either homogeneous of degree d or 0. Hence $P^{(\mathbf{i})}(\mathbf{X})$ is either homogeneous of degree $d - \mathrm{wt}(i)$ or 0.

For item 3, let $P(\mathbf{X}) = H_P(\mathbf{X}) + Q(\mathbf{X})$, where $P(\mathbf{X})$ is of degree d, $H_P(\mathbf{X})$ is homogeneous of degree d, and $Q(\mathbf{X})$ is of degree $< d$. By item 1, $(H_P)^{(\mathbf{i})}(\mathbf{X}) = P^{(\mathbf{i})}(\mathbf{X}) - Q^{(\mathbf{i})}(\mathbf{X})$. By item 2, if $(H_P)^{(\mathbf{i})}(\mathbf{X})$ is nonzero, then it must be homogeneous of degree $d - \mathrm{wt}(i)$. Hence $P^{(\mathbf{i})}(\mathbf{X}) - Q^{(\mathbf{i})}(\mathbf{X})$ is homogeneous of degree $d - \mathrm{wt}(i)$, implying that $P^{(\mathbf{i})}(\mathbf{X}) - Q^{(\mathbf{i})}(\mathbf{X}) = H_{P^{(\mathbf{i})} - Q^{(\mathbf{i})}}(\mathbf{X})$. Since by item 2, the degree of $Q^{(\mathbf{i})}$ is strictly less than $d - \mathrm{wt}(i)$, it must be that $H_{P^{(\mathbf{i})} - Q^{(\mathbf{i})}}(\mathbf{X}) = H_{P^{(\mathbf{i})}}(\mathbf{X})$, and item 3 follows.

For item 4, we expand $P(\mathbf{X} + \mathbf{Z} + \mathbf{W})$ in two ways. First expand

$$P(\mathbf{X} + (\mathbf{Z} + \mathbf{W})) = \sum_{\mathbf{k}} P^{(\mathbf{k})}(\mathbf{X})(\mathbf{Z} + \mathbf{W})^{\mathbf{k}}$$

$$= \sum_{\mathbf{k}} \sum_{\mathbf{i}+\mathbf{j}=\mathbf{k}} P^{(\mathbf{k})}(\mathbf{X}) \binom{\mathbf{k}}{\mathbf{i}} \mathbf{Z}^{\mathbf{j}} \mathbf{W}^{\mathbf{i}}$$

$$= \sum_{\mathbf{i},\mathbf{j}} P^{(\mathbf{i}+\mathbf{j})}(\mathbf{X}) \binom{\mathbf{i}+\mathbf{j}}{\mathbf{i}} \mathbf{Z}^{\mathbf{j}} \mathbf{W}^{\mathbf{i}}.$$

On the other hand, we may write

$$P((\mathbf{X}+\mathbf{Z})+\mathbf{W}) = \sum_{\mathbf{i}} P^{(\mathbf{i})}(\mathbf{X}+\mathbf{Z})\mathbf{W}^{\mathbf{i}} = \sum_{\mathbf{i}} \sum_{\mathbf{j}} \left(P^{(\mathbf{i})}\right)^{(\mathbf{j})}(\mathbf{X})\mathbf{Z}^{\mathbf{j}}\mathbf{W}^{\mathbf{i}}.$$

Comparing coefficients of $\mathbf{Z}^{\mathbf{j}}\mathbf{W}^{\mathbf{i}}$ on both sides, we get the result. ☐

2.3. Properties of multiplicities. We now translate some of the properties of the Hasse derivative into properties of the multiplicities.

LEMMA 2.4 (basic properties of multiplicities). *If* $P(\mathbf{X}) \in \mathbb{F}[\mathbf{X}]$ *and* $\mathbf{a} \in \mathbb{F}^n$ *are such that* $\mathrm{mult}(P, \mathbf{a}) = m$, *then* $\mathrm{mult}(P^{(\mathbf{i})}, \mathbf{a}) \geq m - \mathrm{wt}(\mathbf{i})$.

Proof. By assumption, for any $\mathbf{k}$ with $\mathrm{wt}(\mathbf{k}) < m$, we have $P^{(\mathbf{k})}(\mathbf{a}) = 0$. Now take any $\mathbf{j}$ such that $\mathrm{wt}(\mathbf{j}) < m - \mathrm{wt}(\mathbf{i})$. By item 4 of Proposition 2.3, $(P^{(\mathbf{i})})^{(\mathbf{j})}(\mathbf{a}) = \binom{\mathbf{i}+\mathbf{j}}{\mathbf{i}} P^{(\mathbf{i}+\mathbf{j})}(\mathbf{a})$. Since $\mathrm{wt}(\mathbf{i} + \mathbf{j}) = \mathrm{wt}(\mathbf{i}) + \mathrm{wt}(\mathbf{j}) < m$, we deduce that $(P^{(\mathbf{i})})^{(\mathbf{j})}(\mathbf{a}) = 0$. Thus $\mathrm{mult}(P^{(\mathbf{i})}, \mathbf{a}) \geq m - \mathrm{wt}(\mathbf{i})$. ☐

We now discuss the behavior of multiplicities under composition of polynomial tuples. Let $\mathbf{X} = (X_1, \ldots, X_n)$ and $\mathbf{Y} = (Y_1, \ldots, Y_\ell)$ be formal variables. Let $\mathbf{P}(\mathbf{X}) = (P_1(\mathbf{X}), \ldots, P_m(\mathbf{X})) \in \mathbb{F}[\mathbf{X}]^m$ and $\mathbf{Q}(\mathbf{Y}) = (Q_1(Y), \ldots, Q_n(Y)) \in \mathbb{F}[\mathbf{Y}]^n$. We define the composition polynomial $\mathbf{P} \circ \mathbf{Q}(\mathbf{Y}) \in \mathbb{F}[\mathbf{Y}]^m$ to be the polynomial $\mathbf{P}(Q_1(\mathbf{Y}), \ldots, Q_n(\mathbf{Y}))$. In this situation we have the following proposition.

PROPOSITION 2.5. *Let* $\mathbf{P}(\mathbf{X}), \mathbf{Q}(\mathbf{Y})$ *be as above. Then for any* $\mathbf{a} \in \mathbb{F}^\ell$,

$$\mathrm{mult}(\mathbf{P} \circ \mathbf{Q}, \mathbf{a}) \geq \mathrm{mult}(\mathbf{P}, \mathbf{Q}(\mathbf{a})) \cdot \mathrm{mult}(\mathbf{Q} - \mathbf{Q}(\mathbf{a}), \mathbf{a}).$$

In particular, since $\mathrm{mult}(\mathbf{Q} - \mathbf{Q}(\mathbf{a}), \mathbf{a}) \geq 1$, *we have* $\mathrm{mult}(\mathbf{P} \circ \mathbf{Q}, \mathbf{a}) \geq \mathrm{mult}(\mathbf{P}, \mathbf{Q}(\mathbf{a}))$.

Proof. Let $m_1 = \mathrm{mult}(\mathbf{P}, \mathbf{Q}(\mathbf{a}))$ and $m_2 = \mathrm{mult}(\mathbf{Q} - \mathbf{Q}(\mathbf{a}), \mathbf{a})$. Clearly $m_2 > 0$. If $m_1 = 0$, then the result is obvious. Now assume that $m_1 > 0$ (so that $\mathbf{P}(\mathbf{Q}(\mathbf{a})) = 0$). Then we see that

$$\begin{aligned} \mathbf{P}(\mathbf{Q}(\mathbf{a} + \mathbf{Z})) &= \mathbf{P}\left(\mathbf{Q}(\mathbf{a}) + \sum_{\mathbf{i} \neq 0} \mathbf{Q}^{(\mathbf{i})}(\mathbf{a})\mathbf{Z}^{\mathbf{i}}\right) \\ &= \mathbf{P}\left(\mathbf{Q}(\mathbf{a}) + \sum_{\mathrm{wt}(\mathbf{i}) \geq m_2} \mathbf{Q}^{(\mathbf{i})}(\mathbf{a})\mathbf{Z}^{\mathbf{i}}\right) && \text{(since } \mathrm{mult}(\mathbf{Q} - \mathbf{Q}(\mathbf{a}), \mathbf{a}) = m_2 > 0) \\ &= \mathbf{P}\left(\mathbf{Q}(\mathbf{a}) + h(\mathbf{Z})\right) && \text{(where } h(\mathbf{Z}) = \sum_{\mathrm{wt}(\mathbf{i}) \geq m_2} \mathbf{Q}^{(\mathbf{i})}(\mathbf{a})\mathbf{Z}^{\mathbf{i}}) \\ &= \mathbf{P}(\mathbf{Q}(\mathbf{a})) + \sum_{\mathbf{j} \neq 0} \mathbf{P}^{(\mathbf{j})}(\mathbf{Q}(\mathbf{a}))h(\mathbf{Z})^{\mathbf{j}} \\ &= \sum_{\mathrm{wt}(\mathbf{j}) \geq m_1} \mathbf{P}^{(\mathbf{j})}(\mathbf{Q}(\mathbf{a}))h(\mathbf{Z})^{\mathbf{j}} && \text{(since } \mathrm{mult}(\mathbf{P}, \mathbf{Q}(\mathbf{a})) = m_1 > 0). \end{aligned}$$

