Algebraic Methods in Randomness and Pseudorandomness

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
OCT 05 2010
LIBRARIES

by

Swastik Kopparty

B.S., University of California (2004)
S.M., Massachusetts Institute of Technology (2007)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

ARCHIVES

Doctor of Philosophy in Computer Science and Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2010

© Swastik Kopparty, MMX. All rights reserved.

The author hereby grants to MIT permission to reproduce and distribute publicly paper and electronic copies of this thesis document in whole or in part.

Author .....
Department of Electrical Engineering and Computer Science
September 1, 2010

Certified by .....
Madhu Sudan
Professor
Thesis Supervisor

Accepted by .....
Terry P. Orlando
Chairman, Department Committee on Graduate Students

Algebraic Methods in Randomness and Pseudorandomness

by

Swastik Kopparty

Submitted to the Department of Electrical Engineering and Computer Science
on September 1, 2010, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Computer Science and Engineering

Abstract

Algebra and randomness come together rather nicely in computation. A central example of this relationship in action is the Schwartz-Zippel lemma and its application to the fast randomized checking of polynomial identities. In this thesis, we further this relationship in two ways: (1) by compiling new algebraic techniques that are of potential computational interest, and (2) demonstrating the relevance of these techniques by making progress on several questions in randomness and pseudorandomness.
The technical ingredients we introduce include:

- Multiplicity-enhanced versions of the Schwartz-Zippel lemma and the "polynomial method", extending their applicability to "higher-degree" polynomials.
- Conditions for polynomials to have an unusually small number of roots.
- Conditions for polynomials to have an unusually structured set of roots, e.g., containing a large linear space.

Our applications include:

- Explicit constructions of randomness extractors with logarithmic seed and vanishing "entropy loss".
- Limit laws for first-order logic augmented with the parity quantifier on random graphs (extending the classical 0-1 law).
- Explicit dispersers for affine sources of imperfect randomness with sublinear entropy.

Thesis Supervisor: Madhu Sudan
Title: Professor

Acknowledgments

I am deeply indebted to my advisor, Madhu Sudan, for all the invaluable guidance, wisdom and encouragement that he has given me all these years. Madhu's style of thinking, taste in problems and his views on research, teaching and life, have inspired me greatly and shaped me immeasurably.

Eli Ben-Sasson has been a great friend and mentor to me. I am extremely thankful to him for all the things he has taught me, the many years of encouragement, collaboration, and advice, and for showing me by example that eternal optimism can conquer many a difficult problem. I am very grateful to Alex Samorodnitsky for his valuable friendship and for being a constant source of encouragement. Alex's formidable mastery over $\{0,1\}^n$ has been big inspiration for me. Many thanks to Sergey Yekhanin for many valuable discussions and for fine companionship through the jungle of finite fields.

I was very fortunate to have Chinya Ravishankar as my undergraduate advisor at UCR. Ravi introduced me to, and imbued in me a cultured taste for, a wide variety of topics across computer science and mathematics; it has served me well all these years. I would like to thank Ilya Dumer for hooking me on to coding theory, sharing his vast experience with me, and encouragement ever since. Many thanks to Satish Tripathi, Srikanth Krishnamurthy and Michalis Faloutsos for getting me started on research in the first place and much valuable guidance.

I am grateful to Eli Ben-Sasson, Alex Samorodnitsky, Phokion G. Kolaitis, T. S. Jayram, Henry Cohn, Irit Dinur, Omer Reingold and Sergey Yekhanin for arranging memorable long-term visits and for very enjoyable collaborations.

Thanks to the good guys in the office for all the discussions, advice, opinions and gossip: Elena Grigorescu, Nick Harvey, Brendan Juba, Jelani Nelson, Ben Rossman, Shubhangi Saraf, Tasos Sidiropoulos and Sergey Yekhanin.

I have had many wonderful collaborators: Mukul Agarwal, Eli Ben-Sasson, Arnab Bhattacharyya, Kristian Brander, Henry Cohn, Irit Dinur, Zeev Dvir, Elena Grigorescu, Sandeep Gupta, Venkatesan Guruswami, Danny Gutfreund, Prahladh Harsha,

Avinatan Hassidim, Johan Håstad, T.S. Jayram, Tali Kaufman, Phokion G. Kolaitis, Brendan Juba, Sanjoy Mitter, Jelani Nelson, K.P.S. Bhaskara Rao, Jaikumar Radhakrishnan, Prasad Raghavendra, Chinya Ravishankar, Omer Reingold, Benjamin Rossman, Alex Samorodnitsky, Shubhangi Saraf, Grant Schoenebeck, Madhu Sudan, Sergey Yekhanin and David Zuckerman. I am very grateful to them for everything that they taught me.

This thesis is based on joint works with Eli Ben-Sasson, Zeev Dvir, Phokion G. Kolaitis, Shubhangi Saraf and Madhu Sudan.

My father taught me how to think mathematics ruthlessly.

Everything I did and do is because of Swara, Shubhangi, Mum and Pop.

For my dear sister Swara

Contents

Chapter 1

Introduction

1.1 Randomness in computing

The introduction of randomness to computation led to a revolution in the field of algorithm design. Algorithm designers assumed that their algorithms were given access to a stream of independent, unbiased, random bits, and they found that these algorithms could efficiently solve problems that seemed out of reach of their deterministic counterparts.

The ubiquity of randomness begs the philosophical question: is randomness necessary? Do we really need to use randomness to efficiently solve some problems, or can all efficient randomized computation be replaced by equally efficient deterministic algorithms? A central question of this type is the P vs. BPP question.

Pseudorandomness is the theory of coping with the fictitious nature of pure randomness; of reconciling the idealized resource of truly random bits, with ground realities such as the lack of sources of such randomness, and the need for guarantees. There are several aspects to this theory: *extracting* pure random bits from weak sources of randomness, such as those produced in nature or by physical devices; *derandomizing* randomized algorithms, or more generally, generically derandomizing entire randomized complexity classes through *pseudorandom generators*.

Another topic at the confluence of randomness and computation is *average-case complexity*. Here we consider the behavior of algorithms on randomly chosen inputs.

For a given computational problem, is there an algorithm that solves it on almost all inputs? Such questions are important from a practical viewpoint (because real-life instances are not chosen adversarially) and also from a theoretical viewpoint (a theoretical justification for all of cryptography awaits answers to these questions).

Yet another way that randomness interacts with computer science is via *the probabilistic method*. In many combinatorial and computational problems, the probabilistic method shows us that certain desirable structures exist; however, it gives us no clue as to how to deterministically construct such structures. Here too, some downright basic questions remain unanswered. For example, we know that a random subset of $\{0,1\}^n$ is a "good" error-correcting code. Yet we do not know how to efficiently and deterministically produce a single error-correcting code which is as good!

Traditionally, algebra has played a prominent role in many aspects of randomness in computation. In this thesis, we further this relationship in two ways: (1) by compiling new algebraic techniques that are of potential computational interest, and (2) demonstrating the relevance of these techniques by making progress on several questions in randomness and pseudorandomness.

We now give a few examples of the kinds of algebraic tools that we will bring to bear on problems of interest in randomness and pseudorandomness.

1.2 Some Algebraic Tools

One of the most fruitful aspects of the interaction between algebra and randomness in computation comes from the relationship between a polynomial and the set of its roots. Let us recall the most famous such example. It is a classical theorem that a nonzero n-variate polynomial of degree d over the field $\mathbb{F}_q$ (the finite field of q elements) can evaluate to zero on at most d/q-fraction of the points in $\mathbb{F}_q^n$. This leads to the fundamental Schwartz-Zippel randomized "identity testing" algorithm: to check if a black-box B which computes a polynomial of degree d over a field $\mathbb{F}_q$ is identically 0 (where $d \ll q$), simply pick a random point $x \in \mathbb{F}_q^n$, and check if $B(x)$

equals 0[1].

Following up on this theme, the algebraic tools that we describe below give finer information about the relationship between a polynomial and the set of its zeroes. Later in this chapter we give more detailed introductions to some of their applications in the theory of randomness and pseudorandomness.

1.2.1 Variation 1: Counting roots with multiplicities

Let P be an n-variate polynomial of degree d over the field $\mathbb{F}_q$, with $d > q$. A priori, there is nothing whatsoever that we can say about the set of roots of P in $\mathbb{F}_q^n$. However, we can say something if we slightly expand our definition of root; namely we consider points where P vanishes *with high multiplicity.*

Lemma A Let $P(X_1, \ldots, X_n) \in \mathbb{F}_q[X_1, \ldots, X_n]$ be a nonzero polynomial of degree d. Then

$$\Pr_{x \in \mathbb{F}_q^n}[P \text{ vanishes at } x \text{ with multiplicity at least } m] \leq \frac{d}{mq}.$$

In Chapter 5, we show how this multiplicity-enhanced version of the "Schwartz-Zippel lemma" can be applied to interesting situations in pseudorandomness and combinatorics. As an application, we use this lemma, combined with a multiplicity-enhanced version of the "polynomial method", to give the first explicit constructions of seeded randomness extractors which simultaneously have vanishingly small "entropy loss" and seed-length optimal upto constant factors. We also show how to use such multiplicity-enhanced arguments to derive near-optimal lower bounds on the size of certain extremal geometric configurations in finite fields called Kakeya sets.

The proof of Lemma A itself appears in Chapter 2, along with other useful tools for dealing with polynomials and multiplicities.

[1]Because of its application in the Schwartz-Zippel randomized identity test, the lemma bounding the number of zeroes of a polynomial is often called the "Schwartz-Zippel lemma".

1.2.2 Variation 2: Counting roots of certain polynomials

If we have a polynomial about which we know some more information than simply its degree (for example, we may know something about its coefficients), then we can sometimes deduce more about the number of, and the location of, its roots. The next lemma (which generalizes a lemma of Babai, Nisan and Szegedy for the case $p = 1/2$), demonstrates such a phenomenon.

> **Lemma B** Let $p \in (0,1)$. Let $P(X_1, \ldots, X_n) \in \mathbb{F}_2[X_1, \ldots, X_n]$ be a polynomial of the form $\sum_{i=0}^{m} X_{3i}X_{3i+1}X_{3i+2} + R(X)$, where $R(X)$ is a polynomial of degree at most 2. Pick $x \in \mathbb{F}_2^n$, where each coordinate of x independently equals 1 with probability p and 0 with probability $1 - p$. Then,
>
> $$\Pr[P(x) = 0] \leq 1/2 + 2^{-c_p m},$$
>
> where $c_p > 0$ depends only on p.

In contrast, an arbitrary multilinear polynomial of degree 3 over $\mathbb{F}_2$ could evaluate to 0 on as many as a 7/8-fraction of the points in $\mathbb{F}_2^n$.

In Chapter 3, we show how algebraic results of this kind can be used to give limit laws for the average-case behavior of certain families of algorithms (first-order logic equipped with the "parity quantifier") on random graphs, extending the classical 0-1 law for first-order logic on random graphs. This will also enable us to answer some basic and natural questions about the distribution of subgraph counts mod 2 in random graphs, such as "what is the probability that a random graph has a odd number of triangles?", "what is the probability that a random graph has an even number of 4-cycles?" and "what is the probability of both these events happening simultaneously?".

Algebraic lemmas such as Lemma B formally appear in Chapter 2, where they are proved using the Gowers norms and their generalizations.

1.2.3 Variation 3: Structure of the roots of certain polynomials

If we know more information about a polynomial than just its degree, then it may also be possible to deduce other structural properties about the set of its roots.

To state the next lemma, we first introduce an interesting polynomial. Consider $\mathbb{F}_2^n$ and identify it with the large finite field $\mathbb{F}_{2^n}$ via an arbitrary $\mathbb{F}_2$-linear isomorphism. Let $\mathrm{Tr} : \mathbb{F}_{2^n} \to \mathbb{F}_2$ be the trace map. Consider the function $f : \mathbb{F}_{2^n} \to \mathbb{F}_2$ given by $f(x) = \mathrm{Tr}(x^7)$. Via the identification of $\mathbb{F}_{2^n}$ with $\mathbb{F}_2^n$, this yields a function $f' : \mathbb{F}_2^n \to \mathbb{F}_2$. Now it turns out that this function f' is simply the evaluation of a certain degree-3 polynomial. Call this polynomial $P_0(X_1, \ldots, X_n)$. We can now state the lemma.

> **Lemma C** Let $n > 0$ be an odd integer. Let $P(X_1, \ldots, X_n) \in \mathbb{F}_2[X_1, \ldots, X_n]$ be a polynomial of the form $P_0(X_1, \ldots, X_n) + R(X_1, \ldots, X_n)$, where P_0 is the polynomial described above, and R is any polynomial of degree at most 2. Then for every affine subspace $A \subseteq \mathbb{F}_2^n$ of dimension at least $2n/5 + O(1)$, there exists $x \in A$ such that $P(X) \neq 0$.

In Chapter 4, we prove lemmas of the above kind, and use them to give explicit constructions of randomness dispersers (a weak form of randomness extractor) from affine sources. The main tool that we use to prove these lemmas is a certain kind of polynomial known as a subspace polynomial. An introduction to the theory of subspace polynomials, as well as methods for translating between the world of multivariate polynomials over $\mathbb{F}_2^n$ and univariate polynomials over $\mathbb{F}_{2^n}$, are given in Chapter 2.

1.3 Main results and the role of algebra

We now give a slightly more detailed introduction to the problems considered in a thesis, as well as a glimpse to the role that the algebraic tools mentioned above play in their solution.

1.3.1 The parity quantifier on random graphs

The classical *0-1 law* for random graphs deals with a striking phenomenon at the intersection of logic, finite model theory and random graph theory. It describes a very sharp characterization of the average case behavior of a certain simple family of algorithms, formulas of first-order logic, on random graphs. A first-order formula on graphs is simply a grammatically correct formula using (i) $\forall$, the for-all quantifier, (ii) $\exists$, the there-exists quantifier, (iii) the adjacency relation $E(v, w)$, (iv) the equality relation "v = w", and (v) Boolean operations. A first-order formula determines a graph property: a graph G has the property given by formula φ if and only if φ is true when interpreted on G (the quantifiers quantify over vertices of G). For example, the first-order formula

$$\forall v \; \exists w \; E(v, w),$$

defines the graph property "there are no isolated vertices".

The 0-1 law states that for every first-order property φ in the theory of graphs and every $p \in (0, 1)$, as n approaches infinity, the probability that the random graph $G(n, p)$ satisfies φ approaches either 0 or 1! Furthermore, this limiting probability can be computed given φ.

Since its discovery, 0-1 laws have been discovered for a diverse collection of logics which can express more graph properties than first order logic. The frequently encountered nemesis to all generalizations is the PARITY barrier: any logic that can express the property "there are an odd number of vertices" cannot obey a 0-1 law.

In joint work with Phokion Kolaitis, we study a natural logic equipped with counting (which hence faces the PARITY barrier), and search for phenomena analogous to the 0-1 law for this logic. Specifically, we study $\mathsf{FO}[\oplus]$, first order logic augmented with the parity quantifier. The parity quantifier $\oplus$ is a quantifier which counts mod 2; $\oplus y \varphi(y)$ is true if there are an odd number of y such that $\varphi(y)$ is true. It is well known that $\mathsf{FO}[\oplus]$ fails to have a 0-1 law: for some properties the limiting probability may not exist, while for others the limit may exist, but need not equal 0 or 1. Eluding these two hurdles, we establish the following "modular convergence law":

> For every $\mathsf{FO}[\oplus]$ sentence φ, there are two explicitly computable rational numbers a_0, a_1, such that for $i \in \{0,1\}$, as n approaches infinity, the probability that the random graph $G(2n+i,p)$ satisfies φ approaches a_i.

Our results also extend appropriately to FO equipped with Mod_q quantifiers for prime q.

At the heart of our approach is an algebraic explanation of $\mathsf{FO}[\oplus]$ properties. We show that for every $\mathsf{FO}[\oplus]$ property φ, and for every n, there is a polynomial $Q(X_1,\ldots,X_{\binom{n}{2}}) \in \mathbb{F}_2[X_1,\ldots,X_{\binom{n}{2}}]$ whose degree depends only on φ, such that for most graphs G (under the $G(n,p)$ measure), Q evaluates to 1 on the adjacency matrix of G if and only G has the property φ. This "algebraic explanation" implies that the probability that $G(n,p)$ satisfies φ is essentially the probability that the polynomial Q is nonzero on a random input. The kinds of polynomials Q that show up here turn out to be very structured, and lemmas such as Lemma B above play a key role in understanding their zeroes. Curiously, such lemmas also turn out to be instrumental even in the proof that $\mathsf{FO}[\oplus]$ properties possess algebraic explanations.

Details appear in Chapter 3.

1.3.2 Randomness extraction and dispersion from affine sources

Randomness extraction is the process of obtaining random bits from sources of imperfect randomness. Randomness extraction has typically been studied in two kinds of settings: *deterministically* extracting randomness from structured (but unknown) sources of randomness, and extracting randomenss from general unstructured sources of randomness *using a few bits of pure random seed.*

A determinstic randomness extractor for a family $\mathcal{F}$ of subsets of $\{0,1\}^n$ is an efficiently computable function $E : \{0,1\}^n \to \{0,1\}^m$, which for every subset $X \in \mathcal{F}$ (the "source"), the distribution of $E(x)$ when x is picked uniformly from X is nearly-uniformly distributed. The relevance of such an object comes from the following observation: if we are given a random sample from a set X, and all we know about X is that it is from $\mathcal{F}$ ("X is a structured source"), then by applying E to that sample we

obtain nearly-uniform random bits. For various particular families $\mathcal{F}$ of "structured" subsets, a question of interest has been to explicitly construct randomness extractors for $\mathcal{F}$. The focus of the result described next is the case where $\mathcal{F}$ is the collection of $\mathbb{F}_2$-affine subspaces of a certain dimension.