Thus, since each monomial $\mathbf{Z}^{\mathbf{i}}$ appearing in h has $\mathrm{wt}(\mathbf{i}) \geq m_2$, and since each occurrence of $h(\mathbf{Z})$ in $\mathbf{P}(\mathbf{Q}(\mathbf{a} + \mathbf{Z}))$ is raised to the power $\mathbf{j}$, with $\mathrm{wt}(\mathbf{j}) \geq m_1$, we conclude that $\mathbf{P}(\mathbf{Q}(\mathbf{a}+\mathbf{Z}))$ is of the form $\sum_{\mathrm{wt}(\mathbf{k}) \geq m_1 \cdot m_2} c_{\mathbf{k}}\mathbf{Z}^{\mathbf{k}}$. This shows that $(\mathbf{P} \circ \mathbf{Q})^{(\mathbf{k})}(\mathbf{a}) = 0$ for each $\mathbf{k}$ with $\mathrm{wt}(\mathbf{k}) < m_1 \cdot m_2$, and the result follows. $\square$

COROLLARY 2.6. *Let* $P(\mathbf{X}) \in \mathbb{F}[\mathbf{X}]$, *where* $\mathbf{X} = (X_1, \ldots, X_n)$. *Let* $\mathbf{a}, \mathbf{b} \in \mathbb{F}^n$. *Let* $P_{\mathbf{a},\mathbf{b}}(T)$ *be the polynomial* $P(\mathbf{a} + T \cdot \mathbf{b}) \in \mathbb{F}[T]$. *Then for any* $t \in \mathbb{F}$,

$$\mathrm{mult}(P_{\mathbf{a},\mathbf{b}}, t) \geq \mathrm{mult}(P, \mathbf{a} + t \cdot \mathbf{b}).$$

Proof. Let $\mathbf{Q}(T) = \mathbf{a} + T\mathbf{b} \in \mathbb{F}[T]^n$. Applying the previous proposition to $P(\mathbf{X})$ and $\mathbf{Q}(T)$, we get the desired claim. $\square$

2.4. Strengthening of the Schwartz–Zippel lemma. We are now ready to state and prove the strengthening of the Schwartz–Zippel lemma. In the standard form, this lemma states that the probability that $P(\mathbf{a}) = 0$ when $\mathbf{a}$ is drawn uniformly at random from S^n is at most $d/|S|$, where P is a nonzero degree d polynomial and $S \subseteq \mathbb{F}$ is a finite set. Using $\min\{1, \mathrm{mult}(P, \mathbf{a})\}$ as the indicator variable that is 1 if $P(\mathbf{a}) = 0$, this lemma can be restated as saying $\sum_{\mathbf{a} \in S^n} \min\{1, \mathrm{mult}(P, \mathbf{a})\} \leq d \cdot |S|^{n-1}$. Our version below strengthens this lemma by replacing $\min\{1, \mathrm{mult}(P, \mathbf{a})\}$ with $\mathrm{mult}(P, \mathbf{a})$ in this inequality.

LEMMA 2.7. *Let* $P \in \mathbb{F}[\mathbf{X}]$ *be an* n-*variate nonzero polynomial of total degree at most* d. *Then for any finite* $S \subseteq \mathbb{F}$,

$$\sum_{\mathbf{a} \in S^n} \mathrm{mult}(P, \mathbf{a}) \leq d \cdot |S|^{n-1}.$$

Proof. We prove it by induction on n.

For the base case when $n = 1$, we first show that if $\mathrm{mult}(P, a) = m$, then $(X - a)^m$ divides $P(X)$. To see this, note that by the definition of multiplicity, we have that

$P(a + Z) = \sum_i P^{(i)}(a)Z^i$ and $P^{(i)}(a) = 0$ for all $i < m$. We conclude that Z^m divides $P(a + Z)$, and thus $(X - a)^m$ divides $P(X)$. It follows that $\sum_{a \in S} \text{mult}(P, a)$ is at most the degree of P.

Now suppose that $n > 1$. Let

$$P(X_1, \ldots, X_n) = \sum_{j=0}^{t} P_j(X_1, \ldots, X_{n-1})X_n^j,$$

where $0 \leq t \leq d$, $P_t(X_1, \ldots, X_{n-1}) \neq 0$, and $\deg(P_j) \leq d - j$.

For any $a_1, \ldots, a_{n-1} \in S$, let $m_{a_1, \ldots, a_{n-1}} = \text{mult}(P_t, (a_1, \ldots, a_{n-1}))$. We will show that

(2.2) $$\sum_{a_n \in S} \text{mult}(P, (a_1, \ldots, a_n)) \leq m_{a_1, \ldots, a_{n-1}} \cdot |S| + t.$$

Given this, we may then bound

$$\sum_{a_1, \ldots, a_n \in S} \text{mult}(P, (a_1, \ldots, a_n)) \leq \sum_{a_1, \ldots, a_{n-1} \in S} m_{a_1, \ldots, a_{n-1}} \cdot |S| + |S|^{n-1} \cdot t.$$

By the induction hypothesis applied to P_t, we know that

$$\sum_{a_1, \ldots, a_{n-1} \in S} m_{a_1, \ldots, a_{n-1}} \leq \deg(P_t) \cdot |S|^{n-2} \leq (d - t) \cdot |S|^{n-2}.$$

This implies the result.

We now prove (2.2). Fix $a_1, \ldots, a_{n-1} \in S$ and let $\mathbf{i} = (i_1, \ldots, i_{n-1})$ be such that $\text{wt}(\mathbf{i}) = m_{a_1, \ldots, a_{n-1}}$ and $P_t^{(\mathbf{i})}(X_1, \ldots, X_{n-1}) \neq 0$. Letting $(\mathbf{i}, 0)$ denote the vector $(i_1, \ldots, i_{n-1}, 0)$, we note that

$$P^{(\mathbf{i}, 0)}(X_1, \ldots, X_n) = \sum_{j=0}^{t} P_j^{(\mathbf{i})}(X_1, \ldots, X_{n-1})X_n^j,$$

and hence $P^{(\mathbf{i}, 0)}$ is a nonzero polynomial.

Now by Lemma 2.4 and Corollary 2.6 (applied on the line through $(a_1, \ldots, a_{n-1}, 0)$ in direction $(0, \ldots, 0, a_n)$), we know that

$$\text{mult}(P(X_1, \ldots, X_n), (a_1, \ldots, a_n)) \leq \text{wt}(\mathbf{i}, 0) + \text{mult}(P^{(\mathbf{i}, 0)}(X_1, \ldots, X_n), (a_1, \ldots, a_n))$$

$$\leq m_{a_1, \ldots, a_{n-1}} + \text{mult}(P^{(\mathbf{i}, 0)}(a_1, \ldots, a_{n-1}, X_n), a_n).$$

Summing this up over all $a_n \in S$, and applying the $n = 1$ case of this lemma to the nonzero univariate degree-t polynomial $P^{(\mathbf{i}, 0)}(a_1, \ldots, a_{n-1}, X_n)$, we get (2.2). This completes the proof of the lemma. ☐

The following corollary simply states the above lemma in contrapositive form, with $S = \mathbb{F}_q$.

COROLLARY 2.8. *Let $P \in \mathbb{F}_q[\mathbf{X}]$ be a polynomial of total degree at most d. If $\sum_{\mathbf{a} \in \mathbb{F}_q^n} \text{mult}(P, \mathbf{a}) > d \cdot q^{n-1}$, then $P(\mathbf{X}) = 0$.*

3. A lower bound on the size of Kakeya sets. We now give a lower bound on the size of Kakeya sets in $\mathbb{F}_q^n$. We implement the plan described in section 1. Specifically, in Proposition 3.1, we show that we can find a somewhat low-degree nonzero polynomial that vanishes with high multiplicity on any given Kakeya set, where the degree of the polynomial grows with the size of the set. Next, in Claim 3.3, we show that the homogeneous part of this polynomial vanishes with fairly high multiplicity everywhere in $\mathbb{F}_q^n$. Using the strengthened Schwartz–Zippel lemma, we conclude that the homogeneous polynomial is identically zero if the Kakeya set is too small, leading to the desired contradiction. The resulting lower bound (slightly stronger than Theorem 1.1) is given in Theorem 3.2.

PROPOSITION 3.1. *Given a set $K \subseteq \mathbb{F}^n$ and nonnegative integers m, d such that*

$$\binom{m+n-1}{n} \cdot |K| < \binom{d+n}{n},$$

there exists a nonzero polynomial $P = P_{m,K} \in \mathbb{F}[\mathbf{X}]$ of total degree at most d such that $\mathrm{mult}(P, \mathbf{a}) \geq m$ for every $\mathbf{a} \in K$.

Proof. The number of possible monomials in P is $\binom{d+n}{n}$. Hence there are $\binom{d+n}{n}$ degrees of freedom in the choice for the coefficients for these monomials. For a given point $\mathbf{a}$, the condition that $\mathrm{mult}(P, \mathbf{a}) \geq m$ imposes $\binom{m+n-1}{n}$ homogeneous linear constraints on the coefficients of P. Since the total number of (homogeneous) linear constraints is $\binom{m+n-1}{n} \cdot |K|$, which is strictly less than the number of unknowns, there is a nontrivial solution. ∎

THEOREM 3.2. *If $K \subseteq \mathbb{F}_q^n$ is a Kakeya set, then $|K| \geq \left(\frac{q}{2-1/q}\right)^n$.*

Proof. Let ℓ be a large multiple of q and let

$$m = 2\ell - \ell/q,$$

$$d = \ell q - 1.$$

These three parameters (ℓ, m, and d) will be used as follows: d will be the bound on the degree of a polynomial P which vanishes on K, m will be the multiplicity of the zeroes of P on K, and ℓ will be the multiplicity of the zeroes of the homogeneous part of P, which we will deduce by restricting P to lines passing through K.

Note that by the choices above we have $d < \ell q$ and $(m-\ell)q > d-\ell$. A particular form we will use later is that $(m - w)q > d - w$ for every $w \leq \ell$ (using $q \geq 1$). We prove below that

$$|K| \geq \frac{\binom{d+n}{n}}{\binom{m+n-1}{n}} \geq \alpha^n,$$

where $\alpha \to \frac{q}{2-1/q}$ as $\ell \to \infty$.

Assume for contradiction that $|K| < \frac{\binom{d+n}{n}}{\binom{m+n-1}{n}}$. Then, by Proposition 3.1, there exist $d^* \leq d$ and a nonzero polynomial $P(\mathbf{X}) \in \mathbb{F}[\mathbf{X}]$ of total degree exactly d^* such that $\mathrm{mult}(P, \mathbf{x}) \geq m$ for every $\mathbf{x} \in K$. Note that $d^* \geq \ell$ since $d^* \geq m$ (since P is nonzero and vanishes to multiplicity $\geq m$ at some point), and $m \geq \ell$ by our choice of m. Let $H_P(\mathbf{X})$ be the homogeneous part of $P(\mathbf{X})$ of degree d^*. Note that $H_P(\mathbf{X})$ is nonzero. The following claim shows that H_P vanishes to multiplicity ℓ at each point of $\mathbb{F}_q^n$.

CLAIM 3.3. *For each* $\mathbf{b} \in \mathbb{F}_q^n$,

$$\operatorname{mult}(H_P, \mathbf{b}) \geq \ell.$$

Proof. Fix $\mathbf{i}$ with $\operatorname{wt}(\mathbf{i}) = w \leq \ell - 1$. Let $Q(\mathbf{X}) = P^{(\mathbf{i})}(\mathbf{X})$. Let d' be the degree of the polynomial $Q(\mathbf{X})$, and note that $d' \leq d^* - w$.

Let $\mathbf{a} = \mathbf{a}(\mathbf{b})$ be such that $\{\mathbf{a} + t\mathbf{b} | t \in \mathbb{F}_q\} \subset K$. Then for all $t \in \mathbb{F}_q$, by Lemma 2.4, $\operatorname{mult}(Q, \mathbf{a} + t\mathbf{b}) \geq m - w$. Since $w \leq \ell - 1$ and $(m - \ell) \cdot q > d^* - \ell$, we get that $(m - w) \cdot q > d^* - w \geq d'$.

Let $Q_{\mathbf{a},\mathbf{b}}(T)$ be the polynomial $Q(\mathbf{a} + T\mathbf{b}) \in \mathbb{F}_q[T]$. Then $Q_{\mathbf{a},\mathbf{b}}(T)$ is a univariate polynomial of degree at most d', and by Corollary 2.6, it vanishes at each point of $\mathbb{F}_q$ with multiplicity $m - w$. Since

$$(m - w) \cdot q > d^* - w \geq \deg(Q_{\mathbf{a},\mathbf{b}}(T)),$$

we conclude that $Q_{\mathbf{a},\mathbf{b}}(T) = 0$. Hence the coefficient of $T^{d'}$ in $Q_{\mathbf{a},\mathbf{b}}(T)$ is 0. Let H_Q be the homogeneous component of Q of highest degree. Observe that the coefficient of $T^{d'}$ in $Q_{\mathbf{a},\mathbf{b}}(T)$ is $H_Q(\mathbf{b})$. Hence $H_Q(\mathbf{b}) = 0$.

Now, if $(H_P)^{(\mathbf{i})}(\mathbf{X}) = 0$, then $(H_P)^{(\mathbf{i})}(\mathbf{b}) = 0$. Else $H_Q(\mathbf{X}) = (H_P)^{(\mathbf{i})}(\mathbf{X})$ (by item 3 of Proposition 2.3), and hence, as before, $(H_P)^{(\mathbf{i})}(\mathbf{b}) = H_Q(\mathbf{b}) = 0$. Since this is true for all $\mathbf{i}$ of weight at most $\ell - 1$, we have that $\operatorname{mult}(H_P, \mathbf{b}) \geq \ell$. ◻

Applying Corollary 2.8, and noting that $\ell q^n > d^* q^{n-1}$, we conclude that $H_P(\mathbf{X}) = 0$. This contradicts the fact that $P(\mathbf{X})$ is a nonzero polynomial.

Hence,

$$|K| \geq \frac{\binom{d+n}{n}}{\binom{m+n-1}{n}}.$$

Now, by our choice of d and m,

$$\frac{\binom{d+n}{n}}{\binom{m+n-1}{n}} = \frac{\binom{\ell q - 1 + n}{n}}{\binom{2\ell - \ell/q + n - 1}{n}} = \frac{\prod_{i=1}^n (\ell q - 1 + i)}{\prod_{i=1}^n (2\ell - \ell/q - 1 + i)}.$$

Since this is true for all ℓ such that ℓ is a multiple of q, we get that

$$|K| \geq \lim_{\ell \to \infty} \prod_{i=1}^n \left(\frac{q - 1/l + i/l}{2 - 1/q - 1/l + i/l} \right) = \left(\frac{q}{2 - 1/q} \right)^n. \qquad ◻$$

4. Statistical Kakeya for curves. Next we extend the results of the previous section to a form conducive to analyzing the mergers of Dvir and Wigderson [4]. The extension changes two aspects of the consideration in Kakeya sets, which we refer to as "statistical" and "curves." We describe these terms below.

In the setting of Kakeya sets, we were given a set K such that for *every* direction, there was a line in that direction such that *every* point on the line was contained in K. In the *statistical* setting we replace both occurrences of the "every" quantifier with a weaker "for many" quantifier. So we consider sets that satisfy the condition that for many directions, there exists a line in that direction intersecting K at many points.

A second change we make is that we now consider curves of higher degree and not just lines. We also do not consider curves in various *directions*, but rather curves

passing through a given set of special points. We start with formalizing the terms "curves," "degree," and "passing through a given point."

A *curve of degree* Λ in $\mathbb{F}_q^n$ is a tuple of polynomials $C(X) = (C_1(X), \ldots, C_n(X)) \in \mathbb{F}_q[X]^n$ such that $\max_{i \in [n]} \deg(C_i(X)) = \Lambda$. A curve C naturally defines a map from $\mathbb{F}_q$ to $\mathbb{F}_q^n$. For $\mathbf{x} \in \mathbb{F}_q^n$, we say that a curve C *passes through* $\mathbf{x}$ if there is a $t \in \mathbb{F}_q$ such that $C(t) = \mathbf{x}$.

We now state and prove our statistical version of the Kakeya theorem for curves.

THEOREM 4.1 (statistical Kakeya for curves). *Let $\lambda > 0$, $\eta > 0$. Let $\Lambda > 0$ be an integer such that $\eta q > \Lambda$. Let $S \subseteq \mathbb{F}_q^n$ be an arbitrary set satisfying $|S| = \lambda q^n$. Let $K \subseteq \mathbb{F}_q^n$ be such that for each $\mathbf{x} \in S$, there exists a curve $C_{\mathbf{x}}$ of degree at most Λ that passes through $\mathbf{x}$ and intersects K in at least ηq points. Then,*

$$|K| \geq \left(\frac{\lambda q}{\Lambda \left(\frac{\lambda q - 1}{\eta q} \right) + 1} \right)^n.$$

In particular, if $\lambda \geq \eta$, then we get that $|K| \geq (\frac{\eta q}{\Lambda + 1})^n$.

Observe that when $\lambda = \eta = 1$ and $\Lambda = 1$, we get the same bound as that for Kakeya sets as obtained in Theorem 3.2.

Proof. Let ℓ be a large integer and let

$$d = \lambda \ell q - 1,$$

$$m = \Lambda \frac{\lambda \ell q - \ell}{\eta q} + \ell.$$

By our choice of m and d, we have $\eta q(m - (\ell - 1)) > \Lambda(d - (\ell - 1))$. Since $\eta q > \Lambda$, we have that for all w such that $0 \leq w \leq \ell - 1$, $\eta q(m - w) > \Lambda(d - w)$. Just as in the proof of Theorem 3.2, we will prove that

$$|K| \geq \frac{\binom{d+n}{n}}{\binom{m+n-1}{n}} \geq \alpha^n,$$

where $\alpha \to \frac{\lambda q}{\Lambda(\frac{\lambda q - 1}{\eta q}) + 1}$ as $\ell \to \infty$.

Assume for contradiction that $|K| < \frac{\binom{d+n}{n}}{\binom{m+n-1}{n}}$. Then, as before, by Proposition 3.1, there exists a nonzero polynomial $P(\mathbf{X}) \in \mathbb{F}_q[\mathbf{X}]$ of total degree d^*, where $d^* \leq d$, such that $\text{mult}(P, \mathbf{a}) \geq m$ for every $\mathbf{a} \in K$. We will deduce that in fact P must vanish on all points in S with multiplicity ℓ. We will then get the desired contradiction from Corollary 2.8.

CLAIM 4.2. *For each $\mathbf{x}_0 \in S$,*

$$\text{mult}(P, \mathbf{x}_0) \geq \ell.$$

Proof. Fix any $\mathbf{i}$ with $\text{wt}(\mathbf{i}) = w \leq \ell - 1$. Let $Q(\mathbf{X}) = P^{(\mathbf{i})}(\mathbf{X})$. Note that $Q(\mathbf{X})$ is a polynomial of degree at most $d^* - w$. By Lemma 2.4, for all points $\mathbf{a} \in K$, $\text{mult}(Q, \mathbf{a}) \geq m - w$.