In purely combinatorial terms, this has a very simple description. We seek a polynomial-time computable function $E : \mathbb{F}_2^n \to \{0,1\}^m$ such that for every affine space $A \subseteq \mathbb{F}_2^n$ of dimension at least k, if x is picked uniformly at random from A, then the distribution of $E(x)$ is close to the uniformly distributed over $\{0,1\}^m$. The parameter k measures the amount of entropy needed in the affine random source for the extractor to produce uniform random bits. The probabilistic method guarantees that there exist such functions E (but not necessarily polynomial-time computable) with k as small as $O(\log n)$.

Randomness dispersion is a weakened form of randomness extraction, where one asks only for the support of the random variable $E(x)$ to equal to $\{0,1\}^m$. For affine sources, a 1-bit-output disperser turns out to be exactly equivalent to the following neat Ramsey-like object: a 2-coloring of $\mathbb{F}_2^n$ such that no k-dimensional affine subspace is monochromatic. Again, the probabilistic method guarantees that such colorings exist for k even as small as $O(\log n)$; the problem of interest is to construct these colorings explicitly.

The explicit construction of randomness extractors and dispersers from affine sources has recently received much attention. It turns out that any function $f : \mathbb{F}_2^n \to \{0,1\}$ with small *Fourier coefficients* is a 1-bit-output affine extractor; this leads to affine extractors for affine spaces of dimension $> n/2$. The first breakthrough came in the work of Barak, Kindler, Shaltiel, Sudakov and Wigderson [BKS+05], who gave explicit constructions of affine dispersers from dimension δn for arbitrary $\delta > 0$. Subsequently Bourgain [Bou07] gave explicit constructions of affine extractors from the same dimension. These papers relied on the sum-product theorem for finite fields and other recent results from additive combinatorics.

In joint work with Eli Ben-Sasson, we developed an alternate, algebraic approach to constructing and analyzing affine dispersion phenomena. Via this approach, we

give an efficient deterministic construction of affine dispersers for *sublinear* dimension $k = \Omega(n^{1-\epsilon})$ for some positive $\epsilon > 0$ (one can take $\epsilon = 1/5$).

Our constructions revolve around viewing $\mathbb{F}_2^n$ as $\mathbb{F}_{2^n}$ (as in Lemma C). The method of proof makes use of certain simple-but-powerful objects known as *subspace polynomials*. Subspace polynomials are the enigmatic nexus between the multivariate linear geometry of $\mathbb{F}_2^n$ and the univariate algebra of $\mathbb{F}_{2^n}$. Via subspace polynomials, establishing that certain functions are affine dispersers reduces to understanding the coefficients of certain univariate polynomials. We then achieve such an understanding, and in the course of our proofs, we develop some basic structural results about the zero/nonzero pattern of the coefficients of subspace polynomials.

Details appear in Chapter 4.

1.3.3 Randomness extraction from general sources with negligible entropy loss

The third topic of this thesis addresses *seeded* randomness extraction from general sources of weak randomness.

A *seeded randomness extractor for sources of entropy* k is a function $E : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ such that for every set $X \subseteq \{0,1\}^n$ (the "weak source") with $|X| \geq 2^k$, the distribution of $E(x,u)$ is nearly-uniformly distributed, where x is picked uniformly at random from X and u (the "seed") is picked uniformly at random from $\{0,1\}^d$. The probabilistic method shows that there exist seeded randomness extractors for sources of entropy k with $d = O(\log n)$, while m (the amount of randomness extracted) is almost as large as $k + d$ (the amount of randomness fed into E). The quantity $m/(k+d)$ is referred to as the fraction of entropy extracted.

In joint work with Zeev Dvir, Shubhangi Saraf and Madhu Sudan, we show how to construct randomness extractors that use seeds of length $O(\log n)$ while extracting $1 - o(1)$ fraction of the min-entropy of the source. Previous results could extract only a constant α-fraction (with $\alpha < 1$) of the entropy while maintaining logarithmic seed length.

The crux of our improvement is an algebraic technique which we call the extended method of multiplicities. The "method of multiplicities", as used in prior work, analyzed subsets of vector spaces over finite fields by constructing somewhat low degree interpolating polynomials that vanish on every point in the subset *with high multiplicity*. The typical use of this method involved showing that the interpolating polynomial also vanished on some points outside the subset, and then used simple bounds on the number of zeroes to complete the analysis. Our augmentation to this technique is that we prove, under appropriate conditions, that the interpolating polynomial vanishes *with high multiplicity* outside the set. We then invoke Lemma A, which gives a bound on the number of *high multiplicity* zeroes, to complete the analysis. This novelty leads to significantly tighter analyses.

We use the extended method of multiplicities in the analysis of our improved randomness extractors as follows. For a certain candidate extractor function E (whose definition involves certain geometric objects over finite fields), we suppose that E is not a randomness extractor, and from this "non-extractorness" deduce that certain extremal configurations in vector spaces over finite fields exist. We then rule out the existence of such an extremal configuration using the extended method of multiplicities. Using this method, we also get near-optimal lower bounds on the size of Kakeya sets over finite fields, a topic of much interest in recent years.

Details appear in Chapter 5.

1.3.4 Explicit functions with small correlation with low-degree polynomials

The final topic addressed in this thesis deals with average-case complexity. One of the challenges in computational complexity is to find explicit functions that are hard to compute on average for a "simple" complexity class. A standard measure of the average-case computability of one Boolean function f by another one g is their

correlation defined by

$$\mathrm{Corr}_{\mathbb{F}_2^n}(f,g) = |\mathbb{E}_{x\in\mathbb{F}_2^n}[(-1)^{f(x)+g(x)}]|.$$

Let us informally say that a function f is "exponentially-hard" for a complexity class if it has exponentially small correlation with all functions computed by the class. Given the important role that hard functions play in the study of computational complexity, coming up with explicit constructions of hard functions for natural complexity classes is a well-motivated problem.

One such complexity class for which we would like to find explicit hard functions is the class of functions that can be computed by low-degree polynomials. In addition to being interesting in its own right, this problem is related to important questions in complexity theory because of the result of [Raz87a], who showed that constructing a function that cannot be approximated well by polynomials of degree as high as $\mathrm{poly}\log n$ implies strong average-case lower bounds for the class of bounded-depth circuits with parity gates. Today this problem is wide open.

For the class of polynomials of degree $\ll \log n$, there are two kinds of constructions of exponentially-hard functions known. The first is derived from the multiparty communication lower bounds of [BNS89] and the second is the recent construction of [VW07] that is derived from a XOR-lemma for low-degree polynomials.

In joint work with Eli Ben-Sasson, we find a rich family of explicit functions that are exponentially uncorrelated with polynomials of degree $\ll \log n$, matching results of [BNS89, VW07]. A typical function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ in this family is given as follows: identify $\mathbb{F}_2^n$ with $\mathbb{F}_{2^n}$ via an arbitrary $\mathbb{F}_2$-linear isomorphism (as in Lemma C), and for a certain polynomial $Q(X) \in \mathbb{F}_{2^n}[X]$ of small degree, we set $f(x) = \mathrm{Tr}(Q(x))$.

In coding theory terminology, this gives a kind of threshold phenomenon between some classical algebraic codes. If C_1 is a dual-BCH code and C_2 is a Reed-Muller code (with suitable parameters), both of block-length 2^n, any codeword $c \in C_1$ is either (1) a codeword of C_2, or (2) $1/2 - 2^{-\Omega(n)}$ far from all the codewords of C_2. This result may be viewed as a generalization of the Weil bound for character sums,

which yields this dichotomy in the case where C_2 is the Reed-Muller code of degree 1 polynomials.

Details appear in Chapter 6.

1.3.5 Organization of this Thesis

In Chapter 2, we introduce some basic tools and results on polynomial and their zeroes, upon which the remaining chapters will build. In Chapter 3, we study $\mathsf{FO}[\oplus]$ on random graphs. In Chapter 4, we give explicit constructions of randomness dispersers for affine sources, which we analyze using subspace polynomials. In Chapter 5, we introduce the extended method of multiplicities, and use it to construct randomness extractors with negligible entropy loss and near-tight lower bounds on the size of Kakeya sets over finite fields. In Chapter 6, we describe explicit functions which have exponentially small correlation with low-degree polynomials. Appendix A contains a short exposition of the elementary proof of the Weil bound, which gets used in Chapter 6.

Chapter 2

Polynomials and their zeroes

We now introduce some of the basic tools and results related to polynomials and their zeroes, upon which the remaining chapters will build.

2.1 Derivatives and Multiplicities

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

2.1.1 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 non-negative integers, its *weight*, denoted $\mathrm{wt}(\mathbf{i})$, equals $\sum_{j=1}^{n} i_j$.

Let $\mathbb{F}$ be any field. 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.

For a vector of non-negative integers $\mathbf{i} = \langle i_1, \ldots, i_n \rangle$, let $\mathbf{X}^{\mathbf{i}}$ 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 $\mathrm{wt}(\mathbf{i})$. For

n-tuples of non-negative integers $\mathbf{i}$ and $\mathbf{j}$, we use the notation

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

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

Definition 2.1.1 ((Hasse) Derivative) *For $P(\mathbf{X}) \in \mathbb{F}[\mathbf{X}]$ and non-negative vector $\mathbf{i}$, the $\mathbf{i}$th* (Hasse) derivative *of P, denoted $P^{(\mathbf{i})}(\mathbf{X})$, is the coefficient of $\mathbf{Z}^{\mathbf{i}}$ in the polynomial $\tilde{P}(\mathbf{X},\mathbf{Z}) \stackrel{\text{def}}{=} P(\mathbf{X}+\mathbf{Z}) \in \mathbb{F}[\mathbf{X},\mathbf{Z}]$.*

Thus,

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

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

Definition 2.1.2 (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* $\mathrm{mult}(P,\mathbf{a})$*, is the largest integer M such that for every non-negative vector $\mathbf{i}$ with $\mathrm{wt}(\mathbf{i}) < M$, we have $P^{(\mathbf{i})}(\mathbf{a}) = 0$ (if M may be taken arbitrarily large, we set $\mathrm{mult}(P,\mathbf{a}) = \infty$).*

Note that $\mathrm{mult}(P,\mathbf{a}) \geq 0$ for every $\mathbf{a}$. Also, $P(\mathbf{a}) = 0$ if and only if $\mathrm{mult}(P,\mathbf{a}) \geq 1$.

The above notations and definitions also extend naturally to a tuple $P(\mathbf{X}) = \langle P_1(\mathbf{X}),\ldots,P_m(\mathbf{X})\rangle$ 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 $\mathrm{mult}(P,\mathbf{a}) = \min_{j\in[m]}\{\mathrm{mult}(P_j,\mathbf{a})\}$.

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, [HKT08, pages 144-155]) and seems to have first appeared in the CS literature (without being explicitly referred to by this name) in the work of Guruswami and Sudan [GS99]. 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.

2.1.2 Properties of Hasse Derivatives

The following proposition lists basic properties of the Hasse derivatives. Parts (1)-(3) below are the same as for the analytic 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 crucial for our purposes is that it shows 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.

Proposition 2.1.3 (Basic Properties of Derivatives) *Let $P(\mathbf{X}), Q(\mathbf{X}) \in \mathbb{F}[\mathbf{X}]^m$ 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

Items 1 and 2 are easy to check, and item 3 follows immediately from them. For item 4, we expand $P(\mathbf{X}+\mathbf{Z}+\mathbf{W})$ in two ways. First expand

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

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. ■

2.1.3 Properties of Multiplicities

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

Lemma 2.1.4 (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 2.1.3, $(P^{(\mathbf{i})})^{(\mathbf{j})}(\mathbf{a}) = \binom{\mathbf{i}+\mathbf{j}}{\mathbf{i}} P^{(\mathbf{i}+\mathbf{j})}(\mathbf{a})$. 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 $P(\mathbf{X}) = (P_1(\mathbf{X}), \ldots, P_m(\mathbf{X})) \in \mathbb{F}[\mathbf{X}]^m$ and $Q(\mathbf{Y}) = (Q_1(Y), \ldots, Q_n(Y)) \in \mathbb{F}[\mathbf{Y}]^n$. We define the composition polynomial $P \circ Q(\mathbf{Y}) \in \mathbb{F}[\mathbf{Y}]^m$ to be the polynomial $P(Q_1(\mathbf{Y}), \ldots, Q_n(\mathbf{Y}))$. In this situation we have the following proposition.

Proposition 2.1.5 *Let* $P(\mathbf{X}), Q(\mathbf{Y})$ *be as above. Then for any* $\mathbf{a} \in \mathbb{F}^\ell$,

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

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

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

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

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

Applying the above to $P(\mathbf{X})$ and $Q(T) = \mathbf{a} + T\mathbf{b} \in \mathbb{F}[T]^n$, we get the following corollary.

Corollary 2.1.6 *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 $t \in \mathbb{F}$,*

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

2.1.4 Counting roots with multiplicities

We are now ready to state and prove a bound on the number of high multiplicity zeroes that a polynomial can have, strengthening 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 non-zero degree d polynomial and $S \subseteq \mathbb{F}$ is a finite set.

Lemma 2.1.7 (The Schwartz-Zippel Lemma) *Let $P \in \mathbb{F}_q[\mathbf{X}]$ be a polynomial of total degree at most d, and let $S \subseteq \mathbb{F}_q$. If $\Pr_{\mathbf{a} \in S^n}[P(\mathbf{a}) = 0] > \frac{d}{|S|}$, then $P(\mathbf{X}) = 0$.*

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 $\sum_{\mathbf{a} \in S^n} \min\{1, \mathrm{mult}(P, \mathbf{a})\} \leq d \cdot |S|^{n-1}$. The multiplicity-enhanced version below strengthens this lemma by replacing $\min\{1, \mathrm{mult}(P, \mathbf{a})\}$ with $\mathrm{mult}(P, \mathbf{a})$ in this inequality.

Lemma 2.1.8 *Let $P \in \mathbb{F}[\mathbf{X}]$ be a 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 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 $\sum_{a \in S} \mathrm{mult}(P, a)$ is at most the degree of P.

Now suppose $n > 1$. 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 $a_1, \ldots, a_{n-1} \in S$, let $m_{a_1, \ldots, a_{n-1}} = \mathrm{mult}(P_t, (a_1, \ldots, a_{n-1}))$. We will show that

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

Given this, we may then bound

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

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 Equation (2.2). Fix $a_1,\ldots,a_{n-1}\in S$ and let $\mathbf{i}=(i_1,\ldots,i_{n-1})$ be such that $\mathrm{wt}(\mathbf{i}) = m_{a_1,\ldots,a_{n-1}}$ 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 2.1.4 and Corollary 2.1.6, we know that

$$\begin{aligned}\mathrm{mult}(P(X_1,\ldots,X_n),(a_1,\ldots,a_n)) &\leq \mathrm{wt}(\mathbf{i},0)+\mathrm{mult}(P^{(\mathbf{i},0)}(X_1,\ldots,X_n),(a_1,\ldots,a_n))\\ &\leq m_{a_1,\ldots,a_{n-1}}+\mathrm{mult}(P^{(\mathbf{i},0)}(a_1,\ldots,a_{n-1},X_n),a_n).\end{aligned}$$

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 Equation (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 2.1.9 *Let $P\in\mathbb{F}_q[\mathbf{X}]$ be a polynomial of total degree at most d. If $\sum_{\mathbf{a}\in\mathbb{F}_q^n}\mathrm{mult}(P,\mathbf{a})>d\cdot q^{n-1}$, then $P(\mathbf{X})=0$.*

2.2 The bias of polynomials and the μ-Gowers norm

In this section, we work with polynomials over "small" fields. We will study the probability that a polynomial can evaluate to zero at a random point, where the

point may be picked from a non-uniform distribution. We then show that if the polynomial is of a certain form, then one can get significantly better bounds on this probability. The centerpiece of this result is a measure of pseudorandomness of a function that we call the "μ-Gowers norm".

For contrast, we begin by stating and proving a basic bound on the zeroes of arbitrary nonzero multilinear polynomials. The case $p = 1/2$ of this lemma is the standard bound on the number of zeroes of a multilinear polynomial.

Lemma 2.2.1 (Basic bound on zeroes of multilinear polynomials) *Let $P(X_1, \ldots, X_n) \in \mathbb{F}_2[X_1, \ldots, X_n]$ be a multilinear polynomial of degree at most d. Let $p \in [0, 1]$. Pick $x \in \mathbb{F}_2^n$ where each coordinate of x is picked independently, and $\Pr[x_i = 1] = p$. Then*

$$\Pr_x[P(x) \neq 0] \geq \min\{p^d, (1-p)^d\}.$$

Proof The proof is by induction on n and d.

Let $P(X_1, \ldots, X_n) = P'(X_1, \ldots, X_{n-1}) \cdot X_n + P''(X_1, \ldots, X_{n-1})$, where P' is of degree at most $d-1$ and P'' is of degree at most d.