Let $C_{\mathbf{x}_0}$ be the curve of degree Λ through $\mathbf{x}_0$ that intersects K in at least ηq points. Let $t_0 \in \mathbb{F}_q$ be such that $C_{\mathbf{x}_0}(t_0) = \mathbf{x}_0$. Let $Q_{\mathbf{x}_0}(T)$ be the polynomial $Q \circ C_{\mathbf{x}_0}(T) \in \mathbb{F}_q[T]$. Then $Q_{\mathbf{x}_0}(T)$ is a univariate polynomial of degree at most $\Lambda(d^* - w)$. By Corollary 2.6, for all points $t \in \mathbb{F}_q$ such that $C_{\mathbf{x}_0}(t) \in K$, $Q_{\mathbf{x}_0}(T)$

vanishes at t with multiplicity $m - w$. Since the number of such points t is at least ηq, we get that $Q_{\mathbf{x}_0}(T)$ has at least $\eta q(m - w)$ zeroes (counted with multiplicity). However, by our choice of parameters, we know that

$$\eta q(m - w) > \Lambda(d - w) \geq \Lambda(d^* - w) \geq \deg(Q_{\mathbf{x}_0}(T)).$$

Since the degree of $Q_{\mathbf{x}_0}(T)$ is strictly less than the number of its zeroes, $Q_{\mathbf{x}_0}(T)$ must be identically zero. Thus we get $Q_{\mathbf{x}_0}(t_0) = Q(C_{\mathbf{x}_0}(t_0)) = Q(\mathbf{x}_0) = 0$. Hence $P^{(\mathbf{i})}(\mathbf{x}_0) = 0$. Since this is true for all $\mathbf{i}$ with $\mathrm{wt}(\mathbf{i}) \leq \ell - 1$, we conclude that $\mathrm{mult}(P, \mathbf{x}_0) \geq \ell$. □

Thus P vanishes at every point in S with multiplicity ℓ. As $P(\mathbf{X})$ is a nonzero polynomial, Corollary 2.8 implies that $\ell|S| \leq d^* q^{n-1}$. Hence $\ell\lambda q^n \leq dq^{n-1}$, which contradicts the choice of d.

Thus $|K| \geq \frac{\binom{d+n}{n}}{\binom{m+n-1}{n}}$. By choice of d and m,

$$|K| \geq \frac{\binom{\lambda\ell q - 1 + n}{n}}{\binom{\Lambda\frac{\lambda\ell q - 1 - (\ell - 1)}{\eta q} + \ell + n - 1}{n}}.$$

Picking ℓ arbitrarily large, we conclude that

$$|K| \geq \lim_{\ell \to \infty} \frac{\binom{\lambda\ell q - 1 + n}{n}}{\binom{\Lambda\frac{\lambda\ell q - 1 - (\ell - 1)}{\eta q} + \ell + n - 1}{n}}$$

$$= \lim_{\ell \to \infty} \left(\frac{\ell\lambda q - 1}{\ell\Lambda\left(\frac{\lambda q - 1}{\eta q}\right) + \ell} \right)^n = \left(\frac{\lambda q}{\Lambda\left(\frac{\lambda q - 1}{\eta q}\right) + 1} \right)^n. □$$

5. Improved mergers. In this section we state and prove our main result on randomness mergers.

5.1. Definitions and theorem statement. We start by recalling some basic quantities associated with random variables. The *statistical distance* between two random variables X and Y taking values from a finite domain Ω is defined as

$$\max_{S \subseteq \Omega} |\mathbf{Pr}[\mathsf{X} \in S] - \mathbf{Pr}[\mathsf{Y} \in S]|.$$

We say that X is ϵ-*close* to Y if the statistical distance between X and Y is at most ϵ; otherwise we say that X and Y are ϵ-*far*. The *min-entropy* of a random variable X is defined as

$$H_\infty(\mathsf{X}) \triangleq \min_{x \in \mathrm{supp}(\mathsf{X})} \log_2\left(\frac{1}{\mathbf{Pr}[\mathsf{X} = x]}\right).$$

Note that the notion of having high min-entropy is closed under convex combinations. Specifically, if $0 \leq \alpha \leq 1$ and Y and $\mathbb{Z}$ are random variables supported on Ω with min-entropy at least m and X is the random variable satisfying $\mathbf{Pr}[\mathsf{X} = x] = \alpha\mathbf{Pr}[\mathsf{Y} = x] + (1 - \alpha)\mathbf{Pr}[\mathbb{Z} = x]$, then X also has min-entropy at least m.

We say that a random variable X is ϵ-close to having min-entropy m if there exists a random variable Y of min-entropy m such that X is ϵ-close to Y.

A "merger" of randomness takes a Λ-tuple of random variables and "merges" their randomness to produce a high-entropy random variable, provided that the Λ-tuple is "somewhere-random" as defined below.

DEFINITION 5.1 (somewhere-random source). *For integers Λ and N, a* simple *(N, Λ)-somewhere-random* source *is a random variable* $\mathsf{A} = (\mathsf{A}_1, \ldots, \mathsf{A}_\Lambda)$ *taking values in S^Λ, where S is some finite set of cardinality 2^N, such that for some $i_0 \in [\Lambda]$, the distribution of A_{i_0} is uniform over S. An (N, Λ)-somewhere-random source* is a *convex combination of simple (N, Λ)-somewhere-random sources. (When N and Λ are clear from context, we refer to the source as simply a "somewhere-random source.")*

We are now ready to define a merger.

DEFINITION 5.2 (merger). *For positive integer Λ and set S of size 2^N, a function $f : S^\Lambda \times \{0, 1\}^d \to S$ is called an (m, ϵ)-merger (of (N, Λ)-somewhere-random sources), if for every (N, Λ) somewhere-random source $\mathsf{A} = (\mathsf{A}_1, \ldots, \mathsf{A}_\Lambda)$ taking values in S^Λ, and for B being uniformly distributed over $\{0, 1\}^d$, the distribution of $f((\mathsf{A}_1, \ldots, \mathsf{A}_\Lambda), \mathsf{B})$ is ϵ-close to having min-entropy m.*

A merger thus has five parameters associated with it: N, Λ, m, ϵ, and d. The general goal is to give explicit constructions of mergers of (N, Λ)-somewhere-random sources for every choice of N and Λ, for as large an m as possible, and with ϵ and d being as small as possible. Known mergers attain $m = (1 - \delta) \cdot N$ for arbitrarily small δ, and our goal will be to achieve $\delta = o(1)$ as a function of N, while ϵ is an arbitrarily small positive real number. Thus our main concern is the growth of d as a function of N and Λ. Prior to this work, the best known bounds required either $d = \Omega(\log N + \log \Lambda)$ [4] or $d = \Omega(\Lambda)$ [9]. We require only $d = \Omega(\log \Lambda)$.

THEOREM 5.3. *For every $\epsilon, \delta > 0$ and integers N, Λ, there exists a $((1-\delta) \cdot N, \epsilon)$-merger of (N, Λ)-somewhere-random sources, computable in polynomial time, with seed length*

$$d = \frac{1}{\delta} \cdot \log_2 \left(\frac{2\Lambda}{\epsilon} \right).$$

5.2. The curve merger of [4] and its analysis. The merger that we consider is a very simple one proposed by Dvir and Wigderson [4], and we improve their analysis using our extended method of multiplicities. We note that they used the polynomial method in their analysis; the basic method of multiplicities doesn't seem to improve their analysis.

The curve merger of [4], denoted f_{DW}, is obtained as follows. Let $q \geq \Lambda$ be a prime power, and let n be any integer. Fix some Λ distinct elements of $\mathbb{F}_q$ and let these be denoted $\gamma_1, \ldots, \gamma_\Lambda$. Let $c_i(T) \in \mathbb{F}_q[T]$ be the unique degree $\Lambda - 1$ polynomial with $c_i(\gamma_i) = 1$ and for all $j \neq i$, $c_i(\gamma_j) = 0$. Then the curve merger f_{DW} maps $(\mathbb{F}_q^n)^\Lambda \times \mathbb{F}_q$ to $\mathbb{F}_q^n$ as follows:

$$f_{\mathrm{DW}}((\mathbf{x}_1, \ldots, \mathbf{x}_\Lambda), u) = \sum_{i=1}^{\Lambda} c_i(u) \mathbf{x}_i,$$

where $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_\Lambda) \in (\mathbb{F}_q^n)^\Lambda$ and $u \in \mathbb{F}_q$. In other words, $f_{\mathrm{DW}}((\mathbf{x}_1, \ldots, \mathbf{x}_\Lambda), u)$ picks the (canonical) curve passing through $\mathbf{x}_1, \ldots, \mathbf{x}_\Lambda$ and outputs the uth point on the curve.

THEOREM 5.4. *Let $q \geq \Lambda$ and let A be a somewhere-random source taking values in $(\mathbb{F}_q^n)^\Lambda$. Let B be distributed uniformly over $\mathbb{F}_q$, with A, B independent. Let $\mathsf{C} = f_{\mathrm{DW}}(\mathsf{A}, \mathsf{B})$. Then for*

$$q \geq \left(\frac{2\Lambda}{\epsilon} \right)^{\frac{1}{\delta}},$$

C *is ϵ-close to having min-entropy* $(1 - \delta) \cdot n \cdot \log_2 q$.

Theorem 5.3 easily follows from the above. We note that [4] proved a similar theorem assuming $q \geq \mathrm{poly}(n, \Lambda)$, forcing their seed length to grow logarithmically with n as well. We should also note at this point that a representation of a finite field $\mathbb{F}_q$ with q a power of 2 can be found efficiently (in time polynomial in $\log(q)$) by the results in [16]. This makes our construction explicit even for small ϵ and δ.

Proof of Theorem 5.3. Let $q = 2^d$, so that $q \geq \left(\frac{2\Lambda}{\epsilon}\right)^{\frac{1}{\delta}}$, and let $n = N/d$. Then we may identify $\mathbb{F}_q$ with $\{0,1\}^d$ and $\mathbb{F}_q^n$ with $\{0,1\}^N$. Take f to be the function f_{DW} given earlier. Clearly f is computable in the claimed time. Theorem 5.4 shows that f has the required merger property.   ◻

We now prove Theorem 5.4.

Proof of Theorem 5.4. We first note that it suffices to prove the theorem for the case that A is a simple somewhere-random source. Once this is done, to handle the general case when A is a convex combination of simple somewhere-random sources, we can simply use the fact that $f_{\mathrm{DW}}(\mathsf{A}, \mathbf{B})$ will be a convex combination of random variables that are ϵ-close to having high min-entropy, and this notion is closed under convex combinations.

Let A be a simple somewhere-random source. Let $m = (1 - \delta) \cdot n \cdot \log_2 q$. We wish to show that $f_{\mathrm{DW}}(\mathsf{A}, \mathsf{B})$ is ϵ-close to having min-entropy m.

Suppose not. Then there is a set $K \subseteq \mathbb{F}_q^n$ with $|K| \leq 2^m = q^{(1-\delta) \cdot n} \leq \left(\frac{\epsilon q}{2\Lambda}\right)^n$ such that

$$\Pr_{\mathsf{A}, \mathsf{B}}[f(\mathsf{A}, \mathsf{B}) \in K] \geq \epsilon.$$

To see why, consider the set K of the 2^m most "popular" values in the distribution (i.e., those that have the highest probabilities). If this set is "hit" with probability at most ϵ, then the distribution is clearly ϵ close to having min-entropy at least m.

Suppose that A_{i_0} is uniformly distributed over $\mathbb{F}_q^n$. Let A_{-i_0} denote the random variable

$$(\mathsf{A}_1, \ldots, \mathsf{A}_{i_0 - 1}, \mathsf{A}_{i_0 + 1}, \ldots, \mathsf{A}_\Lambda).$$

By an averaging argument, with probability at least $\lambda = \epsilon/2$ over the choice of A_{i_0}, we have

$$\Pr_{\mathsf{A}_{-i_0}, \mathsf{B}}[f(\mathsf{A}, \mathsf{B}) \in K] \geq \eta,$$

where $\eta = \epsilon/2$. Since A_{i_0} is uniformly distributed over $\mathbb{F}_q^n$, we conclude that there is a set S of cardinality at least λq^n such that for any $\mathbf{x} \in S$,

$$\Pr_{\mathsf{A}, \mathsf{B}}[f(\mathsf{A}, \mathsf{B}) \in K \mid \mathsf{A}_{i_0} = \mathbf{x}] \geq \eta.$$

By fixing the values of A_{-i_0} to preserve the above probability, we conclude that for each $\mathbf{x} \in S$, there is a $\mathbf{y} = \mathbf{y}(\mathbf{x}) = (\mathbf{y}_1, \ldots, \mathbf{y}_\Lambda)$ with $\mathbf{y}_{i_0} = x$ such that $\Pr_{\mathsf{B}}[f(\mathbf{y}, \mathsf{B}) \in K] \geq \eta$. Define the degree $\Lambda - 1$ curve $C_{\mathbf{x}}(T) = f(\mathbf{y}(\mathbf{x}), T) = \sum_{j=1}^\Lambda \mathbf{y}_j c_j(T)$. Recall that the c_j's come from the definition of f_{DW}. Then $C_{\mathbf{x}}$ passes through $\mathbf{x}$, since $C_{\mathbf{x}}(\gamma_{i_0}) = \sum_{j=1}^\Lambda \mathbf{y}_j c_j(\gamma_{i_0}) = \mathbf{y}_{i_0} = \mathbf{x}$, and $\Pr_{\mathsf{B} \in \mathbb{F}_q}[C_{\mathbf{x}}(\mathsf{B}) \in K] \geq \eta$ by definition of $C_{\mathbf{x}}$.

Thus S and K satisfy the hypothesis of Theorem 4.1. We now conclude that

$$|K| \geq \left(\frac{\lambda q}{(\Lambda - 1)\left(\frac{\lambda q - 1}{\eta q}\right) + 1} \right)^n = \left(\frac{\epsilon q/2}{\Lambda - (\Lambda - 1)/\eta q} \right)^n > \left(\frac{\epsilon q}{2\Lambda} \right)^n .$$

This is a contradiction, and the proof of the theorem is complete.   ☐

The somewhere-high-entropy case. It is possible to extend the merger analysis given above also to the case of *somewhere-high-entropy* sources. In this scenario the source is comprised of blocks, one of which has min-entropy at least r. One can then prove an analogue of Theorem 5.4 saying that the output of f_{DW} will be close to having min-entropy $(1 - \delta) \cdot r$ under essentially the same conditions on q. The proof is done by hashing the source using a random linear function into a smaller dimensional space and then applying Theorem 5.4 (in a black box manner). The reason this works is that the merger commutes with the linear map (for details, see [4]). The result is formally stated below. We do not give the details here since they are exactly the same as in [4]. We will not make use of this case in any of our other results.

We say that a random variable $X \in (\mathbb{F}_q^n)^\Lambda$ is a somewhere s-source if there exists an index $i \in [\Lambda]$ such that the min-entropy of X_i is at least $s \cdot \log(q)$ (one can also allow for convex combinations of such sources). Notice that s does not have to be an integer.

THEOREM 5.5. *For every $\epsilon, \delta > 0$, for all integers n, Λ, for every prime power $q \geq \Lambda$, and for every $s \geq 1$, the following holds: Let A be a somewhere-s-source taking values in $(\mathbb{F}_q^n)^\Lambda$. Let B be distributed uniformly over $\mathbb{F}_q$, with A, B independent. Let $C = f_{\mathrm{DW}}(A, B)$. Then for*

$$q \geq \left(\frac{4\Lambda}{\epsilon} \right)^{\frac{2}{\delta}},$$

C *is ϵ-close to having min-entropy $(1 - \delta) \cdot s \cdot \log_2 q$.*

6. Extractors with sublinear entropy loss. In this section we use our improved analysis of the curve merger to show the existence of an explicit extractor with logarithmic seed and sublinear entropy loss.

We will call a random variable X distributed over $\{0, 1\}^n$ with min-entropy k an (n, k)-*source.*

DEFINITION 6.1 (extractor). *A function $E : \{0,1\}^n \times \{0,1\}^d \mapsto \{0,1\}^m$ is a (k, ϵ)-extractor if for every (n, k)-source X, the distribution of $E(X, U_d)$ is ϵ-close to uniform, where U_d is a random variable distributed uniformly over $\{0, 1\}^d$, and X, U_d are independent. An extractor is called* explicit *if it can be computed in polynomial time.*

It is common to refer to the quantity $k - m$ in the above definition as the *entropy loss* of the extractor. The next theorem asserts the existence of an explicit extractor with logarithmic seed and sublinear entropy loss.

THEOREM 6.2 (basic extractor with sublinear entropy loss). *For every $c_1 \geq 1$, for all positive integers $k < n$ with $k \geq \log^2(n)$, there exists an explicit (k, ϵ)-extractor*

$E : \{0,1\}^n \times \{0,1\}^d \mapsto \{0,1\}^m$ *with*

$$d = O(c_1 \cdot \log(n)),$$

$$k - m = O\left(\frac{k \cdot \log\log(n)}{\log(n)}\right),$$

$$\epsilon = O\left(\frac{1}{\log^{c_1}(n)}\right).$$

The extractor of this theorem is constructed by composing several known explicit constructions of pseudorandom objects with the merger of Theorem 5.3. In section 6.1 we describe the construction of our basic extractor. We then show, in section 6.2, how to use the "repeated extraction" technique of Wigderson and Zuckerman [20] to boost this extractor and reduce the entropy loss to $k - m = O(k/\log^c n)$ for any constant c (while keeping the seed logarithmic). The end result is the following theorem.

THEOREM 6.3 (final extractor with sublinear entropy loss). *For every* $c_1, c_2 \geq 1$, *for all positive integers* $k < n$, *there exists an explicit* (k, ϵ)-*extractor* $E : \{0,1\}^n \times \{0,1\}^d \mapsto \{0,1\}^m$ *with*

$$d = O(c_1 c_2 \cdot \log(n)),$$

$$k - m = O\left(\frac{k}{\log^{c_2}(n)}\right),$$

$$\epsilon = O\left(\frac{1}{\log^{c_1}(n)}\right).$$

6.1. Proof of Theorem 6.2. Note that we may equivalently view an extractor $E : \{0,1\}^n \times \{0,1\}^d \rightarrow \{0,1\}^m$ as a randomized algorithm $E : \{0,1\}^n \rightarrow \{0,1\}^m$ which is allowed to use d uniformly random bits. We will present the extractor E as such an algorithm, which takes five major steps.

Before giving the formal proof we give a high level description of our extractor. Our first step is to apply the lossless condenser of [7] to output a string of length $2k$ with min-entropy k (thus reducing our problem to the case $k = \Omega(n)$). The construction continues along the lines of [4]. In the second step, we partition our source (now of length $n' = 2k$) into $\Lambda = \log(n)$ consecutive blocks $X_1, \ldots, X_\Lambda \in \{0,1\}^{n'/\Lambda}$ of equal length. We then consider the Λ possible ways of partitioning the source into a prefix of j blocks and suffix of $\Lambda - j$ blocks for j between 1 and Λ. By a result of Ta-Shma [19], after passing to a convex combination, one of these partitions is a (k', k_2) block source with k' being at least $k - O(k/\Lambda)$ and k_2 being at least polylogarithmic in k. In the third step we use a block source extractor (from [12]) on each of the possible Λ partitions (using the same seed for each partition) to obtain a somewhere-random source with block length k'. The fourth step is to merge this somewhere-random source into a single block of length k' and entropy $k' \cdot (1 - \delta)$ with δ subconstant. In view of our new merger parameters and the fact that Λ (the number of blocks) is small enough, we can get away with choosing $\delta = \log\log(n)/\log(n)$ and keeping the seed logarithmic and the error polylogarithmic. To finish the construction (the fifth step) we need to extract almost all the entropy from a source of length k' and entropy $k' \cdot (1 - \delta)$. This can be done (using techniques from [12]) with logarithmic seed and an additional entropy loss of $O(\delta \cdot k')$.