For $i \in \{0, 1\}$, let $P_i(X_1, \ldots, X_{n-1})$ be the polynomial $P(X_1, \ldots, X_{n-1}, i)$. Observe that $P_0 + P' = P_1$.

- **Case 1:** $P' = 0$. In this case, $P(X_1, \ldots, X_n) = P''(X_1, \ldots, X_{n-1})$, and in this case, $\Pr_x[P(x) \neq 0] = \Pr_x[P''(x) \neq 0] \geq \min\{p^d, (1-p)^d\}$ (by induction hypothesis).

- **Case 2:** $P' \neq 0$ and $\deg(P'') \leq d - 1$. In this case, both the polynomials $P(X_1, \ldots, X_{n-1}, 0)$ and $P(X_1, \ldots, X_{n-1}, 1)$ are of degree at most $d-1$. Since they differ by $P' \neq 0$, at most one of them can be identically 0. Thus $\Pr_x[P(x) \neq 0] \geq \min\{p, (1-p)\} \cdot \min\{p^{d-1}, (1-p)^{d-1}\} = \min\{p^d, (1-p)^d\}$, as desired.

- **Case 3:** In this case, P'' has degree exactly d, and thus both P_0 and P_1 are nonzero polynomials of degree at most d. Thus $\Pr_x[P(x) \neq 0] \geq \min\{\Pr_x[P(x) = 0 \mid x_n = 0], \Pr_x[P(x) = 0 \mid x_n = 1]\} \geq \min\{p^d, (1-p)^d\}$.

■

An interesting corollary of this lemma is that low-degree polynomials over $\mathbb{F}_2$ cannot vanish on a large fraction of Hamming balls of radius $\Omega(n)$. This does not follow from just a bound on the number of zeroes of such polynomials.

Below we denote by $B(x, \delta)$ the Hamming ball of radius δn centered at x.

Corollary 2.2.2 *Let $P(X_1, \ldots, X_n) \in \mathbb{F}_2[X_1, \ldots, X_n]$ be a nonzero multilinear polynomial of degree d. Then for every $0 < p < \delta < 1/2$ and every $x \in \mathbb{F}_2^n$,*

$$\Pr_{y \in B(x,\delta)}[P(y) \neq 0] \geq p^d - o_n(1).$$

The aim of the rest of this section is to prove a substantially strengthened bound on the probability that a nonzero polynomial of degree d evaluates to 0 at a random point, when the highest degree monomials of the polynomial take a certain special form. We state this lemma below. This lemma is a strengthening of a lemma of Babai, Nisan, Szegedy (which deals with the case $p = 1/2, q = 2$) and of Grolmusz (which deals with the case $p = 1/2$ and general q).

Lemma 2.2.3 *Let $q > 1$ be an integer and let $p \in (0, 1)$. Let $E_1, \ldots, E_r$ be pairwise disjoint subsets of $[m]$ each of cardinality d. Let $Q(Z_1, \ldots, Z_m) \in \mathbb{Z}_q[Z_1, \ldots, Z_m]$ be a polynomial of the form*

$$\left(\sum_{j=1}^{r} a_j \prod_{i \in E_j} Z_i \right) + R(\mathbf{Z}),$$

where each $a_j \neq 0$ and $\deg(R(\mathbf{Z})) < d$. Let $\mathbf{z} = (z_1, \ldots, z_m) \in \mathbb{Z}_q^m$ be the random variable where, independently for each i, we have $\Pr[z_i = 1] = p$ and $\Pr[z_i = 0] = 1 - p$. Then,

$$\left| \mathbb{E}\left[\omega^{Q(\mathbf{z})}\right] \right| \leq 2^{-\Omega_{q,p,d}(r)}.$$

In particular, if q is prime, then

$$\Pr[Q(\mathbf{z}) \neq 0] \geq 1 - 1/q - 2^{-\Omega_{q,p,d}(r)}.$$

2.2.1 The μ-Gowers norm

The proof of Lemma 2.2.3 will use a variant of the Gowers norms. Let $Q : \mathbb{Z}_q^m \to \mathbb{Z}_q$ be any function, and define $f : \mathbb{Z}_q^m \to \mathbb{C}$ by $f(x) = \omega^{Q(x)}$. The Gowers norm of f is an analytic quantity that measures how well Q correlates with degree d polynomials: the correlation of Q with polynomials of degree $d-1$ under the uniform distribution is bounded from above by the d^{th}-Gowers norm of f. Thus to show that a certain Q is uncorrelated with all degree $d-1$ polynomials under the uniform distribution, it suffices to bound the d^{th}-Gowers norm of f. In Lemma 2.2.3, we wish to show that a certain Q is uncorrelated with all degree $d-1$ polynomials under a distribution μ that need not be uniform. To this end, we define a variant of the Gowers norm, which we call the μ-Gowers norm, and show that if the $(d,\mu)^{\text{th}}$-Gowers norm of f is small, then Q is uncorrelated with all degree $d-1$ polynomials *under* μ. We then complete the proof of Lemma 2.2.3 by bounding the $(d,\mu)^{\text{th}}$-Gowers norm of the relevant f.

We first define the μ-Gowers norm and develop some of its basic properties.

Let H be an abelian group and let μ be a probability distribution on H. For each $d \geq 0$, define a probability distribution $\mu^{(d)}$ on H^{d+1} inductively by $\mu^{(0)} = \mu$, and, for $d \geq 1$, let $\mu^{(d)}(x, t_1, \ldots, t_d)$ equal

$$\frac{\mu^{(d-1)}(x, t_1, \ldots, t_{d-1}) \cdot \mu^{(d-1)}(x + t_d, t_1, \ldots, t_{d-1})}{\sum_{z \in H} \mu^{(d-1)}(z, t_1, \ldots, t_{d-1})}.$$

Equivalently, to sample $(x, t_1, \ldots, t_d)$ from $\mu^{(d)}$, first take a sample $(x, t_1, \ldots, t_{d-1})$ from $\mu^{(d-1)}$, then take a sample $(y, t_1', \ldots, t_{d-1}')$ from $\mu^{(d-1)}$ conditioned on $t_i' = t_i$ for each $i \in [d-1]$, and finally set $t_d = y - x$ (our sample is then $(x, t_1, \ldots, t_{d-1}, t_d)$). Notice that the distribution of a sample $(x, t_1, \ldots, t_d)$ from $\mu^{(d)}$ is such that for each $S \subseteq [d]$, the distribution of the point $x + \sum_{i \in S} t_i$ is precisely μ.

For a function $f : H \to \mathbb{C}$ and $\mathbf{t} \in H^d$, we define its d^{th}-*discrete-derivative in directions* $\mathbf{t}$ to be the function $D_{\mathbf{t}} f : H \to \mathbb{C}$ given by

$$D_{\mathbf{t}} f(x) = \prod_{S \subseteq [d]} f(x + \sum_{i \in S} t_i)^{\circ S},$$

where $a^{\circ S}$ equals the complex conjugate $\bar{a}$ if $|S|$ is odd, and $a^{\circ S}$ equals a otherwise. From the definition it immediately follows that $D_{(\mathbf{t},u)}f(x) = D_{\mathbf{t}}f(x)\overline{D_{\mathbf{t}}f(x+u)}$ (where $(\mathbf{t},u)$ denotes the vector $(t_1,\ldots,t_d,u) \in H^{d+1}$).

We now define the μ-Gowers norm.

Definition 2.2.4 (μ-Gowers Norm) *If μ is a distribution on H, and $f : H \to \mathbb{C}$, we define its (d,μ)-Gowers norm by*

$$\|f\|_{U^d,\mu} = \left|\mathbb{E}_{(x,\mathbf{t})\sim\mu^{(d)}}\left[(D_{\mathbf{t}}f)(x)\right]\right|^{\frac{1}{2^d}}.$$

When μ is the uniform distribution over H, we recover the usual Gowers norm, denoted by $\|f\|_{U^d}$.

When H is of the form $\mathbb{Z}_q^m$, then the (d,μ)-Gowers norm of a function is supposed to estimate the correlation, under μ, of that function with polynomials of degree $d-1$. Intuitively, this happens because the Gowers norm of f measures how often the d^{th} discrete derivative of f vanishes.

The next few lemmas enumerate some of the useful properties that μ-Gowers norms enjoy.

Lemma 2.2.5 *Let $f : H \to \mathbb{C}$. Then,*

$$|\mathbb{E}_{x\sim\mu}[f(x)]| \leq \|f\|_{U^d,\mu}.$$

Proof We prove that for every d, $\|f\|_{U^d,\mu} \leq \|f\|_{U^{d+1},\mu}$. The lemma follows by noting that $\|f\|_{U^0,\mu} = |\mathbb{E}_{x\sim\mu}[f(x)]|$.

The proof proceeds (following Gowers [Gow01] and Green-Tao [GT08]) via the

Cauchy-Schwarz inequality,

$$\begin{aligned}
\|f\|_{U^d,\mu}^{2^{d+1}} &= \left|\mathbb{E}_{(x,\mathbf{t})\sim\mu^{(d)}}\left[D_{\mathbf{t}}f(x)\right]\right|^2 \\
&\leq \mathbb{E}_{\mathbf{t}}\left[\left|\mathbb{E}_x\left[D_{\mathbf{t}}f(x)\right]\right|^2\right] && \text{by Cauchy-Schwarz} \\
&= \mathbb{E}_{\mathbf{t}}\mathbb{E}_{x,y}\left[D_{\mathbf{t}}f(x)\overline{D_{\mathbf{t}}f(y)}\right] && \text{where } y \text{ is an independent sample of } x \text{ given } \mathbf{t}\,. \\
&= \mathbb{E}_{x,\mathbf{t},u}\left[D_{\mathbf{t}}f(x)\overline{D_{\mathbf{t}}f(x+u)}\right] && \text{where } u = y - x \\
&= \mathbb{E}_{(x,\mathbf{t},u)\sim\mu^{(d+1)}}\left[D_{\mathbf{t}}f(x)\overline{D_{\mathbf{t}}f(x+u)}\right] && \text{by definition of } \mu^{(d+1)} \\
&= \mathbb{E}_{(x,\mathbf{t},u)\sim\mu^{(d+1)}}\left[D_{(\mathbf{t},u)}f(x)\right] \\
&= \|f\|_{U^{d+1},\mu}^{2^{d+1}}.
\end{aligned}$$

This proves the lemma. ■

Definition 2.2.6 *For each $i \in [r]$, let $g_i : H \to \mathbb{C}$. We define $(\bigotimes_{i=1}^r g_i) : H^r \to \mathbb{C}$ by*

$$\left(\bigotimes_{i=1}^r g_i\right)(x_1,\ldots,x_r) = \prod_{i=1}^r g_i(x_i).$$

For each $i \in [r]$, let μ_i be a probability measure on H. We define the probability measure $\bigotimes_{i=1}^r \mu_i$ on H^r by

$$\left(\bigotimes_{i=1}^r \mu_i\right)(x_1,\ldots,x_r) = \prod_{i=1}^r \mu_i(x_i).$$

Lemma 2.2.7 $\|\bigotimes_{i=1}^r g_i\|_{U^d,\bigotimes_{i=1}^r \mu_i} = \prod_{i=1}^r \|g_i\|_{U^d,\mu_i}$.

Proof Follows by expanding both sides and using the fact that $\left(\bigotimes_{i=1}^r \mu_i\right)^{(d)} = \bigotimes_{i=1}^r \left(\mu_i^{(d)}\right)$.

■

Lemma 2.2.8 *Let $q > 1$ be an integer and let $\omega \in \mathbb{C}$ be a primitive q^{th}-root of unity. For all $f : \mathbb{Z}_q^n \to \mathbb{C}$, all probability measures μ on $\mathbb{Z}_q^n$, and all polynomials*

$h \in \mathbb{Z}_q[Y_1, \ldots, Y_n]$ *of degree* $< d$,

$$\|f\omega^h\|_{U^d,\mu} = \|f\|_{U^d,\mu}.$$

The above lemma follows from the fact that $(D_{\mathbf{t}}f) = (D_{\mathbf{t}}(f \cdot \omega^h))$.

Lemma 2.2.9 *Let* $a \in \mathbb{Z}_q \setminus \{0\}$ *and let* $g : \mathbb{Z}_q^d \to \mathbb{C}$ *be given by* $g(y) = \omega^{a\prod_{i=1}^d y_i}$. *Let* μ *be a probability distribution on* $\mathbb{Z}_q^d$ *with* $\mathrm{supp}(\mu) \supseteq \{0,1\}^d$. *Then* $\|g\|_{U^d,\mu} < 1 - \epsilon$, *where* $\epsilon > 0$ *depends only on* q, d *and* μ.

Proof As $\{0,1\} \subseteq \mathrm{supp}(\mu)$, the distribution $\mu^{(d)}$ give some positive probability $\delta > 0$ to the point $(x_0, \mathbf{e}) = (x_0, e_1, \ldots, e_d)$, where $x_0 = 0 \in \mathbb{Z}_q^d$, and $e_i \in \mathbb{Z}_q^d$ is the vector with 1 in the ith coordinate and 0 in all other coordinates (and δ depends only on q, d and μ). Then $(D_{\mathbf{e}}g)(x_0) = \prod_{S\subseteq[d]} g(\sum_{i\in S} e_i)^{\circ S} = \omega^{\pm a} \neq 1$ (since whenever $S \neq [d]$, we have $g(\sum_{i\in S} e_i) = 1$). On the other hand, whenever $\mathbf{t} \in (\mathbb{Z}_q^d)^d$ has some coordinate equal to 0, which also happens with positive probability depending only on d, μ and q, we have $(D_{\mathbf{t}}g(x)) = 1$. Thus in the expression

$$\|g\|_{U^d,\mu} = \left|\mathbb{E}_{(x,\mathbf{t})\sim\mu^{(d)}}\left[(D_{\mathbf{t}}f)(x)\right]\right|^{\frac{1}{2^d}},$$

since every term in the expectation has absolute value at most 1, and we just found two terms with positive probability with values 1 and $\omega^{\pm a} \neq 1$, we conclude that $\|g\|_{U^d,\mu} < 1 - \epsilon$ for some ϵ depending only on q, μ and d. ■

We now put together the above ingredients.

Theorem 2.2.10 *Let* $f : (\mathbb{Z}_q^d)^r \to \mathbb{C}$ *be given by*

$$f(x_1, \ldots, x_r) = \omega^{\sum_{j=1}^r a_j \prod_{i=1}^d x_{ij}},$$

where $a_j \in \mathbb{Z}_q \setminus \{0\}$ *for all* $j \in [r]$. *Let* μ *be a probability distribution on* $\mathbb{Z}_q^d$ *with* $\mathrm{supp}(\mu) \supseteq \{0,1\}^d$. *Then for all polynomials* $h \in \mathbb{Z}_q[(Y_{ij})_{i\in[d],j\in[r]}]$, *with* $\deg(h) < d$, *we have*

$$\left|\mathbb{E}_{x\sim\mu^{\otimes r}}\left[f(x)\omega^{h(x)}\right]\right| \leq c^r,$$

where $c < 1$ depends only on q, d and μ.

Proof Let $g_j : \mathbb{Z}_q^d \to \mathbb{C}$ be given by $g_j(y) = \omega^{a_j \prod_{i=1}^d y_i}$ (as in in Lemma 2.2.9), and take $c = 1 - \epsilon$ from that Lemma. Notice that $f = \otimes_{j=1}^r g_j$. Therefore by Lemma 2.2.7, we have

$$\|f\|_{U^d,\mu^{\otimes r}} = \prod_{j=1}^r \|g_j\|_{U^d,\mu} \leq c^r.$$

As the degree of h is at most $d-1$, Lemma 2.2.8 implies that

$$\|f\omega^h\|_{U^d,\mu^{\otimes r}} = \|f\|_{U^d,\mu^{\otimes r}} \leq c^r.$$

Lemma 2.2.5 now implies that

$$\left|\mathbb{E}_{x\sim\mu^{\otimes r}}\left[f(x)\omega^{h(x)}\right]\right| \leq c^r,$$

as desired. ■

We can now complete the proof of Lemma 2.2.3.

Proof of Lemma 2.2.3: By fixing the variables Z_i for $i \notin \cup_j E_j$, and then averaging over all such fixings, it suffices to consider the case $[m] = \cup_j E_j$. Then the polynomial $Q(Z_1, \ldots, Z_m) = \left(\sum_{j=1}^r a_j \prod_{i\in E_j} Z_i\right) + R(Z)$ can be rewritten in the form (after renaming the variables):

$$\sum_{j=1}^r a_j \prod_{i=1}^d X_{ij} + h(\mathbf{X}),$$

where $\deg(h) < d$. Let μ be the p-biased probability measure on $\{0,1\}^d \subseteq \mathbb{Z}_q^d$. Theorem 2.2.10 now implies that

$$\left|\mathbb{E}_{x\sim\mu^{\otimes r}}\left[\omega^{Q(x)}\right]\right| \leq 2^{-\Omega_{q,p,d}(r)},$$

as desired. ■

2.3 $\mathbb{F}_p^n$ versus $\mathbb{F}_{p^n}$

In this section, we build up some machinery for translating between the worlds of $\mathbb{F}_p^n$ and $\mathbb{F}_{p^n}$.