Proof of Theorem 6.2. We now formally prove Theorem 6.2. Let $\mathcal{X}$ be the input (n, k)-*source.* It would be convenient for us to assume during the proof that k is not too small. This is accomplished by noticing that for small k, there already exist very good extractors. Formally, using Theorem 6.14 below, we can assume that

$$k > 2^{\sqrt{\log n}}.$$

We begin by reducing to the case where $n = O(k)$ using the lossless condensers of [7].

THEOREM 6.4 (lossless condenser [7]).   *For all positive integers* $k < n$ *with* $k = \omega(\log(n))$, *there exists an explicit function* $C_{\mathrm{GUV}} : \{0,1\}^n \times \{0,1\}^{d_1} \mapsto \{0,1\}^{n'}$ *with* $n' = 2k$, $d_1 = O(\log(n))$, *such that for every* (n, k)-*source* X, $C(\mathsf{X}, \mathsf{U}_{d_1})$ *is* $(1/n)$-*close to an* (n', k)-*source, where* U_{d_1} *is distributed uniformly over* $\{0,1\}^{d_1}$, *and* $\mathsf{X}, \mathsf{U}_{d_1}$ *are independent.*

Step 1: Pick U_{d_1} uniformly from $\{0,1\}^{d_1}$. Compute $\mathcal{X}_1 = C_{\mathrm{GUV}}(\mathcal{X}, \mathsf{U}_{d_1})$.

By the above theorem, $\mathcal{X}_1$ is $(1/n)$-close to an (n', k)-*source,* where $n' = 2k$. The seed length used in this step is $d_1 = O(\log(n))$. Our next goal is to produce a *somewhere-block source.* We now define these formally.

DEFINITION 6.5 (block source).   *Let* $\mathsf{X} = (\mathsf{X}_1, \mathsf{X}_2)$ *be a random source over* $\{0,1\}^{n_1} \times \{0,1\}^{n_2}$. *We say that* X *is a* (k_1, k_2)-*block source if* X_1 *is an* (n_1, k_1)-*source and for each* $x_1 \in \{0,1\}^{n_1}$ *the conditional random variable* $\mathsf{X}_2 | \mathsf{X}_1 = x_1$ *is an* (n_2, k_2)-*source.*

DEFINITION 6.6 (somewhere-block source).   *Let* $\mathsf{X} = (\mathsf{X}_1, \ldots, \mathsf{X}_\Lambda)$ *be a random variable such that each* X_i *is distributed over* $\{0,1\}^{n_{i,1}} \times \{0,1\}^{n_{i,2}}$. *We say that* X *is a* simple somewhere-(k_1, k_2)-block source *if there exists* $i \in [\Lambda]$ *such that* X_i *is a* (k_1, k_2)-*block source. We say that* X *is a* somewhere-(k_1, k_2)-block source *if* X *is a convex combination of simple somewhere-(k_1, k_2)-block sources.*

We now state a result of Ta-Shma [19] which converts an arbitrary source into a somewhere-block source. This is the first step in the proof of Theorem 1 of [19, p. 44]. (Theorem 1 shows how to convert any arbitrary source to a somewhere-block source, and then does more by showing how one could extract from such a source.)

Let Λ be an integer, and assume for simplicity of notation that n' is divisible by Λ. Let

$$\mathsf{X} = (\mathsf{X}_1, \ldots, \mathsf{X}_\Lambda) \in \left(\{0,1\}^{n'/\Lambda} \right)^\Lambda$$

denote the partition of X into Λ blocks. For every $1 \leq j < \Lambda$ we denote

$$\mathsf{Y}_j = (\mathsf{X}_1, \ldots, \mathsf{X}_j),$$

$$\mathsf{Z}_j = (\mathsf{X}_{j+1}, \ldots, \mathsf{X}_\Lambda).$$

Consider the function $B_{\mathrm{TS}}^\Lambda : \{0,1\}^{n'} \to (\{0,1\}^{n'})^{\Lambda-1}$, where

$$B_{\mathrm{TS}}^\Lambda(X) = ((\mathsf{Y}_1, \mathsf{Z}_1), (\mathsf{Y}_2, \mathsf{Z}_2), \ldots, (\mathsf{Y}_{\Lambda-1}, \mathsf{Z}_{\Lambda-1})).$$

The next theorem shows that the source $((\mathsf{Y}_j, \mathsf{Z}_j))_{j \in [\Lambda-1]}$ is close to a somewhere-block source.

THEOREM 6.7 (see [19]).   *Let* Λ *be an integer. Let* $k = k_1 + k_2 + s$. *Then the function* $B_{\mathrm{TS}}^\Lambda : \{0,1\}^{n'} \to (\{0,1\}^{n'})^{\Lambda-1}$ *is such that for any* (n', k)-*source* X, *letting*

$X' = B_{\mathrm{TS}}^\Lambda(X)$, we have that X' is $O(n' \cdot 2^{-s})$-close to a somewhere-$(k_1 - O(n'/\Lambda), k_2)$-block source.

Step 2: Set $\Lambda = \log(n)$. Compute $\mathcal{X}_2 = (\mathcal{X}_{2,1}, \mathcal{X}_{2,2}, \ldots, \mathcal{X}_{2,\Lambda}) = B_{\mathrm{TS}}^\Lambda(\mathcal{X}_1)$.

Note that $\mathcal{X}_2$ is a random variable taking values in $(\{0,1\}^{n'})^{\Lambda-1}$.

Plugging $k_2 = O(\log^4(n')) = O(\log^4(k))$, $s = O(\log n)$, and $k_1 = k - k_2 - s$ into the above theorem, we conclude that $\mathcal{X}_2$ is $n'^{-\Omega(1)}$-close to a somewhere-(k', k_2)-block source, where

$$k' = k_1 - O(n'/\log(n)) = k - k_2 - s - O(k/\log(n)) = k - O(k/\log(n)),$$

where for the last inequality we use the fact that $k > \log^2(n)$ and so both s and k_2 are bounded by $O(k/\log(n))$.

We next use the block source extractor from [12] to convert the above somewhere-block source to a somewhere-random source.

THEOREM 6.8 (see [12]). *Let $n' = n_1 + n_2$ and let k', k_2 be such that $k_2 > \log^4(n_1)$. Then there exists an explicit function $E_{\mathrm{RSW}} : \{0,1\}^{n_1} \times \{0,1\}^{n_2} \times \{0,1\}^{d_2} \mapsto \{0,1\}^{m''}$ with $m'' = k'$, $d_2 = O(\log(n'))$, such that for any (k', k_2)-block source X, $E_{\mathrm{RSW}}(\mathsf{X}, \mathsf{U}_{d_2})$ is $(n_1)^{-\Omega(1)}$-close to the uniform distribution over $\{0,1\}^{m''}$, where U_{d_2} is distributed uniformly over $\{0,1\}^{d_2}$, and $\mathsf{X}, \mathsf{U}_{d_2}$ are independent.*

Set $d_2 = O(\log(n'))$ as in Theorem 6.8.

Step 3: Pick U_{d_2} uniformly from $\{0,1\}^{d_2}$.
$\forall j \in [\Lambda]$, compute $\mathcal{X}_{3,j} = E_{\mathrm{RSW}}(\mathcal{X}_{2,j}, U_{d_2})$.

Set $\mathcal{X}_3 = (\mathcal{X}_{3,1}, \ldots, \mathcal{X}_{3,\Lambda})$. By the above theorem, $\mathcal{X}_3$ is $n'^{-\Omega(1)}$-close to a somewhere-random source.

Note that $\mathcal{X}_3$ is a random variable taking values in $(\{0,1\}^{m''})^{\Lambda-1}$, where $m'' = k' = k - O(k/\log(n))$, one of $\{\mathcal{X}_{3,1}, \ldots, \mathcal{X}_{3,\Lambda}\}$ is $(n_1)^{-\Omega(1)}$-close to being uniform, and $n_1 \geq n'/\Lambda \geq 2k/\log n$ (by definition of the function B_{TS}^Λ). Since the original source we started with had min-entropy k, and $\mathcal{X}_3$ is $n_1^{-\Omega(1)}$ close to having min-entropy at least $k - O(k/\log(n))$, so far we have lost only a sublinear fraction of the min-entropy of the original source.

We are now ready to use the merger M from Theorem 5.3. We invoke that theorem with entropy loss $\delta = \log\log(n)/\log(n)$ and error $\epsilon = \frac{1}{\log^{c_1}(n)}$, and hence M has a seed length of

$$d_3 = O\left(\frac{1}{\delta} \log \frac{\Lambda}{\epsilon}\right) = O(c_1 \log(n)).$$

Step 4: Pick U_{d_3} uniformly from $\{0,1\}^{d_3}$.
Compute $\mathcal{X}_4 = M(\mathcal{X}_3, \mathsf{U}_{d_3})$.

By Theorem 5.3, $\mathcal{X}_4$ is $O(\frac{1}{\log^{c_1}(n)})$-close to a $(k', (1-\delta)k')$-source, where $k' = k - O(k/\log k)$. Note that $\delta = o(1)$, and thus $\mathcal{X}_4$ has nearly full entropy. We now apply an extractor for sources with extremely high entropy rate, given by the following lemma.