2.3.1 $\mathbb{F}$-Degree

Let $\mathbb{F}$ be a finite field. Let $f : \mathbb{F}^m \to \mathbb{F}$ be any function, and let $g(X_1, \ldots, X_m) \in \mathbb{F}[X_1, \ldots, X_m]$ be the unique polynomial with individual degrees bounded by $|\mathbb{F}| - 1$ such that for all $x \in \mathbb{F}^m$, $g(x) = f(x)$. We define the $\mathbb{F}$-*degree of* f, denoted $\deg_{\mathbb{F}}(f)$, to be the total degree of $g(X_1, \ldots, X_n)$.

We proceed to define $\mathbb{F}$-degree (also denoted $\deg_{\mathbb{F}}$) for more general functions. Let $f = (f_1, \ldots, f_n) : \mathbb{F}^m \to \mathbb{F}^n$ be any function. We define $\deg_{\mathbb{F}}(f)$ to be $\max_{i \in [n]} \deg_{\mathbb{F}}(f_i)$.

If V, W are $\mathbb{F}$-vector-spaces of dimension m, n respectively, and $f : V \to W$ is any function, we define $\deg_{\mathbb{F}}(f)$ to be $\deg_{\mathbb{F}}(\varphi_n^{-1} \circ f \circ \varphi_m)$, where $\varphi_m : \mathbb{F}^m \to V$ and $\varphi_n : \mathbb{F}^n \to W$ are arbitrary linear isomorphisms. This definition is independent of the choice of φ_m, φ_n.

Note that for functions $f : V \to W$ and $g : W \to W'$, $\deg_{\mathbb{F}}(g \circ f) \leq \deg_{\mathbb{F}}(g) \cdot \deg_{\mathbb{F}}(f)$.

A case of special importance for us is when V and W are also $\mathbb{K}$-vector-spaces, where $\mathbb{K}$ is a field containing $\mathbb{F}$ (and hence the $\mathbb{K}$-vector-space structure of V, W is compatible with their $\mathbb{F}$-vector-space structure). In this case, we may think of functions $f : V \to W$ with $\deg_{\mathbb{F}} \leq d$ as *tuples* of degree-d multivariate polynomials over $\mathbb{F}$. The following formula computes the $\mathbb{F}$-degree of a function from $\mathbb{K}$ to $\mathbb{K}$. The formula is in terms of the base-p sum-of-digits function $\mathrm{wt}_p(i)$, which equals the sum of the digits of the base-p representation of the integer i.

Lemma 2.3.1 (The $\deg_{\mathbb{F}}(f)$ Formula) *Let $\mathbb{F} \subseteq \mathbb{K}$ be finite fields. Let $f : \mathbb{K} \to \mathbb{K}$ be given by*

$$f(x) = \sum_{i \in S} a_i x^i,$$

where $a_i \neq 0$ for all $i \in S$. Then $\deg_{\mathbb{F}}(f) = \max_{i \in S} \mathrm{wt}_{|\mathbb{F}|}(i)$.

Proof This can be seen in many ways. We give a quick proof based on dimension counting. For an "explicit" proof, see [KaS].

We first prove

$$\deg_{\mathbb{F}}(f) \leq \max_{i:a_i \neq 0} \mathrm{wt}_{|\mathbb{F}|}(i). \tag{2.3}$$

Let $0 \leq i < |\mathbb{K}| - 1$ with $i = \sum_{j=0}^{[\mathbb{K}:\mathbb{F}]-1} b_j |\mathbb{F}|^j$ and $0 \leq b_j < |\mathbb{F}|$. We first consider the case $f(x) = x^i = \prod_{j=0}^{[\mathbb{K}:\mathbb{F}]-1} \prod_{k=0}^{b_j} x^{|\mathbb{F}|^j}$. We express $f : \mathbb{K} \to \mathbb{K}$ as a composition of two maps $f' : \mathbb{K} \to \prod_j \mathbb{K}^{b_j}$ and $f'' : \mathbb{K}^{\sum_j b_j} \to \mathbb{K}$, where $f'(x)_{j,r} = x^{|\mathbb{F}|^j}$ (for $r \in [b_j]$), and $f''(y_1, \ldots, y_{\sum_j b_j}) = \prod_{t=1}^{\sum_j b_j} y_t$. We see that $\deg(f') = 1$ (by the $\mathbb{F}$-linearity of the map $x \to x^{|\mathbb{F}|^j}$), and $\deg(f'') \leq \sum_j b_j = \mathrm{wt}_{|\mathbb{F}|}(i)$. Therefore $\deg_{\mathbb{F}}(f) \leq \mathrm{wt}_{|\mathbb{F}|}(i)$. For general $f = \sum_i a_i x^i$, the above case implies that $\deg_{\mathbb{F}}(f) \leq \max_{i:a_i \neq 0} \mathrm{wt}_{|\mathbb{F}|}(i)$.

Let S_d be the $\mathbb{F}$-linear space $\{g : \mathbb{K} \to \mathbb{K} \mid \deg_{\mathbb{F}}(g) \leq d\}$. We see that its $\mathbb{F}$-dimension equals

$$\dim\left(\{h : \mathbb{F}^{[\mathbb{K}:\mathbb{F}]} \to \mathbb{F} \mid \deg_{\mathbb{F}}(h) \leq d\}\right) \cdot [\mathbb{K} : \mathbb{F}] = \binom{[\mathbb{K} : \mathbb{F}] + d}{d} \cdot [\mathbb{K} : \mathbb{F}].$$

Let T_d be the $\mathbb{K}$-linear space

$$\{g : \mathbb{K} \to \mathbb{K} \mid g(x) \text{ is of the form } \sum_{\substack{0 \leq i \leq |\mathbb{K}|-1 \\ \mathrm{wt}_{|\mathbb{F}|}(i) \leq d}} \alpha_i x^i \text{ for all } x \in \mathbb{K}\}.$$

We see that its $\mathbb{K}$-dimension is $|\{0 \leq i \leq |\mathbb{K}| - 1 : \mathrm{wt}_{|\mathbb{F}|}(i) \leq d\}|$, which equals $\binom{[\mathbb{K}:\mathbb{F}]+d}{d}$. Hence its $\mathbb{F}$-dimension equals $\binom{[\mathbb{K}:\mathbb{F}]+d}{d} \cdot [\mathbb{K} : \mathbb{F}]$.

Equation (2.3) implies that $T_d \subseteq S_d$. But we just saw that $\dim_{\mathbb{F}}(S_d) = \dim_{\mathbb{F}}(T_d)$. Thus $S_d = T_d$, and the lemma follows. ■

The basic bound on the number of zeroes of a function expressible as the trace of a low-degree polynomial from a big field $\mathbb{K}$ to its prime subfield $\mathbb{F}$ is given by the Weil bound.

Theorem 2.3.2 (The Weil Bound) *Let $\mathbb{K}$ be a finite field of characteristic p. Let $f(X) \in \mathbb{K}[X]$ with $\deg_{\mathbb{K}}(f) = \ell$. Suppose f is not of the form $g(X)^p - g(X) + c$, where*

$g(X) \in \mathbb{K}[X]$ and $c \in \mathbb{K}$. Let $\omega \in \mathbb{C}$ be a primitive p^{th} root of unity. Let $\mathrm{Tr} : \mathbb{K} \to \mathbb{F}_p$ denote the finite field trace map. Then

$$\left|\mathbb{E}_{x \in \mathbb{K}}\left[\omega^{\mathrm{Tr}(f(x))}\right]\right| \leq \frac{\ell}{|\mathbb{K}|^{1/2}}.$$

In Appendix A, we give a exposition of an elementary proof of this theorem for fields of characteristic 2, following Bombieri and Stepanov.

2.3.2 Discrete Directional Derivatives

We now revisit the notion of discrete directional derivative, earlier discussed in Section 2.2.1, and study the relationship between the $\mathbb{F}$-degree, the $\mathbb{K}$-degree and the operation of taking derivatives for functions $f : \mathbb{K} \to \mathbb{K}$, where $\mathbb{F}$ is a subfield of $\mathbb{K}$.

Let $f : V \to W$ be a function between $\mathbb{F}$ vector spaces V, W. For $a \in V$, we define *the discrete directional derivative of f in direction a*, $D_a f : V \to W$ by the equation

$$(D_a f)(x) = f(x + a) - f(x).$$

For $\mathbf{a} = (a_1, \ldots, a_k) \in V^k$, we define (inductively) $D_{\mathbf{a}} f : V \to W$ to be the function $D_{a_1}(D_{(a_2,\ldots,a_k)} f)$. It can be seen that for vectors $\mathbf{a} \in V^{k_1}$ and $\mathbf{b} \in V^{k_2}$,

$$D_{\mathbf{a}}(D_{\mathbf{b}} f) = D_{\mathbf{b}}(D_{\mathbf{a}} f) = D_{\mathbf{c}} f,$$

where $\mathbf{c} = (a_1, \ldots, a_{k_1}, b_1, \ldots, b_{k_2}) \in V^{k_1+k_2}$. Explicitly, we have

$$D_{\mathbf{a}} f(x) = \sum_{I \subseteq [k]} (-1)^{k-|I|} f\left(x + \sum_{i \in I} a_i\right). \tag{2.4}$$

In particular, it can be seen that $D_{\mathbf{a}}$ is a *linear* operator on functions from V to W. If $h : W \to W'$ is an $\mathbb{F}$-linear map of $\mathbb{F}$-vector-spaces, then we have the commutativity relation

$$D_{\mathbf{a}}(h \circ f) = h \circ (D_{\mathbf{a}} f).$$

We now summarize some facts describing the interplay between taking derivatives of functions and their degree.

Fact 2.3.3 *Let $\mathbb{F}$ be a field and let V and W be $\mathbb{F}$-vector-spaces. Let $f : V \to W$. Let $h : V \times V^k \to W$ be given by $h(x, \mathbf{a}) = D_{\mathbf{a}} f(x)$.*

1. *If $k \leq d$, then for all $\mathbf{a} \in V^k$, $\deg_{\mathbb{F}}(D_{\mathbf{a}} f) \leq \deg_{\mathbb{F}}(f) - k$.*
2. *If $k \leq d$, then $\deg_{\mathbb{F}}(h) \leq \deg_{\mathbb{F}}(f)$.*
3. *If $k > d$, then for all $\mathbf{a} \in V^k$, $D_{\mathbf{a}} f = 0$.*

Proof It suffices to show the result for $k = 1$, $V = \mathbb{F}^m$ and $W = \mathbb{F}$ (as we may then induct on k). We first consider the case when $f(x) = \prod_{i=1}^m x_i^{e_i}$, where $0 \leq e_i \leq |\mathbb{F}| - 1$ for each i. Then $d = \sum_{i=1}^m e_i$. We may now compute $D_a f(x) = \prod_{i=1}^m (x_i + a_i)^{e_i} - \prod_{i=1}^m (x_i)^{e_i}$, which has total $\mathbb{F}$-degree at most $\sum_{i=1}^m e_i - 1 = d - 1$. Similarly, the $\mathbb{F}$-degree of h is at most $\sum_{i=1}^m e_i = d$.

The case of general f which is a sum of monomials now follows from the above case and linearity of D_a. ■

The following fact (which may be proved by induction on k) gives finer information about the function $D_{\mathbf{a}} f$. Note that the multinomial coefficients below may equal 0 over the field $\mathbb{F}$.

Fact 2.3.4 *If $f : \mathbb{F} \to \mathbb{F}$ is given by $f(x) = x^e$, and $\mathbf{a} \in \mathbb{F}^k$, then*

$$D_{\mathbf{a}} f(x) = \sum_{r + r_1 + r_2 + \ldots r_k = e} \binom{e}{r, r_1, \ldots, r_k} x^r \prod_{i=1}^k a_i^{r_i},$$

and hence $D_{\mathbf{a}} f(x)$ may be written as $\sum_{r \leq e} x^r h_r(a_1, \ldots, a_k)$, where $h_r(A_1, \ldots, A_k) \in \mathbb{F}[A_1, \ldots, A_k]$ is a homogeneous polynomial of degree $e - r$

In particular, this fact implies that if $\mathbb{F}$ is of characteristic 2, $\mathrm{wt}_2(e) \geq k$ and $2 \not| e$, then there is an $r < e/2$ such that $h_r(A_1, \ldots, A_k)$ is a nonzero homogeneous polynomial of degree $e - r$, and $2 \not| e - r$. To see this, if $e = \sum_{j=1}^d 2^{e_j}$ with $0 = e_1 < e_2 < \ldots < e_d$, then we may take $r = \sum_{j=2}^{d-k+1} 2^{e_j}$. Then for $r_1 = 2^{e_1}$ and $r_i = 2^{e_{d-k+i}}$ for $2 \leq i \leq k$, it can be checked that that the coefficient of $x^r \prod_{i=1}^k a_i^{r_i}$ is nonzero modulo 2.

2.4 Subspace polynomials

In this section, we give a brief introduction to the theory of subspace polynomials. A detailed study of subspace polynomials was first carried out in the work of [Ore33, Ore34]. We refer the reader interested in a more thorough introduction to the subject to [LN97, Chapter 4] and to [Ber68, Chapter 11].

A polynomial $P \in \mathbb{F}_{p^n}[X]$ is said to be $\mathbb{F}_p$-*linearized* if it is of the form:

$$P(X) = \sum_{i=0}^{n-1} a_i X^{p^i}, a_i \in \mathbb{F}_{p^n}$$

(when p is clear from context, we will simply refer to them as linearized polynomials). P being linearized is equivalent to having $P(\beta b + \gamma c) = \beta P(b) + \gamma P(c)$ for all $b, c \in \mathbb{F}_{p^n}$ and $\beta, \gamma \in \mathbb{F}_p$. By extension, a polynomial is said to be *affine linearized* if $P(X) = \hat{P}(X) + \hat{a}$ where $\hat{P}$ is linearized and $\hat{a} \in \mathbb{F}_{p^n}$. The affine linearized polynomials over $\mathbb{F}_{p^n}$ are precisely the polynomials of $\mathbb{F}_p$-degree at most 1.

The next lemma, which follows from Lemma 2.3.1, shows that every affine transformation corresponds to an affine linearized polynomial.

Lemma 2.4.1 *Let $\phi : \mathbb{F}_p^n \to \mathbb{F}_{p^n}$ be an $\mathbb{F}_p$-linear isomorphism. There is a one-to-one correspondence between affine transformations from $\mathbb{F}_p^n$ to $\mathbb{F}_p^n$ and affine linearized polynomials in $\mathbb{F}_{p^n}[X]$, i.e., for every affine transformation $T : \mathbb{F}_p^n \to \mathbb{F}_p^n$ there exists a unique affine linearized polynomial P_T satisfying $P_T(\phi(b)) = T(\phi(b))$ for all $b \in \mathbb{F}_p^n$.*

We shall take particular interest in a special class of linearized polynomials that split completely in $\mathbb{F}_{p^n}$ to a set of roots that forms a $\mathbb{F}_p$-affine subspace of $\mathbb{F}_{p^n}$.

Definition 2.4.2 (Kernel-subspace polynomial) *Let $L \subseteq \mathbb{F}_{p^n}$ be an affine subspace of dimension d. Define $P_L(X) \in \mathbb{F}_{p^n}[X]$, the* kernel-subspace polynomial *of L, to be*

$$P_L(X) = \prod_{\alpha \in L} (X - \alpha).$$

We have the following interesting fact.

Lemma 2.4.3 (Kernel-subspace polynomials are affine) *If $L \subseteq \mathbb{F}_{p^n}$ is an affine subspace of dimension d then $P_L(X)$ is a monic affine linearized polynomial of degree p^d. Furthermore, P_L is linearized iff L is a linear space.*

Every kernel-subspace polynomial P_L corresponds to an affine transformation whose kernel is L, so by linearity $P(\mathbb{F}_{p^n})$ is an affine subspace of $\mathbb{F}_{p^n}$ of dimension $n - \dim(L)$. Surprisingly, every $\mathbb{F}$-subspace of $\mathbb{F}_{p^n}$ arises as the image of $\mathbb{F}_{p^n}$ under P_L for some $\mathbb{F}$-subspace L. These *image-subspace* polynomials will be the starting point of our analysis of affine dispersers.

The next lemma shows the existence of an image-subspace polynomial for every subspace. We include the beautiful proof of this lemma from [Ber68].

Lemma 2.4.4 (Existence of an image-subspace polynomial) *If $L \subseteq \mathbb{F}_{p^n}$ is an affine subspace of dimension d then there exists a monic affine linearized polynomial $Q_L(X)$ with* $\deg(Q_L) = p^{n-d}$, *called the* image-subspace polynomial *of L, such that*

$$L = Q_L(\mathbb{F}_{p^n}) \triangleq \{Q_L(c) \mid c \in \mathbb{F}_{p^n}\}.$$

Moreover, if $P_L(X)$ is the subspace polynomial of L then

$$P_L(Q_L(X)) \equiv Q_L(P_L(X)) \equiv X^{p^n} - X. \tag{2.5}$$

Thus the kernel of $Q_L : \mathbb{F}_{p^n} \to \mathbb{F}_{p^n}$ is the image of $P_L : \mathbb{F}_{p^n} \to \mathbb{F}_{p^n}$. In particular $Q_L(X)$ has p^{n-d} roots in $\mathbb{F}_{p^n}$, and is thus also a kernel subspace polynomial of some $(n-d)$-dimensional subspace.

Proof Let $L' = P_L(\mathbb{F}_{p^n})$ be the image of $P_L(X)$. Define $Q_L(X)$ to be $P_{L'}(X)$, the kernel-subspace polynomial of L'.

Notice that $Q_L(P_L(X))$ is a monic polynomial of degree p^n that vanishes on $\mathbb{F}_{p^n}$, hence $Q_L(P_L(X)) = X^{p^n} - X$. Thus $P_L(Q_L(P_L(X)))) = P_L(X^{p^n} - X) = P_L(X^{p^n}) - P_L(X) = P_L(X)^{p^n} - P_L(X)$. Letting $g(Y)$ be the polynomial $P_L(Q_L(Y)) - (Y^{p^n} - Y)$, we have just proved that $g(P_L(X)) = 0$. This implies $g(Y) = 0$, since $\deg(g(P_L(X))) = (\deg g(Y)) \cdot (\deg(P_L(X)))$.

So $P_L(Q_L(y)) = 0$ for each $y \in \mathbb{F}_{p^n}$. In particular, we see that the image of Q_L is contained in L, and by dimension counting, the image of Q_L equals L. ■

Let us pause to appreciate the strength of this theorem. This theorem says that for any linear space $L \subseteq \mathbb{F}_q$, there is a polynomial Q_L of degree $q/|L|$ which maps $\mathbb{F}_q$ onto L. This is the smallest possible degree that a polynomial mapping $\mathbb{F}_q$ to L can have: any such polynomial Q must have some point $x \in L$ with $|\{y \in \mathbb{F}_q \mid Q(y) = x\}| \geq q/|L|$. Such polynomials of degree $q/|L|$ do not exist for arbitrary subsets $L \subseteq \mathbb{F}_q$.

Chapter 3

Random Graphs and the Parity Quantifier

3.1 Introduction

For quite a long time, combinatorialists have studied the asymptotic probabilities of properties on classes of finite structures, such as graphs and partial orders. Assume that $\mathcal{C}$ is a class of finite structures and let $\Pr_n$, $n \geq 1$, be a sequence of probability measures on all structures in $\mathcal{C}$ with n elements in their domain. If Q is a property of some structures in $\mathcal{C}$ (that is, a decision problem on $\mathcal{C}$), then the *asymptotic probability* $\Pr(Q)$ *of* Q *on* $\mathcal{C}$ is defined as $\Pr(Q) = \lim_{n\to\infty} \Pr_n(Q)$, provided this limit exists. In this chapter, we will be focusing on the case when $\mathcal{C}$ is the class $\mathcal{G}$ of all finite graphs, and $\Pr_n = G(n,p)$ for constant p; this is the probability distribution on n-vertex undirected graphs where between each pair of nodes an edge appears with probability p, independently of other pairs of nodes. For example, for this case, the asymptotic probabilities $\Pr(\textsc{Connectivity}) = 1$ and $\Pr(\textsc{Hamiltonicity}) = 1$; in contrast, if $\Pr_n = G(n,p(n))$ with $p(n) = 1/n$, then $\Pr(\textsc{Connectivity}) = 0$ and $\Pr(\textsc{Hamiltonicity}) = 0$.

Instead of studying separately one property at a time, it is natural to consider formalisms for specifying properties of finite structures and to investigate the connection between the expressibility of a property in a certain formalism and its asymptotic

probability. The first and most celebrated such connection was established by Glebskii et al. [GKLT69] and, independently, by Fagin [Fag76], who showed that a *0-1 law* holds for first-order logic[1] FO on the random graph $G(n,p)$ with p a constant in $(0,1)$; this means that if Q is a property of graphs expressible in FO and $\Pr_n = G(n,p)$ with p a constant in $(0,1)$, then $\Pr(Q)$ exists and is either 0 or 1. This result became the catalyst for a series of investigations in several different directions. Specifically, one line of investigation [SS87, SS88] investigated the existence of 0-1 laws for first-order logic FO on the random graph $G(n,p(n))$ with $p(n) = n^{-\alpha}$, $0 < \alpha < 1$. Since first-order logic on finite graphs has limited expressive power (for example, FO cannot express CONNECTIVITY and 2-COLORABILITY), a different line of investigation pursued 0-1 laws for extensions of first-order logic on the random graph $G(n,p)$ with p a constant in $(0,1)$. In this vein, it was shown in [BGK85, KV87] that the 0-1 law holds for extensions of FO with fixed-point operators, such as least fixed-point logic LFP, which can express CONNECTIVITY and 2-COLORABILITY. As regards to higher-order logics, it is clear that the 0-1 law fails even for existential second-order logic ESO, since it is well known that ESO = NP on finite graphs [Fag74]. In fact, even the *convergence law* fails for ESO, that is, there are ESO-expressible properties Q of finite graphs such that $\Pr(Q)$ does *not* exist. For this reason, a separate line of investigation pursued 0-1 laws for syntactically-defined subclasses of NP. Eventually, this investigation produced a complete classification of the quantifier prefixes of ESO for which the 0-1 law holds [KV87, KV90, PS89], and provided a unifying account for the asymptotic probabilities of such NP-complete problems as k-COLORABILITY, $k \geq 3$.

Let L be a logic for which the 0-1 law (or even just the convergence law) holds on the random graph $G(n,p)$ with p a constant in $(0,1)$. An immediate consequence of this is that L cannot express any *counting* properties, such as EVEN CARDINALITY ("there is an even number of nodes"), since $\Pr_{2n}(\text{EVEN CARDINALITY}) = 1$

[1]Recall that the formulas of first-order logic on graphs are obtained from atomic formulas $E(x,y)$ (interpreted as the adjacency relation) and equality formulas $x = y$ using Boolean combinations, existential quantification, and universal quantification; the quantifiers are interpreted as ranging over the set of vertices of the graph (and not over sets of vertices or sets of edges, etc.).

and $\Pr_{2n+1}(\text{EVEN CARDINALITY}) = 0$. In this chapter, we turn the tables around and systematically investigate the asymptotic probabilities of properties expressible in extensions of FO with *counting quantifiers* Mod_q^i, where q is a prime number. The most prominent such extension is FO[$\oplus$], which is the extension of FO with the *parity quantifier* Mod_2^1. The syntax of FO[$\oplus$] augments the syntax of FO with the following formation rule: if $\varphi(y)$ is a FO[$\oplus$]-formula, then $\oplus y\varphi(y)$ is also a FO[$\oplus$]-formula; this formula is true if the number of y's that satisfy $\varphi(y)$ is odd (analogously, $\mathsf{Mod}_q^i y\varphi(y)$ is true if the number of y's that satisfy $\varphi(y)$ is congruent to $i \mod q$). A typical property on graphs expressible in FO[$\oplus$] (but not in FO) is $\mathcal{P} := \{G :$ every vertex of G has odd degree$\}$, since a graph is in $\mathcal{P}$ if and only if it satisfies the FO[$\oplus$]-sentence $\forall x \oplus y E(x,y)$.

There are two notable "reasons" to which one can attribute the failure of the 0-1 law for FO[$\oplus$] on the random graph $G(n,p)$, with p a constant. The first, most glaring, reason is that FO[$\oplus$] can express the property EVEN CARDINALITY, whose asymptotic probability does not converge. The other, more subtle, reason comes from properties that express "subgraph counting" mod 2. For a fixed graph H, FO[$\oplus$] can express the property $\mathcal{P}_H$: "the number of induced copies of H is even". It turns out that as $n \to \infty$, for a typical connected graph H, the probability that $G(n,p)$ has $\mathcal{P}_H$ tends to $1/2$ (we shall prove this later in the chapter). Thus in this case, asymptotic probability converges, but does not equal 0 or 1. The above two phenomena must be accounted for in any law describing the asymptotic probabilities of FO[$\oplus$] sentences on $G(n,p)$.

The main result of this chapter (see Theorem 3.2.1) is a *modular convergence law* for FO[$\oplus$] on $G(n,p)$ with p a constant in $(0,1)$. This law asserts that if φ is a FO[$\oplus$]-sentence, then there are two explicitly computable rational numbers a_0, a_1, such that, as $n \to \infty$, the probability that the random graph $G(2n+i,p)$ satisfies φ approaches a_i, for $i = 0, 1$. Moreover, a_0 and a_1 are of the form $r/2^s$, where r and s are non-negative integers. We also establish that an analogous modular convergence law holds for every extension FO[Mod_q] of FO with the counting quantifiers $\{\mathsf{Mod}_q^i : i \in [q-1]\}$, where q is a prime. It should be noted that results in [HKL96] imply that the modular

convergence law for FO[$\oplus$] does *not* generalize to extensions of FO[$\oplus$] with fixed-point operators. This is in sharp contrast to the aforementioned 0-1 law for FO which carries over to extensions of FO with fixed-point operators.

3.1.1 Methods

Earlier 0-1 laws have been established by a combination of standard methods and techniques from mathematical logic and random graph theory. In particular, on the side of mathematical logic, the tools used include the compactness theorem, Ehrenfeucht-Fraïssé games, and quantifier elimination. Here, we establish the modular convergence law by combining quantifier elimination with, interestingly, algebraic methods related to multivariate polynomials over finite fields. In what follows in this section, we present an overview of the methods and techniques that we will use.

The distribution of subgraph frequencies mod q, polynomials and Gowers norms

Let us briefly indicate the relevance of polynomials to the study of FO[$\oplus$] on random graphs. A natural example of a statement in FO[$\oplus$] is a formula φ such that G satisfies φ if and only if the number of copies of H in G is odd, for some graph H (where by copy we mean an induced subgraph, for now). Thus understanding the asymptotic probability of φ on $G(n,p)$ amounts to understanding the distribution of the number of copies (mod 2) of H in $G(n,p)$.

In this spirit, we ask: what is the probability that in $G(n,1/2)$ there is an odd number of triangles (where we count *unordered* triplets of vertices $\{a,b,c\}$ such that a,b,c are all pairwise adjacent[2])?

We reformulate this question in terms of the following "triangle polynomial", that takes the adjacency matrix of a graph as input and returns the parity of the number

[2]Counting the number of *unordered* triples is not expressible in FO[$\oplus$], we ask this question only for expository purposes (nevertheless, we do give an answer to this question in Section 3.3).

of triangles in the graph; $P_\triangle : \{0,1\}^{\binom{n}{2}} \to \{0,1\}$, where

$$P_\triangle((x_e)_{e\in\binom{n}{2}}) = \sum_{\{e_1,e_2,e_3\} \text{ forming a } \triangle} x_{e_1}x_{e_2}x_{e_3},$$

where the arithmetic is mod 2. Note that for the random graph $G(n,1/2)$, each entry of the adjacency matrix is chosen independently and uniformly from $\{0,1\}$. Thus the probability that a random graph $G \in G(n,1/2)$ has an odd number of triangles is precisely equal to $\Pr_{x\in\mathbb{Z}_2^n}[P_\triangle(x) = 1]$. Thus we have reduced our problem to studying the distribution of the evaluation of a certain polynomial at a random point, a topic of much study in pseudorandomness and algebraic coding theory, and we may now appeal to tools from these areas.

In Section 3.3, via the above approach, we show that the probability that $G(n,1/2)$ has an odd number of triangles equals $1/2 \pm 2^{-\Omega(n)}$. Similarly, for any connected graph $F \neq \mathrm{K}_1$ (the graph consisting of one vertex), the probability that $G(n,1/2)$ has an odd number of copies[3] of F is also $1/2 \pm 2^{-\Omega(n)}$ (when $F = \mathrm{K}_1$, there is no randomness in the number of copies of F in $G(n,1/2)$!). In fact, we show that for any collection of distinct connected graphs $F_1, \ldots, F_\ell$ ($\neq \mathrm{K}_1$), the joint distribution of the number of copies mod 2 of $F_1, \ldots, F_\ell$ in $G(n,1/2)$ is $2^{-\Omega(n)}$-close to the uniform distribution on $\mathbb{Z}_2^\ell$, i.e., the events that there are an odd number of F_i are essentially independent of one another.

Generalizing the above to $G(n,p)$ and counting mod q for arbitrary $p \in (0,1)$ and arbitrary integers q motivates the study of new kinds of questions about polynomials, that we believe are interesting in their own right. For $G(n,p)$ with arbitrary p, we need to study the distribution of $P(x)$, for certain polynomials P, when $x \in \mathbb{Z}_2^m$ is distributed according to the *p-biased measure*. Even more interestingly, for the study of $\mathsf{FO}[\mathsf{Mod}_q]$, where we are interested in the distribution of the number of triangles mod q, one needs to understand the distribution of $P(x)$ (P is now a polynomial over $\mathbb{Z}_q$) where x is chosen uniformly from $\{0,1\}^m \subseteq \mathbb{Z}_q^m$ (as opposed to x being chosen uniformly from all of $\mathbb{Z}_q^m$, which is traditionally studied). In Section 3.4, we

[3]with a certain precise definition of "copy".

develop all the relevant polynomial machinery in order to answer these questions. This involves generalizing some classical results of Babai, Nisan and Szegedy [BNS89] on correlations of polynomials. The key technical innovation here is our definition of a μ-Gowers norm (where μ is a measure on $\mathbb{Z}_q^m$) that measures the correlation, under μ, of a given function with low-degree polynomials (letting μ be the uniform measure, we recover the standard Gowers norm). After generalizing several results about the standard Gowers norm to the μ-Gowers norm case, we can then use a technique of Viola and Wigderson [VW07] to establish the generalization of [BNS89] that we need.

Quantifier elimination

Although we studied the distribution of subgraph frequencies mod q as an attempt to determine the limiting behavior of only a special family of $\mathsf{FO}[\mathsf{Mod}_q]$ properties, it turns out that this case, along with the techniques developed to handle it, play a central role in the proof of the full modular convergence law. In fact, we reduce the modular convergence law for general $\mathsf{FO}[\mathsf{Mod}_q]$ properties to the above case. We show that for any $\mathsf{FO}[\mathsf{Mod}_q]$ sentence φ, with high probability over $G \in G(n,p)$, the truth of φ on G is determined by the number of copies in G, mod q, of each small subgraph. Then by the results described earlier on the equidistribution of these numbers (except for the number of K_1, which depends only on $n \mod q$), the full modular convergence law for $\mathsf{FO}[\mathsf{Mod}_q]$ follows.

In Section 3.6, we establish such a reduction using the method of elimination of quantifiers. To execute this, we need to analyze $\mathsf{FO}[\mathsf{Mod}_q]$ formulas which may contain free variables (i.e., not every variable used is quantified). Specifically, we show that for every $\mathsf{FO}[\mathsf{Mod}_q]$ formula $\varphi(\alpha_1, \ldots, \alpha_k)$, with high probability over $G \in G(n,p)$, it holds that for all vertices $w_1, \ldots, w_k$ of G, the truth of $\varphi(w_1, \ldots, w_k)$ is entirely determined by the following data: (a) which of the w_i, w_j pairs are adjacent, (b) which of the w_i, w_j pairs are equal to one another, and (c) the number of copies "rooted" at $w_1, \ldots, w_k$, mod q, of each small *labelled graph*. This statement is a generalization of what we needed to prove, but lends itself to inductive proof (*this* is quantifier elimination). This leads us to studying the distribution (via the polynomial approach

described earlier) of the number of copies of labelled graphs in G; questions of the form, given two specified vertices v, w (the "roots"), what is the probability that there are an odd number of paths of length 4 in $G \in G(n,p)$ from v to w? After developing the necessary results on the distribution of labelled subgraph frequencies, combined with some elementary combinatorics, we can eliminate quantifiers and thus complete the proof of the modular convergence law.

3.1.2 Comparison with $\mathsf{AC}^0[\oplus]$

Every $\mathsf{FO}[\oplus]$ property naturally defines a family of boolean functions $f_n : \{0,1\}^{\binom{n}{2}} \to \{0,1\}$, such that a graph G satisfies φ if and only if $f_n(A_G) = 1$, where A_G is the adjacency matrix of G. This family of functions is easily seen to be contained in $\mathsf{AC}^0[\oplus]$, which is AC^0 with parity gates (each $\forall$ becomes an AND gate, $\exists$ becomes a OR gate and $\oplus$ becomes a parity gate). This may be summarized by saying that $\mathsf{FO}[\oplus]$ is a highly uniform version of $\mathsf{AC}^0[\oplus]$.

Currently, all our understanding of the power of $\mathsf{AC}^0[\oplus]$ comes from the Razborov-Smolensky [Raz87b, Smo87] approach to proving circuit lower bounds on $\mathsf{AC}^0[\oplus]$. At the heart of this approach is the result that for every $\mathsf{AC}^0[\oplus]$ function f, there is a low-degree polynomial P such that for $1 - \epsilon(n)$ fraction of inputs, the evaluations of f and P are equal. Note that this result automatically holds for $\mathsf{FO}[\oplus]$ (since $\mathsf{FO}[\oplus] \subseteq \mathsf{AC}^0[\oplus]$).

We show that for the special case when $f : \{0,1\}^{\binom{n}{2}} \to \{0,1\}$ comes from an $\mathsf{FO}[\oplus]$ property φ, a significantly improved approximation may be obtained: *(i)* We show that the degree of P may be chosen to be a constant depending only on φ, whereas the Razborov-Smolensky approximation required P to be of polylog(n) degree, *(ii)* The error parameter $\epsilon(n)$ may be chosen to be exponentially small in n, whereas the Razborov-Smolensky method only yields $\epsilon(n) = 2^{-\log^{O(1)} n}$. *(iii):* Finally, the polynomial P can be chosen to be symmetric under the action of S_n on the $\binom{n}{2}$ coordinates, while in general, the polynomial produced by the Razborov-Smolensky approach need not be symmetric (due to the randomness involved in the choices).