LEMMA 6.9. *For any k' and $\delta > 0$, there exists an explicit $(k'(1-\delta), k'^{-\Omega(1)})$-extractor $E_{\mathrm{HIGH}} : \{0,1\}^{k'} \times \{0,1\}^{d_4} \mapsto \{0,1\}^{(1-3\delta)k'}$ with $d_4 = O(\log(k'))$.*

Proof. The proof of this lemma will follow from Theorem 6.8. Initially, the input is partitioned into two blocks of length $k'(1-\delta)-2\log^4 k'$ and $\delta k'+2\log^4 k'$. Intuitively, this partition should be a block source, since fixing the first block cannot "kill" all of the entropy of the source. Formally, using Lemma 6.2 from [13], one knows that this partition is $1/k'$-close to a $(k'(1-2\delta)-2\log^4 k',\log^4 k')$-block source. This block source is then passed through the block source extractor of Theorem 6.8, which can extract $k'(1-2\delta)-2\log^4 k' \geq k'(1-3\delta)$ bits with polynomially small error. $\square$

> **Step 5:** Pick U_{d_4} uniformly from $\{0,1\}^{d_4}$.
> Compute $\mathcal{X}_5 = E_{\text{HIGH}}(\mathcal{X}_4, \mathsf{U}_{d_4})$. *Output $\mathcal{X}_5$.*

This completes the description of the extractor E. It remains for us to note that d, the total number of random bits used, is at most $d_1 + d_2 + d_3 + d_4 = O(c_1 \log n)$. The output $\mathcal{X}_5$ is ϵ-close to uniformly distributed over

$$\{0,1\}^{(1-3\delta)k'} = \{0,1\}^{k-O(k \cdot \frac{\log\log n}{\log n})},$$

where ϵ can be bounded by the *sum* of all the errors in all of the above steps. The errors introduced in all steps, other than the merger step (Step 3), were polynomially small in k. Since $k > 2^{\sqrt{\log n}}$, this is negligible compared to the $(\log^{c_1}(n))^{-1}$ error introduced in the merger step. Thus, the final error of the construction is dominated by this last quantity and so the error is as required. This completes the proof of Theorem 6.2. $\square$

We summarize the transformations in the following diagram.
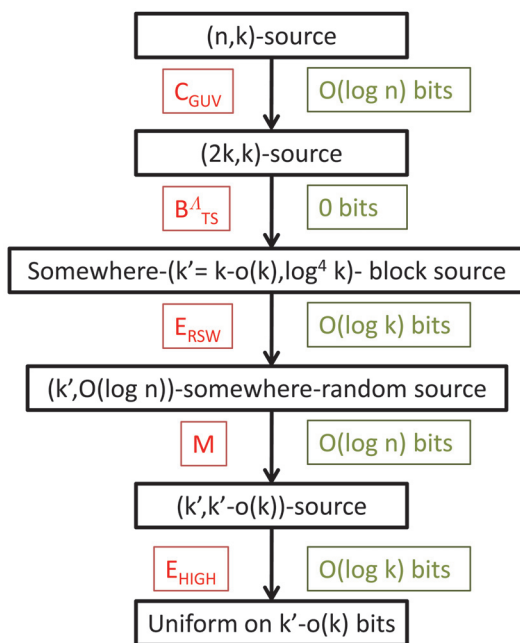
6.2. Improving the output length by repeated extraction. We now use some ideas from [12] and [20] to extract an even larger fraction of the min-entropy out of the source. This will prove Theorem 6.3. We first prove a variant of the theorem with a restriction on k. This restriction will be later removed using known constructions of extractors for low min-entropy.

THEOREM 6.10 (explicit extractor with improved sublinear entropy loss). *For every $c_1, c_2 \geq 1$, for all positive integers $k < n$ with $k = \log^{\omega(1)}(n)$, there exists an explicit (k, ϵ)-extractor $E : \{0,1\}^n \times \{0,1\}^d \mapsto \{0,1\}^m$ with*

$$d = O(c_1 c_2 \cdot \log(n)),$$

$$k - m = O\left(\frac{k}{\log^{c_2}(n)}\right),$$

$$\epsilon = O\left(\frac{1}{\log^{c_1}(n)}\right).$$

We first transform the extractor given in Theorem 6.2 into a *strong* extractor (defined below) via [12, Theorem 8.2] (which gives a generic way of getting a strong extractor from any extractor). We then use a trick from [20] that repeatedly uses the same extractor with independent seeds to extract the "remaining entropy" from the source, thus improving the entropy loss.

DEFINITION 6.11. *A (k, ϵ)-extractor $E : \{0,1\}^n \times \{0,1\}^d \mapsto \{0,1\}^m$ is strong if for every (n, k)-source X, the distribution of $(E(X, U_d), U_d)$ is ϵ-close to the uniform distribution over $\{0,1\}^{m+d}$, where U_d is distributed uniformly over $\{0,1\}^d$, and X, U_d are independent.*

THEOREM 6.12 (see [12, Theorem 8.2]). *Any explicit (k, ϵ)-extractor $E : \{0,1\}^n \times \{0,1\}^d \mapsto \{0,1\}^m$ can be transformed into an explicit strong $(k, O(\sqrt{\epsilon}))$-extractor $E' : \{0,1\}^n \times \{0,1\}^{O(d)} \mapsto \{0,1\}^{m-d-2\log(1/\epsilon)-O(1)}$.*

THEOREM 6.13 (see [20, Lemma 2.4]). *Let $E_1 : \{0,1\}^n \times \{0,1\}^{d_1} \mapsto \{0,1\}^{m_1}$ be an explicit strong (k, ϵ_1)-extractor, and let $E_2 : \{0,1\}^n \times \{0,1\}^{d_2} \mapsto \{0,1\}^{m_2}$ be an explicit strong $(k - (m_1 + r), \epsilon_2)$-extractor. Then the function*

$$E_3 : \{0,1\}^n \times \left(\{0,1\}^{d_1} \times \{0,1\}^{d_2}\right) \mapsto \{0,1\}^{m_1+m_2}$$

defined by

$$E_3(x, y_1, y_2) = E_1(x, y_1) \circ E_2(x, y_2)$$

is a strong $(k, \epsilon_1 + \epsilon_2 + 2^{-r})$-extractor.

Proof of Theorem 6.10. Let E be the (k, ϵ)-extractor with seed $O(c_1 \log n)$ of Theorem 6.2. By Theorem 6.12, we get an explicit strong $(k, \sqrt{\epsilon})$-extractor E' with entropy loss $O(k \frac{\log \log n}{\log n})$. We now iteratively apply Theorem 6.13 as follows. Let $E^{(0)} = E'$. For each $1 < i \leq O(c_2)$, let $E^{(i)} : \{0,1\}^n \times \{0,1\}^{d_i} \to \{0,1\}^{m_i}$ be the strong (k, ϵ_i)-extractor produced by Theorem 6.13 when we take $E_1 = E^{(i-1)}$ and E_2 to be the strong $(k - m_{i-1} - c_1 \log n, 1/\log^{c_1}(n))$-extractor with seed length $O(c_1 \log n)$ given by Theorems 6.2 and 6.12. Thus,

$$d_i = O(i c_1 \log n),$$

$$\epsilon_i = O\left(\frac{i}{\log^{c_1}(n)}\right),$$

$$m_i = m_{i-1} + (k - m_{i-1} - c_1 \log n)\left(1 - O\left(\frac{\log \log n}{\log n}\right)\right).$$

Thus the entropy loss of $E^{(i)}$ is given by

$$k - m_i = (k - m_{i-1})\left(1 - \left(1 - O\left(\frac{\log\log n}{\log n}\right)\right)\right) + O(c_1 \log n) = O\left(\frac{k \cdot \log\log(n)^i}{\log^i(n)}\right).$$

$E^{(O(c_2))}$ is the desired extractor. $\square$

Remark. In fact [7] and [11] show how to extract all the min-entropy with poly-logarithmic seed length. Combined with the lossless condenser of [7], this gives an extractor that uses logarithmic seed to extract all the min-entropy from sources that have min-entropy rate at most $2^{O(\sqrt{\log n})}$.

THEOREM 6.14 (corollary of [7, Theorem 4.21]). *For all positive integers $n \geq k$ such that $k = 2^{O(\sqrt{\log n})}$, and for all $\epsilon > 0$, there exists an explicit (k, ϵ)-extractor $E : \{0,1\}^n \times \{0,1\}^d \mapsto \{0,1\}^m$ with $d = O(\log(n))$ and $m = k + d - 2\log(1/\epsilon) - O(1)$.*

This result combined with Theorem 6.10 gives an extractor with improved sub-linear entropy loss that works for sources of all entropy rates, thus completing the proof of Theorem 6.3.

7. Bounds on the list size for list-decoding Reed–Solomon codes. In this section, we give a simple algebraic proof of an upper bound on the list size for list-decoding Reed–Solomon codes within the Johnson radius.

Before stating and proving the theorem, we need some definitions. For a bivariate polynomial $P(X,Y) \in \mathbb{F}[X,Y]$, we define its (a,b)-degree to be the maximum of $ai + bj$ over all (i,j) such that the monomial $X^i Y^j$ appears in $P(X,Y)$ with a nonzero coefficient. Let $N(k,d,\theta)$ be the number of monomials $X^i Y^j$ which have $(1,k)$-degree at most d and $j \leq \theta d/k$. We have the following simple fact.