These strengthened approximation results allow us, using known results about

pseudorandomness against low-degree polynomials, to show that *(i)* there exist explicit pseudorandom generators that fool $\mathsf{FO}[\oplus]$ sentences, and *(ii)* there exist explicit functions f such that for any $\mathsf{FO}[\oplus]$ formula φ, the probability over $G \in G(n,p)$ that $f(G) = \varphi(G)$ is at most $\frac{1}{2} + 2^{-\Omega(n)}$. The first result follows from the pseudorandom generators against low-degree polynomials due to Bogdanov-Viola [BV07], Lovett [Lov08] and Viola [Vio08]. The second result follows from the result of Babai, Nisan and Szegedy [BNS89], and our generalization of it, giving explicit functions that are uncorrelated with low degree polynomials.

Obtaining similar results for $\mathsf{AC}^0[\oplus]$ is one of the primary goals of modern day "low-level" complexity theory.

Organization of this chapter: In the next section, we formally state our main results and some of its corollaries. In Section 3.3, we determine the distribution of subgraph frequencies mod q. In Section 3.4, we prove a theorem, which is needed for the previous section, which gives a simple criterion for a polynomial to be unbiased. In Section 3.5, we state the theorem which implements the quantifier elimination and describe the plan for its proof. This plan is then executed in Sections 3.6, 3.7 and 3.8. We conclude with some open questions.

3.2 The Main Result and its Corollaries

We now state our main theorem.

Theorem 3.2.1 *Let q be a prime. Then for every $\mathsf{FO}[\mathsf{Mod}_q]$-sentence φ, there exist rationals $a_0, \ldots, a_{q-1}$ such that for every $p \in (0,1)$ and every $i \in \{0, 1, \ldots, q-1\}$,*

$$\lim_{\substack{n \to \infty \\ n \equiv i \bmod q}} \Pr_{G \in G(n,p)}[G \text{ satisfies } \varphi] = a_i.$$

Remark The proof of Theorem 3.2.1 also yields:

- Given the formula φ, the numbers $a_0, \ldots, a_{q-1}$ can be computed.

- Each a_i is of the form r/q^s, where r, s are nonnegative integers.
- For every sequence of numbers $b_0, \ldots, b_{q-1} \in [0,1]$, each of the form r/q^s, there is a $\mathsf{FO}[\mathsf{Mod}_q]$-sentence φ such that for each i, the number a_i given by the theorem equals b_i.

Before we describe the main steps in the proof of Theorem 3.2.1, we make a few definitions.

For graphs $F = (V_F, E_F)$ and $G = (V_G, E_G)$, an *(injective) homomorphism* from F to G is an (injective) map $\chi : V_F \to V_G$ that maps edges to edges, i.e., for any $(u,v) \in E_F$, we have $(\chi(u), \chi(v)) \in E_G$. Note that we do not require that χ maps non-edges to non-edges. We denote by $[F](G)$ the number of injective homomorphisms from F to G, and we denote by $[F]_q(G)$ this number mod q. We let $\mathsf{aut}(F) := [F](F)$ be the number of automorphisms of F.

The following lemma (which follows from Lemma 3.6.5 in Section 3.6), shows that for some graphs F, as G varies, the number $[F](G)$ cannot be arbitrary.

Lemma 3.2.2 *Let F be a connected graph and G be any graph. Then* $\mathsf{aut}(F) \mid [F](G)$.

For the rest of this section, let q be a fixed prime. Let Conn^a be the set of connected graphs on at most a vertices. For any graph G, let the *subgraph frequency vector* $\mathsf{freq}_G^a \in \mathbb{Z}_q^{\mathsf{Conn}^a}$ be the vector such that its value in coordinate F ($F \in \mathsf{Conn}^a$) equals $[F]_q(G)$, the number of injective homomorphisms from F to G mod q. Let $\mathsf{FFreq}(a)$, the set of *feasible frequency vectors*, be the subset of $\mathbb{Z}_q^{\mathsf{Conn}^a}$ consisting of all f such that for all $F \in \mathsf{Conn}^a$, $f_F \in \mathsf{aut}(F) \cdot \mathbb{Z}_q := \{\mathsf{aut}(F) \cdot x \mid x \in \mathbb{Z}_q\}$. By Lemma 3.2.2, for every G and a, $\mathsf{freq}_G^a \in \mathsf{FFreq}(a)$, i.e., the subgraph frequency vector is always a feasible frequency vector.

We can now state the two main technical results that underlie Theorem 3.2.1.

The first states that on almost all graphs G, every $\mathsf{FO}[\mathsf{Mod}_q]$ formula can be expressed in terms of the subgraph frequencies, $[F]_q(G)$, over all small connected graphs F.

Theorem 3.2.3 (Subgraph frequencies mod q **determine** $\mathsf{FO}[\mathsf{Mod}_q]$ **formulae)** *For every* $\mathsf{FO}[\mathsf{Mod}_q]$*-sentence* φ *of quantifier depth* t*, there exists an integer* $c = c(t,q)$ *and a function* $\psi : \mathbb{Z}_q^{\mathsf{Conn}^c} \to \{0,1\}$ *such that for all* $p \in (0,1)$*,*

$$\Pr_{G \in G(n,p)} [(G \text{ satisfies } \varphi) \Leftrightarrow (\psi(\mathsf{freq}_G^c) = 1)] \geq 1 - \exp(-n).$$

This result is complemented by the following result, that shows the distribution of subgraph frequencies mod q in a random graph $G \in G(n,p)$ is essentially uniform in the space of all feasible frequency vectors, up to the obvious restriction that the number of vertices (namely the frequency of K_1 in G) should equal $n \mod q$.

Theorem 3.2.4 (Distribution of subgraph frequencies mod q **depends only on** $n \mod q$**)** *Let* $p \in (0,1)$*. Let* $G \in G(n,p)$*. Then for any constant* a*, the distribution of* freq_G^a *is* $\exp(-n)$*-close to the uniform distribution over the set*

$$\{f \in \mathsf{FFreq}(a) : f_{K_1} \equiv n \mod q\}.$$

Theorem 3.2.4 is proved in Section 3.3 by studying the bias of multivariate polynomials over finite fields via a generalization of the Gowers norm. Theorem 3.2.3 is proved in Section 3.6 using two main ingredients:

1. A generalization of Theorem 3.2.4 that determines the joint distribution of the frequencies of "labelled subgraphs" with given roots (see Section 3.8).

2. A variant of quantifier elimination (which may be called quantifier conversion) designed to handle Mod_q quantifiers, that crucially uses the probabilistic input from the previous ingredient (see Section 3.6).

Proof of Theorem 3.2.1: Follows by combining Theorem 3.2.3 and Theorem 3.2.4. ■

We quickly give some examples of the finer information about modular convergence that can be derived from Theorem 3.2.3 and Theorem 3.2.4.

Observe that $\mathsf{FFreq}(a) \subseteq \mathbb{Z}_q^{\mathsf{Conn}^a}$ is a product set: indeed, it equals $\prod_{F \in \mathsf{Conn}^a} (\mathsf{aut}(F) \cdot \mathbb{Z}_q)$. In particular, we see that $|\{f \in \mathsf{FFreq}(a) \mid f_{K_1} \equiv n \mod q\}|$ is a power of q. This implies that the numbers a_i in Theorem 3.2.1 are all of the form α/q^β.

Next, observe that the property "$[F]_q(G) = i$" is expressible in $\mathsf{FO}[\mathsf{Mod}_q]$. This observation, combined with Theorem 3.2.4, easily implies that for every collection of numbers $b_0, \ldots, b_{q-1}$ of the form α/q^β, there is an $\mathsf{FO}[\mathsf{Mod}_q]$ statement for which each number a_i given by Theorem 3.2.1 equals the corresponding b_i.

3.2.1 Pseudorandomness against $\mathsf{FO}[\mathsf{Mod}_q]$

We now point out three simple corollaries of our study of $\mathsf{FO}[\mathsf{Mod}_q]$ on random graphs.

Corollary 3.2.5 ($\mathsf{FO}[\mathsf{Mod}_q]$ is well approximated by low-degree polynomials) *For every $\mathsf{FO}[\mathsf{Mod}_q]$-sentence φ, there is a constant d, such that for each $n \in \mathbb{N}$, there is a degree d polynomial $P((X_e)_{e \in \binom{n}{2}}) \in \mathbb{Z}_q[(X_e)_{e \in \binom{n}{2}}]$, such that for all $p \in (0,1)$,*

$$\Pr_{G \in G(n,p)}[(G \text{ satisfies } \varphi) \Leftrightarrow P(A_G) = 1] \geq 1 - 2^{-\Omega(n)},$$

where $A_G \in \{0,1\}^{\binom{n}{2}}$ is the adjacency matrix of G.

Proof Follows from Theorem 3.2.3 and the observation that for any graph F of constant size, there is a polynomial $Q((X_e)_{e \in \binom{n}{2}})$ of constant degree, such that $Q(A_G) = [F]_q(G)$ for all graphs G. ■

Corollary 3.2.6 (Pseudorandom generators against $\mathsf{FO}[\mathsf{Mod}_q]$) *For each $s \in \mathbb{N}$ and constant $\epsilon > 0$, there is a constant $c \geq 0$ such that for each n, there is a family $\mathcal{F}$ of $\Theta(n^c)$ graphs on n vertices, computable in time $\mathrm{poly}(n^c)$, such that for all $\mathsf{FO}[\mathsf{Mod}_q]$-sentences φ of size at most s, and for all $p \in (0,1)$,*

$$|\Pr_{G \in \mathcal{F}}[G \text{ satisfies } \varphi] - \Pr_{G \in G(n,p)}[G \text{ satisfies } \varphi]| < \epsilon.$$

Proof For $q = 2$ and $p = 1/2$, this follows from the previous corollary and the result of Viola [Vio08] (building on results of Bogdanov-Viola [BV07] and Lovett [Lov08]) explicitly constructing a set of points fooling low-degree polynomials under the uniform distribution. For $q = 2$ and general p, note that the same family $\mathcal{F}$ from the $p = 1/2$ case works, since the distribution of subgraph frequencies given in Theorem 3.2.4 is independent of p.

For general q, a slight complication arises because adjacency matrices of graphs have entries from $\{0, 1\}$, while the result of Viola for polynomials over $\mathbb{Z}_q$ constructs points the uniform distribution over $\mathbb{Z}_q^m$. Nevertheless, we get by with a trick[4]. Let $P(X_1, \ldots, X_m)$ be a polynomial over $\mathbb{Z}_q$ and let $P'(Y_1, \ldots, Y_m) = P(Y_1^{q-1}, \ldots, Y_m^{q-1})$. Then the distribution of the evaluation of P' at a uniformly random point from $\mathbb{Z}_q^m$ is identical to the distribution of the evaluation of P at a point x chosen p-biasedly from $\{0, 1\}^m$, (where $p = (q-1)/q$). Thus for general q and $p = (q-1)/q$, taking the set of points given by Viola [Vio08] fooling low degree polynomials over $\mathbb{Z}_q$ under the uniform distribution over $\mathbb{Z}_q^m$, and then raising each coordinate to the power $q - 1$, yields the desired family of graphs $\mathcal{F}$. Then for general q and general p the same family of graphs works, arguing just as in the $q = 2$ case. ■

The analogue of the previous corollary for FO was proved in [GS71, BEH81] (see also [BR05, NNT05]).

The next corollary gives explicit functions which are hard on average for $\mathsf{FO}[\oplus]$. At present, we do not know how to extend it to $\mathsf{FO}[\mathsf{Mod}_q]$ for general q.

Corollary 3.2.7 (Explicit functions exponentially hard for $\mathsf{FO}[\oplus]$) *There is an explicit function $f : \{0,1\}^{\binom{n}{2}} \to \{0,1\}$ such that for every $\mathsf{FO}[\oplus]$-sentence φ,*

$$\Pr_{G \in G(n,p)}[(G \text{ satisfies } \varphi) \Leftrightarrow (f(A_G) = 1)] < \frac{1}{2} + 2^{-\Omega(n)}.$$

Proof Follows from Corollary 3.2.5, and the result of Babai, Nisan, Szegedy [BNS89] (for $p = 1/2$) and its generalization, Lemma 2.2.3 (for general p), constructing func-

[4]We are grateful to Salil Vadhan for pointing this out to us.

tions exponentially uncorrelated with low degree polynomials under the p-biased measure. It actually follows from our proofs that, one may even choose a function f that is a graph property (namely, invariant under the action of S_n on the coordinates). ■

3.3 The Distribution of Subgraph Frequencies mod q

In this section, we prove Theorem 3.2.4 on the distribution of subgraph frequencies in $G(n,p)$.

We first make a few definitions. If F is a connected graph and G is any graph, a *copy* of F in G is a set $E \subseteq E_G$ such that there exists an injective homomorphism χ from F to G such that $E = \chi(E_F) := \{(\chi(v), \chi(w)) \mid (v,w) \in E_F\}$. We denote the set of copies of F in G by $\mathsf{Cop}(F,G)$, the cardinality of $\mathsf{Cop}(F,G)$ by $\langle F \rangle(G)$, and this number mod q by $\langle F \rangle_q(G)$. We have the following basic relation (which follows from Lemma 3.6.5 in Section 3.6).

Lemma 3.3.1 *If F is a connected graph with $|E_F| \geq 1$, then*

$$[F](G) = \mathsf{aut}(F) \cdot \langle F \rangle(G).$$

For notational convenience, we view $G(n,p)$ as a graph whose vertex set is $[n]$ and whose edge set is a subset of $\binom{[n]}{2}$.

We can now state the general equidistribution theorem from which Theorem 3.2.4 will follow easily (We use the notation $\Omega_{q,p,d}(n)$ to denote the expression $\Omega(n)$, where the implied constant depends only on q, p and d). Note that this theorem holds for arbitrary integers q, not necessarily prime.

Theorem 3.3.2 (Equidistribution of graph copies) *Let $q > 1$ be an integer and let $p \in (0,1)$. Let $F_1, \ldots, F_\ell \in \mathsf{Conn}^a$ be distinct graphs with $1 \leq |E_{F_i}| \leq d$.*

Let $G \in G(n,p)$. Then the distribution of $(\langle F_1 \rangle_q(G), \ldots, \langle F_\ell \rangle_q(G))$ on $\mathbb{Z}_q^\ell$ is $2^{-\Omega_{q,p,d}(n)+\ell}$-close to uniform in statistical distance.

Using this theorem, we complete the proof of Theorem 3.2.4.

Proof of Theorem 3.2.4: Let $F_1, \ldots, F_\ell$ be an enumeration of the elements of Conn^a except for K_1. By Theorem 3.3.2, the distribution of $g = (\langle F_i \rangle_q(G))_{i=1}^{\ell}$ is $2^{-\Omega(n)}$ close to uniform over $\mathbb{Z}_q^\ell$. Given the vector g, we may compute the vector freq_G^a by:

- $(\mathsf{freq}_G^a)_{\mathrm{K}_1} = n \mod q$.
- For $F \in \mathsf{Conn}^a \setminus \{\mathrm{K}_1\}$, $(\mathsf{freq}_G^a)_F = g_F \cdot \mathsf{aut}(F)$ (by Lemma 3.3.1).

This implies that the distribution of freq_G^a is $2^{-\Omega(n)}$-close to uniformly distributed over $\{f \in \mathsf{FFreq}(a) : f_{\mathrm{K}_1} = n \mod q\}$. ■

The rest of this section is devoted to a proof of Theorem 3.3.2

Consider the special case $\ell = 1$, $F_1 = \mathrm{K}_3$ (the triangle), $q = 2$, $p = 1/2$ of Theorem 3.3.2. The theorem asserts that the distribution of $\langle F_1 \rangle_2(G)$ (for $G = G(n, 1/2)$) is $2^{-\Omega(n)}$-close to the uniform distribution over $\mathbb{Z}_2$. As described in the introduction, this reduces to showing that the polynomial $P_\triangle$ is unbiased on uniformly random inputs, where

$$P_\triangle((X_e)_{e \in \binom{[n]}{2}}) = \sum_{\{e_1,e_2,e_3\} \text{ forming a } \triangle} X_{e_1} X_{e_2} X_{e_3},$$

(recall that we view $G(n, 1/2)$ as having vertex set $[n]$).

We now sketch the proof in this special case. Let $r = \lfloor n/3 \rfloor$. Pick disjoint sets $V_1, \ldots, V_r \subseteq [n]$ with $|V_i| = 3$ for each i. Let $E_i = \binom{V_i}{2}$; E_i is the set of edges involved in the triangle formed by the vertices in V_i. Now for every $e \in \binom{[n]}{2} \setminus (\bigcup_i E_i)$, let us fix X_e to an arbitrary value in $\mathbb{Z}_2$. After this fixing, the polynomial $P_\triangle$ becomes a polynomial only in the variables $\{X_e \mid e \in \bigcup_i E_i\}$. Closer inspection reveals that this polynomial is of the form:

$$\sum_{i=1}^{r} X_{3i-2} X_{3i-1} X_{3i} + R(X),$$

where R is a polynomial of degree at most 2. At this point we invoke an elegant result of Babai-Nisan-Szegedy [BNS89], originally discovered in the context of com-

munication complexity, which asserts that polynomials of the above kind (where R is an arbitrary polynomial of degree at most 2), take the values $0, 1$ with roughly equal probability ($\approx 1/2$). Finally, since this unbiasedness occurs for an arbitrary fixing of the variables $\{X_e \mid e \in \binom{[n]}{2} \setminus (\bigcup_i E_i)\}$, it follows that this unbiasedness also holds for the original polynomial $P_\triangle$.