FACT 7.1. *For any $k < d$ and $\theta \in [0,1]$, $N(k,d,\theta) > \theta \cdot (2 - \theta) \cdot \frac{d^2}{2k}$.*

Proof. Letting $t = \lfloor \theta d/k \rfloor$, note that

$$N(k,d,\theta) = \sum_{j=0}^{t}(d+1-kj) = (t+1)(d+1) - k\binom{t+1}{2} = (t+1)(d+1-kt/2).$$

Using $t \leq \theta d/k \leq t+1$, we get $N(k,d,\theta) \geq (\theta d/k)(d+1-\theta d/2) > \theta \cdot (2-\theta) \cdot \frac{d^2}{2k}$. $\square$

Now we prove the main theorem of this section. The proof is an enhancement of the original analysis of the Guruswami–Sudan algorithm using the extended method of multiplicities.

THEOREM 7.2 (list size bound for Reed–Solomon codes). *Let*

$$(\alpha_1, \beta_1), \ldots, (\alpha_n, \beta_n) \in \mathbb{F}^2.$$

Let $R, \gamma \in [0,1]$ with $\gamma^2 > R$. Let $k = Rn$. Let $f_1(X), \ldots, f_L(X) \in \mathbb{F}[X]$ be polynomials of degree at most k such that for each $j \in [L]$ we have $|\{i \in [n] : f_j(\alpha_i) = \beta_i\}| > \gamma n$. Then $L \leq \frac{2\gamma}{\gamma^2 - R}$.

Proof. Let $\epsilon > 0$ be a parameter. Let $\theta = \frac{2}{\left(1 + \frac{\gamma^2}{R}\right)}$. Let m be a large integer (to be chosen later), and let $d = (1 + \epsilon) \cdot m \cdot \sqrt{\frac{nk}{\theta \cdot (2 - \theta)}}$. We first interpolate a nonzero polynomial $P(X,Y) \in \mathbb{F}[X,Y]$ of $(1,k)$-degree at most d and Y-degree at most $\theta d/k$, that vanishes with multiplicity at least m at each of the points (α_i, β_i). Such a polynomial exists if $N(k,d,\theta)$, the number of monomials available, is larger than the number of homogeneous linear constraints imposed by the vanishing conditions

$$(7.1) \qquad \frac{m(m+1)}{2} \cdot n < N(k,d,\theta).$$

This can be made to hold by picking m sufficiently large, since by Fact 7.1,

$$N(k, d, \theta) > \theta \cdot (2 - \theta) \frac{d^2}{2k} = \frac{(1 + \epsilon)^2 m^2}{2} \cdot n.$$

Having obtained the polynomial $P(X, Y)$, we also view it as a univariate polynomial $Q(Y) \in \mathbb{F}(X)[Y]$ with coefficients in $\mathbb{F}(X)$, the field of rational functions in X.

Now let $f(X)$ be any polynomial of degree at most k such that, letting $I = \{i \in [n] : f(\alpha_i) = \beta_i\}$, $|I| \geq A$. We claim that the polynomial $Q(Y)$ vanishes at $f(X)$ with multiplicity at least $m - d/A$. Indeed, fix an integer $j < m - d/A$, and let $R_j(X) = Q^{(j)}(f(X)) = P^{(0,j)}(X, f(X))$. Notice that the degree of $R_j(X)$ is at most d. By Proposition 2.5 and Lemma 2.4,

$$\mathrm{mult}(R_j, \alpha_i) \geq \mathrm{mult}(P^{(0,j)}, \quad (\alpha_i, \beta_i)) \geq \mathrm{mult}(P, (\alpha_i, \beta_i)) - j.$$

Thus

$$\sum_{i \in I} \mathrm{mult}(R_j, \alpha_i) \geq |I| \cdot (m - j) \geq A \cdot (m - j) > d.$$

By Lemma 2.7, we conclude that $R_j(X) = 0$. Since this holds for every $j < m - d/A$, we conclude that $\mathrm{mult}(Q, f(X)) \geq m - d/A$.

We now complete the proof of the theorem. By the above discussion, for each $j \in [L]$, we know that $\mathrm{mult}(Q, f_j(X)) \geq m - \frac{d}{\gamma n}$. Thus, by Lemma 2.7 (applied to the nonzero polynomial $Q(Y) \in \mathbb{F}(X)[Y]$ and the set of evaluation points $S = \{f_j(X) : j \in [L]\}$),

$$\deg(Q) \geq \sum_{j \in [L]} \mathrm{mult}(Q, f(X)) \geq \left(m - \frac{d}{\gamma n} \right) \cdot L.$$

Since $\deg(Q) \leq \theta d/k$, we get

$$\theta d/k \geq \left(m - \frac{d}{\gamma n} \right) \cdot L.$$

Using $d = (1 + \epsilon) \cdot m \cdot \sqrt{\frac{nk}{\theta \cdot (2 - \theta)}}$ and $\theta = \frac{2}{1 + \frac{\gamma^2}{R}}$, we get

$$L \leq \frac{\theta}{k \cdot \frac{m}{d} - \frac{k}{\gamma n}}$$

$$= \frac{\theta}{\frac{1}{1 + \epsilon} \sqrt{\frac{k}{n} \cdot \theta \cdot (2 - \theta)} - \frac{k}{\gamma n}}$$

$$= \frac{1}{\frac{1}{1 + \epsilon} \sqrt{R \left(\frac{2}{\theta} - 1 \right)} - \frac{R}{\theta \gamma}} = \frac{1}{\frac{\gamma}{1 + \epsilon} - \left(\frac{\gamma}{2} + \frac{R}{2\gamma} \right)}.$$

Letting $\epsilon \to 0$, we get $L \leq \frac{2\gamma}{\gamma^2 - R}$, as desired.   □

REFERENCES

[1] Y. CASSUTO AND J. BRUCK, *A Combinatorial Bound on the List Size*, Technical report ETR058, Paradise Laboratory, California Institute of Technology, 2004.

[2] Z. DVIR, *On the size of Kakeya sets in finite fields*, J. Amer. Math. Soc., 22 (2009), pp. 1093–1097.

[3] Z. DVIR AND A. SHPILKA, *An improved analysis of linear mergers*, Comput. Complexity, 16 (2007), pp. 34–59.

[4] Z. DVIR AND A. WIGDERSON, *Kakeya sets, new mergers and old extractors*, in Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2008, pp. 625–633.

[5] V. GURUSWAMI AND A. RUDRA, *Explicit codes achieving list decoding capacity: Error-correction with optimal redundancy*, IEEE Trans. Inform. Theory, 54 (2008), pp. 135–150.

[6] V. GURUSWAMI AND M. SUDAN, *Improved decoding of Reed-Solomon and algebraic-geometric codes*, IEEE Trans. Inform. Theory, 45 (1999), pp. 1757–1767.

[7] V. GURUSWAMI, C. UMANS, AND S. P. VADHAN, *Unbalanced expanders and randomness extractors from Parvaresh–Vardy codes*, J. ACM, 56 (2009), 20.

[8] J. W. P. HIRSCHFELD, G. KORCHMAROS, AND F. TORRES, *Algebraic Curves over a Finite Field*, Princeton Ser. Appl. Math., Princeton University Press, Princeton, NJ, 2008.

[9] C. J. LU, O. REINGOLD, S. VADHAN, AND A. WIGDERSON, *Extractors: Optimal up to constant factors*, in Proceedings of the 35th Annual ACM Symposium on Theory of Computing (STOC), 2003, pp. 602–611.

[10] F. PARVARESH AND A. VARDY, *Correcting errors beyond the Guruswami-Sudan radius in polynomial time*, in Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2005, pp. 285–294.

[11] R. RAZ, O. REINGOLD, AND S. P. VADHAN, *Extracting all the randomness and reducing the error in Trevisan's extractors*, J. Comput. System Sci., 65 (2002), pp. 97–128.

[12] O. REINGOLD, R. SHALTIEL, AND A. WIGDERSON, *Extracting randomness via repeated condensing*, SIAM J. Comput., 35 (2006), pp. 1185–1209.

[13] O. REINGOLD, S. VADHAN, AND A. WIGDERSON, *Entropy waves, the zig-zag graph product, and new constant-degree expanders and extractors*, in Proceedings of the 41st IEEE Symposium on Foundations of Computer Science (FOCS), 2000, pp. 3–13.

[14] S. SARAF AND M. SUDAN, *Improved lower bound on the size of Kakeya sets over finite fields*, Anal. PDE, 1 (2008), pp. 375–379.

[15] R. SHALTIEL, *Recent developments in extractors*, Bull. Eur. Assoc. Theor. Comput. Sci. EATCS, 77 (2002), pp. 67–95.

[16] V. SHOUP, *New algorithms for finding irreducible polynomials over finite fields*, in Proceedings of the 29th Annual Symposium on Foundations of Computer Science (FOCS), 1988, pp. 283–290.

[17] M. SUDAN, *Decoding of Reed-Solomon codes beyond the error-correction bound*, J. Complexity, 13 (1997), pp. 180–193.

[18] A. TA-SHMA, *On extracting randomness from weak random sources*, in Proceedings of the 28th Annual ACM Symposium on Theory of Computing (STOC), 1996, pp. 276–285.

[19] A. TA-SHMA, *Refining Randomness*, Ph.D. thesis, The Hebrew University, Jerusalem, 1996.

[20] A. WIGDERSON AND D. ZUCKERMAN, *Expanders that beat the eigenvalue bound: Explicit construction and applications*, Combinatorica, 19 (1999), pp. 125–138.

[21] T. WOLFF, *Recent work connected with the Kakeya problem*, in Prospects in Mathematics, Princeton, NJ, 1999, pp. 129–162.

[22] D. ZUCKERMAN, *Linear degree extractors and the inapproximability of max clique and chromatic number*, Theory Comput., 3 (2007), pp. 103–128.