A virtually identical argument shows the unbiasedness of the number of copies mod 2 of any other connected graph. Another very similar argument shows the unbiasedness of $\sum\langle F_i\rangle_2(G)$ for any collection of distinct connected graphs F_i. Combining these unbiasedness results yields the full joint equidistribution result of Theorem 3.3.2.

We now proceed with the details.

3.3.1 Preliminary lemmas

The following lemma, which is used in the proof of Theorem 3.3.2 (and again in Section 3.8 to study the distribution of labelled subgraph frequencies), gives a simple sufficient criterion for the distribution of values of a polynomial to be "unbiased". The proof appears in Section 3.4.

Lemma 3.3.3 *Let $q > 1$ be an integer and let $p \in (0, 1)$. Let[5] $\mathcal{F} \subseteq 2^{[m]}$. Let $d > 0$ be an integer. Let $Q(Z_1, \ldots, Z_m) \in \mathbb{Z}_q[Z_1, \ldots, Z_m]$ be a polynomial of the form*

$$\sum_{S \in \mathcal{F}} a_S \prod_{i \in S} Z_i + Q'(\mathbf{Z}),$$

where $\deg(Q') < d$. Suppose there exist $\mathcal{E} = \{E_1, \ldots, E_r\} \subseteq \mathcal{F}$ such that:

- *$|E_j| = d$ for each j,*
- *$a_{E_j} \neq 0$ for each j.*
- *$E_j \cap E_{j'} = \emptyset$ for each j, j',*
- *For each $S \in \mathcal{F} \setminus \mathcal{E}$, $|S \cap (\cup_j E_j)| < d$.*

[5] If S is a set, we use the notation 2^S to denote its power set.

Let $\mathbf{z} = (z_1, \ldots, z_m) \in \mathbb{Z}_q^m$ *be the random variable where, independently for each* i*, we have* $\Pr[z_i = 1] = p$ *and* $\Pr[z_i = 0] = 1 - p$*. Then,*

$$\left|\mathbb{E}\left[\omega^{Q(\mathbf{z})}\right]\right| \leq 2^{-\Omega_{q,p,d}(r)},$$

where $\omega \in \mathbb{C}$ *is a primitive* q^{th}*-root of unity.*

The lemma below is a useful tool for showing that a distribution on $\mathbb{Z}_q^\ell$ is close to uniform.

Lemma 3.3.4 (Vazirani XOR lemma) *Let* $q > 1$ *be an integer and let* $\omega \in \mathbb{C}$ *be a primitive* q^{th}*-root of unity. Let* $\mathbf{X} = (X_1, \ldots, X_\ell)$ *be a random variable over* $\mathbb{Z}_q^\ell$*. Suppose that for every nonzero* $c \in \mathbb{Z}_q^\ell$*,*

$$\left|\mathbb{E}\left[\omega^{\sum_{i\in[\ell]} c_i X_i}\right]\right| \leq \epsilon.$$

Then $\mathbf{X}$ *is* $q^\ell \cdot \epsilon$*-close to uniformly distributed over* $\mathbb{Z}_q^\ell$*.*

3.3.2 Proof of the equidistribution theorem

Proof of Theorem 3.3.2: By the Vazirani XOR Lemma (Lemma 3.3.4), it suffices to show that for each nonzero $c \in \mathbb{Z}_q^\ell$, we have $\left|\mathbb{E}\left[\omega^R\right]\right| \leq 2^{-\Omega_{q,p,d(n)}}$, where $R := \sum_{i\in[\ell]} c_i \langle F_i \rangle_q(G)$, and $\omega \in \mathbb{C}$ is a primitive q^{th}-root of unity.

We will show this by appealing to Lemma 3.3.3. Let $m = \binom{n}{2}$. Let $\mathbf{z} \in \{0,1\}^{\binom{[n]}{2}}$ be the random variable where, for each $e \in \binom{[n]}{2}$, $z_e = 1$ if and only if e is present in G. Thus, independently for each e, $\Pr[z_e = 1] = p$.

We may now express R in terms of the z_e. Let K_n denote the complete graph on the vertex set $[n]$. Thus $\mathsf{Cop}(F_i, \mathrm{K}_n)$ is the set of E that could potentially arise as copies of F_i in G. Then we may write,

$$\begin{aligned} R = \sum_{i\in[\ell]} c_i \langle F_i \rangle_q(G) &= \sum_{i\in[\ell]} c_i \sum_{E\in\mathsf{Cop}(F_i,\mathrm{K}_n)} \prod_{e\in E} z_e \\ &= \sum_{E\in\mathcal{F}} c_E \prod_{e\in E} z_e, \end{aligned}$$

where $\mathcal{F} \subseteq 2^{\binom{[n]}{2}}$ is the set $\bigcup_{i:c_i \neq 0} \mathsf{Cop}(F_i, \mathrm{K}_n)$, and for $E \in \mathcal{F}$, $c_E = c_i$ for the unique i satisfying $E \in \mathsf{Cop}(F_i, \mathrm{K}_n)$ (note that since the F_i are nonisomorphic connected graphs, the $\mathsf{Cop}(F_i, \mathrm{K}_n)$ are pairwise disjoint).

Let $Q(\mathbf{Z}) \in \mathbb{Z}_q[\mathbf{Z}]$, where $\mathbf{Z} = (Z_e)_{e \in \binom{[n]}{2}}$ be the polynomial $\sum_{E \in \mathcal{F}} c_E \prod_{e \in E} Z_e$. Then $R = Q(\mathbf{z})$. We wish to show that

$$\left|\mathbb{E}\left[\omega^{Q(\mathbf{z})}\right]\right| \leq 2^{-\Omega_{q,p,d}(n)}. \tag{3.1}$$

We do this by demonstrating that the polynomial $Q(\mathbf{Z})$ satisfies the hypotheses of Lemma 3.3.3.

Let $d^* = \max_{i:c_i \neq 0} |E_{F_i}|$. Let $i_0 \in [\ell]$ be such that $c_{i_0} \neq 0$ and $|E_{F_{i_0}}| = d^*$. Let $\chi_1, \chi_2, \ldots, \chi_r \in \mathsf{Inj}(F_{i_0}, \mathrm{K}_n)$ be a collection of homomorphisms such that for all distinct $j, j' \in [r]$, we have $\chi_j(V_{F_{i_0}}) \cap \chi_{j'}(V_{F_{i_0}}) = \emptyset$. Such a collection can be chosen greedily so that $r = \Omega(\frac{n}{d})$. Let $E_j \in \mathsf{Cop}(F_{i_0}, \mathrm{K}_n)$ be given by $\chi_j(E_{F_{i_0}})$. Let $\mathcal{E}$ be the family of sets $\{E_1, \ldots, E_r\} \subseteq \mathcal{F}$. We observe the following properties of the E_j:

1. For each $j \in [r]$, $|E_j| = d^*$ (since χ_j is injective).

2. For each $j \in [r]$, $c_{E_j} = c_{i_0} \neq 0$.

3. For distinct $j, j' \in [r]$, $E_j \cap E_{j'} = \emptyset$ (by choice of the χ_j).

4. For every $S \in \mathcal{F} \setminus \mathcal{E}$, $|S \cap (\cup_j E_j)| < d^*$. To see this, take any $S \in \mathcal{F} \setminus \mathcal{E}$ and suppose $|S \cap (\cup_j E_j)| \geq d^*$. Let $i' \in [\ell]$ be such that $c_{i'} \neq 0$ and $S \in \mathsf{Cop}(F_{i'}, \mathrm{K}_n)$. Let $\chi \in \mathsf{Inj}(F_{i'}, \mathrm{K}_n)$ with $\chi(E_{F_{i'}}) = S$. By choice of d^*, we know that $|S| \leq d^*$. Therefore, the only way that $|S \cap (\cup_j E_j)|$ can be $\geq d^*$ is if (1) $|S| = d^*$, and (2) $S \cap (\cup_j E_j) = S$, or in other words, $S \subseteq (\cup_j E_j)$. However, since the $\chi_j(V_{F_{i_0}})$ are all pairwise disjoint, this implies that $S \subseteq E_j$ for some j. But since $|E_j| = |S|$, we have $S = E_j$, contradicting our choice of S. Therefore, $|S \cap (\cup_j E_j)| < d^*$ for any $S \in \mathcal{F} \setminus \mathcal{E}$.

It now follows that $Q(\mathbf{Z})$, $\mathcal{F}$ and $\mathcal{E}$ satisfy the hypothesis of Lemma 3.3.3. Consequently, (recalling that $r = \Omega(n/d)$ and $d^* \leq d$) Equation (3.1) follows, completing the proof of the theorem. ■

Remark We just determined the joint distribution of the number of injective homomorphisms, mod q, from all small connected graphs to $G(n,p)$. This information can be used in conjunction with Lemma 3.6.2 to determine the joint distribution of the number of injective homomorphisms, mod q, from *all* small graphs to $G(n,p)$.

Many intriguing basic questions about the distribution of subgraph frequencies mod q remain. For example, it would be interesting to determine whether the statistical distance $2^{-\Omega(n)}$ in Theorem 3.3.2 can be replaced by $2^{-\Omega(n^2)}$. It would also be interesting to know what happens in the graph $G(n, n^{-\alpha})$, where some constant size graphs may not appear as subgraphs even once.

3.4 A criterion for unbiasedness

Our main goal in this section is to give a full proof of Lemma 3.3.3, which gives a criterion for a polynomial to be unbiased.

Our proof of Lemma 3.3.3 will go through Lemma 2.2.3, (which was proved in Section 2.2.1).

Lemma 2.2.3 (restated) *Let $q > 1$ be an integer and let $p \in (0,1)$. Let $E_1, \ldots, E_r$ be pairwise disjoint subsets of $[m]$ each of cardinality d. Let $Q(Z_1, \ldots, Z_m) \in \mathbb{Z}_q[Z_1, \ldots, Z_m]$ be a polynomial of the form*

$$\left(\sum_{j=1}^{r} a_j \prod_{i \in E_j} Z_i \right) + R(\mathbf{Z}),$$

where each $a_j \neq 0$ and $\deg(R(\mathbf{Z})) < d$. Let $\mathbf{z} = (z_1, \ldots, z_m) \in \mathbb{Z}_q^m$ be the random variable where, independently for each i, we have $\Pr[z_i = 1] = p$ and $\Pr[z_i = 0] = 1-p$. Then,

$$\left| \mathbb{E}\left[\omega^{Q(\mathbf{z})}\right] \right| \leq 2^{-\Omega_{q,p,d}(r)}.$$

Given Lemma 2.2.3, we may now prove Lemma 3.3.3.

Proof of Lemma 3.3.3: Let $U = \cup_{j=1}^{r} E_j$. Fix any $x \in \{0,1\}^{[m]\setminus U}$, and let $Q_x(\mathbf{Y}) \in$

$\mathbb{Z}_q[(Y_i)_{i\in U}]$ be the polynomial

$$\sum_{S\in\mathcal{F}} a_S \left(\prod_{j\in S\cap([m]\setminus U)} x_j\right)\left(\prod_{i\in S\cap U} Y_i\right) + Q'(x,\mathbf{Y})$$

so that $Q_x(y) = Q(x,y)$ for each $y \in \mathbb{Z}_q^U$. Notice that the degree (in $\mathbf{Y}$) of the term corresponding to $S \in \mathcal{F}$ is $|S \cap U|$. By assumption, unless $S = E_j$ for some j, we must have $|S \cap U| < d$.

Therefore the polynomial $Q_x(\mathbf{Y})$ is of the form:

$$\sum_{j=1}^{r} a_{E_j} \prod_{i\in E_j} Y_i + R(\mathbf{Y}),$$

where $\deg(R(\mathbf{Y})) < d$. By Lemma 2.2.3,

$$\left|\mathbb{E}\left[\omega^{Q_x(\mathbf{y})}\right]\right| < 2^{-\Omega_{q,p,d}(r)},$$

where $\mathbf{y} \in \{0,1\}^U$ with each $y_i = 1$ independently with probability p.

As $Q_x(y) = Q(x,y)$, we get

$$\left|\mathbb{E}\left[\omega^{Q(\mathbf{z}^x)}\right]\right| < 2^{-\Omega_{q,p,d}(r)},$$

where $\mathbf{z}^x \in \mathbb{Z}_q^n$ is the random variable $\mathbf{z}$ conditioned on the event $z_j = x_j$ for every $j \in [m] \setminus U$. Now, the distribution of $\mathbf{z}$ is a convex combination of the distributions of $\mathbf{z}^x$ as x varies over $\{0,1\}^{[m]\setminus U}$. This allows us to deduce that

$$\left|\mathbb{E}\left[\omega^{Q(\mathbf{z})}\right]\right| \leq 2^{-\Omega_{q,p,d}(r)},$$

as desired. ■

3.5 Outline of the Proof

Now that we have understood the distribution of subgraph frequencies mod q, we now approach the main part of the proof of the modular convergence law, Theorem 3.2.3, which relates $\mathsf{FO}[\mathsf{Mod}_q]$ sentences to subgraph frequencies mod q.

The proof of Theorem 3.2.3, will be via a more general theorem amenable to inductive proof, Theorem 3.5.8. Just as Theorem 3.2.3 states that for almost all $G \in G(n,p)$, the truth of any $\mathsf{FO}[\mathsf{Mod}_q]$ sentence on G is determined by subgraph frequencies, freq^c_G, Theorem 3.5.8 states that for almost all graphs $G \in G(n,p)$, for any $w_1, \ldots, w_k \in V_G$ the truth of any $\mathsf{FO}[\mathsf{Mod}_q]$ *formula* $\varphi(w_1, \ldots, w_k)$ on G is determined by (1) the internal adjacency and equality information about $w_1, \ldots, w_k$ (which we will call the *type*), and (2) the *subgraph frequencies of labelled graphs rooted at* $\mathbf{w}$. In the next subsection, we formalize these notions.

3.5.1 Labelled graphs and labelled subgraph frequencies

Let I be a finite set. We begin with some preliminaries on I-labelled graphs.

Definition 3.5.1 (I-labelled graphs) *An* I-labelled graph *is a graph* $F = (V_F, E_F)$ *where some vertices are labelled by elements of* I*, such that (a) for each* $i \in I$*, there is exactly one vertex labelled* i*. We denote this vertex* $F(i)$*, and (b) the graph induced on the set of labelled vertices is an independent set. We denote the set of labelled vertices of* F *by* $\mathcal{L}(F)$*.*

Definition 3.5.2 (Homomorphisms and Copies) *A* homomorphism *from an* I*-labelled graph* F *to a pair* $(G, \mathbf{w})$*, where* G *is a graph and* $\mathbf{w} \in V_G^I$*, is a homomorphism* $\chi \in \mathsf{Hom}(F, G)$ *such that for each* $i \in I$*,* χ *maps* $F(i)$ *to* w_i*. A homomorphism from* F *to* $(G, \mathbf{w})$ *is called* injective *if for any distinct* $v, w \in V_F$*, such that* $\{v, w\} \not\subseteq \mathcal{L}(F)$*, we have* $\chi(v) \neq \chi(w)$*. A* copy *of* F *in* $(G, \mathbf{w})$ *is a set* $E \subseteq E_G$ *such that there exists an injective homomorphism* χ *from* F *to* $(G, \mathbf{w})$ *such that* $E = \chi(E_F) := \{(\chi(v), \chi(w)) \mid (v, w) \in E_F\}$*. An* automorphism *of* F *is an injective homomorphism from* F *to* $(F, \mathbf{w})$*, where* $w_i = F(i)$ *for each* $i \in I$*.*

Definition 3.5.3 (Hom, Inj, Cop, Aut **for labelled graphs**) *Let F be an I-labelled graph, and G be any graph. Let $\mathbf{w} \in V_G^I$. We define* $\mathsf{Hom}(F,(G,\mathbf{w}))$ *to be the set of homomorphisms from F to $(G,\mathbf{w})$. We define* $\mathsf{Inj}(F,(G,\mathbf{w}))$ *to be the set of injective homomorphisms from F to $(G,\mathbf{w})$. We define* $\mathsf{Cop}(F,(G,\mathbf{w}))$ *to be the set of copies of F in $(G,\mathbf{w})$. We define* $\mathsf{Aut}(F)$ *to be the set of automorphisms of F. We let $[F](G,\mathbf{w})$ (respectively $\langle F\rangle(G,\mathbf{w})$, $\mathsf{aut}(F)$) be the cardinality of* $\mathsf{Inj}(F,(G,\mathbf{w}))$ (respectively $\mathsf{Cop}(F,(G,\mathbf{w}))$, $\mathsf{Aut}(F)$).

Finally, let $[F]_q(G,\mathbf{w}) = [F](G,\mathbf{w}) \mod q$ and $\langle F\rangle_q(G,\mathbf{w}) = \langle F\rangle(G,\mathbf{w}) \mod q$.

Definition 3.5.4 (Label-connected) *For F an I-labelled graph, we say F is* label-connected *if $F \setminus \mathcal{L}(F)$ is connected. Define Conn_I^t to be the set of all I-labelled label-connected graphs with at most t unlabelled vertices. For $i \in I$, we say an I-labelled graph F is* **dependent on label** *i if $F(i)$ is not an isolated vertex.*

Definition 3.5.5 (Partitions) *If I is a set, an* I-partition *is a set of subsets of I that are pairwise disjoint, and whose union is I. If Π is an I partition, then for $i \in I$ we denote the unique element of Π containing i by $\Pi(i)$. If V is any set and $\mathbf{w} \in V^I$, we say $\mathbf{w}$ respects Π if for all $i,i' \in I$, $w_i = w_{i'}$ iff $\Pi(i) = \Pi(i')$.*

The collection of all partitions of I is denoted $\mathsf{Partitions}(I)$.

If $I \subseteq J$, $\Pi \in \mathsf{Partitions}(I)$ and $\Pi' \in \mathsf{Partitions}(J)$, we say Π' extends Π if for all $i_1, i_2 \in I$, $\Pi(i_1) = \Pi(i_2)$ if and only if $\Pi'(i_1) = \Pi'(i_2)$.

Definition 3.5.6 (Types) *An I-*type *τ is a pair (Π_τ, E_τ) where $\Pi_\tau \in \mathsf{Partitions}(I)$ and $E_\tau \subseteq \binom{\Pi_\tau}{2}$. For a graph G and $\mathbf{w} \in V_G^I$, we define the* type *of $\mathbf{w}$ in G, denoted $\mathsf{type}_G(\mathbf{w})$, to be the I-*type *τ, where $\mathbf{w}$ respects Π_τ, and for all $i,i' \in I$, $\{\Pi_\tau(i), \Pi_\tau(i')\} \in E_\tau$ if and only if w_i and $w_{i'}$ are adjacent in G.*

The collection of all I-types is denoted $\mathsf{Types}(I)$.

If $I \subseteq J$, and $\tau \in \mathsf{Types}(I)$ and $\tau' \in \mathsf{Types}(J)$, we say τ' *extends* τ if $\Pi_{\tau'}$ extends Π_τ and for each $i_1, i_2 \in I$, $\{\Pi_\tau(i_1), \Pi_\tau(i_2)\} \in E_\tau$ if and only if $\{\Pi_{\tau'}(i_1), \Pi_{\tau'}(i_2)\} \in E_{\tau'}$.

Definition 3.5.7 (Labelled subgraph frequency vector) *Let G be a graph and I be any set. Let $\mathbf{w} \in V_G^I$. We define the* labelled subgraph frequency vector at $\mathbf{w}$, $\mathsf{freq}_G^a(\mathbf{w}) \in \mathbb{Z}_q^{\mathsf{Conn}_I^a}$, *to be the vector such that for each $F \in \mathsf{Conn}_I^a$,*

$$(\mathsf{freq}_G^a(\mathbf{w}))_F = [F]_q(G, \mathbf{w}).$$

Remark We will often deal with $[k]$-labelled graphs. By abuse of notation we will refer to them as k-labelled graphs. If $\mathbf{w} \in V^{[k]}$ and $u \in V$, when we refer to the tuple $(\mathbf{w}, v)$, we mean the $[k+1]$-tuple whose first k coordinates are given by $\mathbf{w}$ and whose $k+1$st coordinate is v. Abusing notation even further, when we deal with a $[k+1]$-labelled graph F, then by $[F](G, \mathbf{w}, v)$, we mean $[F](G, (\mathbf{w}, v))$. Similarly Conn_k^t denotes $\mathsf{Conn}_{[k]}^t$.

3.5.2 The quantifier eliminating theorem

We now state Theorem 3.5.8, from which Theorem 3.2.3 follows easily. Informally, it says that an $\mathsf{FO}[\mathsf{Mod}_q]$-formula $\varphi(\mathbf{w})$ is essentially determined by the type of $\mathbf{w}$, $\mathsf{type}_G(\mathbf{w})$, and the labelled subgraph frequencies at $\mathbf{w}$, $\mathsf{freq}_G^c(\mathbf{w})$.

Theorem 3.5.8 *For all primes q and integers $k, t > 0$, there is a constant $c = c(k, t, q)$ such that for every $\mathsf{FO}[\mathsf{Mod}_q]$ formula $\varphi(\alpha_1, \ldots, \alpha_k)$ with quantifier depth t, there is a function $\psi : \mathsf{Types}(k) \times \mathbb{Z}_q^{\mathsf{Conn}_k^c} \to \{0, 1\}$ such that for all $p \in (0, 1)$, the quantity*

$$\Pr_{G \in G(n,p)} \left[\begin{array}{l} \forall w_1, \ldots, w_k \in V_G, \\ \quad (G \text{ satisfies } \varphi(w_1, \ldots, w_k)) \Leftrightarrow (\psi(\mathsf{type}_G(\mathbf{w}), \mathsf{freq}_G^c(\mathbf{w})) = 1) \end{array} \right] \geq 1 - 2^{-\Omega(n)}.$$

Putting $k = 0$, we recover Theorem 3.2.3.

We now give a brief sketch of the proof of Theorem 3.5.8 (the detailed proof appears in Section 3.6). The proof is by induction on the size of the formula φ. When the formula φ has no quantifiers, then the truth of $\varphi(\mathbf{w})$ on G is completely determined by

$\mathsf{type}_G(\mathbf{w})$. The case where φ is of the form $\varphi_1(\alpha_1, \ldots, \alpha_k) \wedge \varphi_2(\alpha_1, \ldots, \alpha_k)$ is easily handled via the induction hypothesis. The case where $\varphi(\alpha_1, \ldots, \alpha_k) = \neg\varphi_1(\alpha_1, \ldots, \alpha_k)$ is similar.

The key cases for us to handle are thus *(i)* $\varphi(\alpha_1, \ldots, \alpha_k)$ is of the form $\mathsf{Mod}_q^i\beta, \varphi'(\alpha_1, \ldots, \alpha_k, \beta)$, and *(ii)* $\varphi(\alpha_1, \ldots, \alpha_k)$ is of the form $\exists\beta, \varphi'(\alpha_1, \ldots, \alpha_k, \beta)$. We now give a sketch of how these cases may be handled.

For case (i), let $\psi' : \mathsf{Types}(k+1) \times \mathbb{Z}_q^{\mathsf{Conn}_{k+1}^b}$ be the function given by the induction hypothesis for the formula φ'. Thus for most graphs $G \in G(n,p)$ (namely the ones for which ψ' is good for φ'), $\varphi(w_1, \ldots, w_k)$ is true if and only the number of vertices $v \in V_G$ such that $\psi'(\mathsf{type}_G(\mathbf{w}, v), \mathsf{freq}_G^b(\mathbf{w}, v)) = 1$ is congruent to i mod q. In Theorem 3.6.1 (whose proof appears in Section 3.7), we show that the number of such vertices v can be determined solely as a function of $\mathsf{type}_G(\mathbf{w})$ and $\mathsf{freq}_G^a(\mathbf{w})$ for suitable a. This fact allows us to define ψ in a natural way, and this completes case (i).

Case (ii) is the most technically involved case. As before, we get a function ψ' corresponding to φ' by the induction hypothesis. We show that one can define ψ essentially as follows: define $\psi(\tau, f) = 1$ if there exists some $(\tau', f') \in \mathsf{Types}(k+1) \times \mathbb{Z}_q^{\mathsf{Conn}_{k+1}^b}$ that "extends" (τ, f) for which $\psi'(\tau', f') = 1$; otherwise $\psi(\tau, f) = 0$. Informally, we show that if it is conceivable that there is a vertex v such that $\varphi'(\mathbf{w}, v)$ is true, then $\varphi(\mathbf{w})$ is almost surely true. Proving this statement requires us to get a characterization of the distribution of labelled subgraph frequencies, significantly generalizing Theorem 3.2.4. This is done in Theorem 3.6.12 (whose proof appears in Section 3.8).

3.6 Quantifier Elimination

In this section, we give a full proof of Theorem 3.5.8. Before doing so, we state the main technical theorems: Theorem 3.6.1 (which is needed for eliminating Mod_q quantifiers), and Theorem 3.6.12 (which is needed for eliminating $\exists$ quantifiers). We do this in the following two subsections.

3.6.1 Counting extensions

The next theorem plays a crucial role in the elimination of the Mod_q quantifiers. This is the only step where the assumption that q is a prime plays a role in the modular convergence law.

Theorem 3.6.1 *Let q be a prime, let $k, b > 0$ be integers and let $a \geq (q-1) \cdot b \cdot |\mathsf{Conn}^b_{k+1}| + 1$. There is a function*

$$\lambda : \mathsf{Types}(k+1) \times \mathbb{Z}_q^{\mathsf{Conn}^b_{k+1}} \times \mathsf{Types}(k) \times \mathbb{Z}_q^{\mathsf{Conn}^a_k} \to \mathbb{Z}_q$$

such that for all $\tau' \in \mathsf{Types}(k+1)$, $f' \in \mathbb{Z}_q^{\mathsf{Conn}^b_{k+1}}$, $\tau \in \mathsf{Types}(k)$, $f \in \mathbb{Z}_q^{\mathsf{Conn}^a_k}$, it holds that for every graph G, and every $w_1, \ldots, w_k \in V_G$ with $\mathsf{type}_G(\mathbf{w}) = \tau$ and $\mathsf{freq}^a_G(\mathbf{w}) = f$, the cardinality of the set

$$\{v \in V_G : \mathsf{type}_G(\mathbf{w}, v) = \tau' \wedge \mathsf{freq}^b_G(\mathbf{w}, v) = f'\}$$

is congruent to $\lambda(\tau', f', \tau, f) \mod q$.

The proof appears in Section 3.7. The principal ingredient in its proof is the following lemma, which states that the numbers $[F](G, \mathbf{w})$, as F varies over small label-connected graphs, determine the number $[F'](G, \mathbf{w})$ for all small graphs F'.

Lemma 3.6.2 (Label-connected subgraph frequencies determine all subgraph frequencies) *For every k-labelled graph F' with $|V_{F'} \setminus \mathcal{L}(F')| \leq t$, there is a polynomial $\delta_{F'} \in \mathbb{Z}[(X_F)_{F \in \mathsf{Conn}^t_k}]$ such that for all graphs G and $\mathbf{w} \in V_G^k$,*

$$[F'](G, \mathbf{w}) = \delta_{F'}(x),$$

where $x \in \mathbb{Z}^{\mathsf{Conn}^t_k}$ is given by $x_F = [F](G, \mathbf{w})$.

3.6.2 The distribution of labelled subgraph frequencies mod q

In this subsection, we state the theorem that will help us eliminate $\exists$ quantifiers. Let us first give an informal description of the theorem. We are given a tuple $\mathbf{w} \in [n]^k$, and distinct $u_1, \ldots, u_s \in [n] \setminus \{w_1, \ldots, w_k\}$. Let G be sampled from $G(n,p)$ (recall that we think of $G(n,p)$ as a random graph whose vertex set is $[n]$: thus the w_i and u_j are vertices of G). The theorem completely describes the joint distribution of the labelled subgraph frequency vectors at all the tuples $\mathbf{w}$, $(\mathbf{w}, u_1), \ldots, (\mathbf{w}, u_s)$; namely it pins down the distribution of $(\mathsf{freq}^a_G(\mathbf{w}), \mathsf{freq}^b_G(\mathbf{w}, u_1), \ldots, \mathsf{freq}^b_G(\mathbf{w}, u_s))$. We first give a suitable definition of the set of *feasible frequency vectors*, and then claim that (a) the $\mathsf{freq}^a_G(\mathbf{w})$ is essentially uniformly distributed over the set of its feasible frequency vectors, and (b) conditioned on $\mathsf{freq}^a_G(\mathbf{w})$, the distributions of $\mathsf{freq}^b_G(\mathbf{w}, u_1), \ldots, \mathsf{freq}^b_G(\mathbf{w}, u_s)$ are all essentially independent and uniformly distributed over the set of those feasible frequency vectors that are "consistent" with freq^a_G.

To define the set of feasible frequency vectors (which will equal the set of all possible values that $\mathsf{freq}^a_G(\mathbf{w})$ may assume), there are two factors that come into play. The first factor, one that we already encountered while dealing with unlabelled graphs, is a divisibility constraint: the number $[F](G, \mathbf{w})$ is always divisible by a certain integer depending on F, and hence for some F, it cannot assume arbitrary values mod q. The second factor is a bit subtler: when $w_1, \ldots, w_k$ are not all distinct, for certain pairs F, F' of label-connected k-labelled graphs, $[F](G, \mathbf{w})$ is forced to equal $[F'](G, \mathbf{w})$. Let us see a simple example of such a phenomenon. Let $k = 2$ and let $w_1 = w_2$. Let the 2-labelled graph F be a path of length 2 with ends labelled 1 and 2. Let the 2-labelled graph F' be the disjoint union of an edge, one of whose ends is labelled 1, and an isolated vertex labelled 2. Then in any graph G, $[F](G, \mathbf{w}) = [F'](G, \mathbf{w}) =$ the degree of w_1.

In the rest of this subsection, we will build up some notation and results leading up to a definition of feasible frequency vectors and the statement of the main technical theorem describing the distribution of labelled subgraph frequency vectors.

Definition 3.6.3 (Quotient of a labelled graph by a paritition) *Let F be a I-labelled graph and let $\Pi \in \mathsf{Partitions}(I)$. We define F/Π to be the Π-labelled graph obtained from F by (a) for each $J \in \Pi$, identifying all the vertices with labels in J and labelling this new vertex J, and (b) deleting duplicate edges. If F and F' are I-labelled graphs and $\Pi \in \mathsf{Partitions}(I)$, we say F and F' are Π-equivalent if $F/\Pi \cong F'/\Pi$.*

Let $\mathbf{w} \in V_G^I$. Let $\Pi \in \mathsf{Partitions}(I)$ be such that $\mathbf{w}$ respects Π. Define $(\mathbf{w}/\Pi) \in V_G^\Pi$ by: for each $J \in \Pi$, $(\mathbf{w}/\Pi)_J = w_j$, where j is any element of J (this definition is independent of the choice of $j \in J$). Observe that as J varies over Π, the vertices $(w/\Pi)_J$ are all distinct.

The next two lemmas show that the numbers $[F](G, \mathbf{w})$ must satisfy certain constraints. These constraints will eventually motivate our definition of feasible frequency vectors.

Lemma 3.6.4 *If G is a graph and $\mathbf{w} \in V_G^I$, with $\mathbf{w}$ respecting $\Pi \in \mathsf{Partitions}(I)$, then for any I-labelled F,*

$$[F](G, \mathbf{w}) = [F/\Pi](G, (\mathbf{w}/\Pi)). \tag{3.2}$$

Proof We define a bijection $\alpha : \mathsf{Inj}(F/\Pi, (G, \mathbf{w}/\Pi)) \to \mathsf{Inj}(F, (G, \mathbf{w}))$. Let $\pi \in \mathsf{Hom}(F, F/\Pi)$ be the natural homomorphism sending each unlabelled vertex in V_F to its corresponding vertex in $V_{F/\Pi}$, and, for each $i \in I$ sending $F(i)$ to $(F/\Pi)(\Pi(i))$. We define $\alpha(\chi)$ to be $\chi \circ \pi$.

Take distinct $\chi, \chi' \in \mathsf{Inj}(F/\Pi, (G, \mathbf{w}/\Pi))$. Let $u \in V_{F/\Pi}$ with $\chi(u) \neq \chi'(u)$. Note that u cannot be an element of $\mathcal{L}(F/\Pi)$, for if $u = (F/\Pi)(\Pi(i))$, then $\chi(u) = \chi'(u) = w_i$. Thus $u \notin \mathcal{L}(F/\Pi)$. Let $v \in V_F$ be the vertex $\pi^{-1}(u)$ (which is uniquely specified since $u \notin \mathcal{L}(F/\Pi)$). Thus we have $\chi(\pi(v)) = \chi(u) \neq \chi'(u) = \chi'(\pi(v))$. Thus $\alpha(\chi) \neq \alpha(\chi')$, and α is one-to-one.

To show that α is onto, take any $\chi \in \mathsf{Inj}(F, (G, \mathbf{w}))$. Define $\chi' \in \mathsf{Inj}(F/\Pi, (G, \mathbf{w}/\Pi))$ by:

1. $\chi'(u) = \chi(\pi^{-1}(u))$ if $u \notin \mathcal{L}(F/\Pi)$.

2. $\chi'(u) = w_j$ for any $j \in J$, if $u = (F/\Pi)(J)$ with $J \in \Pi$.

Then $\alpha(\chi') = \chi$. ■

Lemma 3.6.5 *Let G be a graph and $\mathbf{w} \in V_G^I$. Suppose all the $(w_i)_{i \in I}$ are distinct. Let F be an I-labelled label-connected graph with $|E_F| \geq 1$. Then*

$$[F](G, \mathbf{w}) = \mathsf{aut}(F) \cdot \langle F \rangle (G, \mathbf{w}).$$

Proof We give a bijection $\alpha : \mathsf{Aut}(F) \times \mathsf{Cop}(F, (G, \mathbf{w})) \to \mathsf{Inj}(F, (G, \mathbf{w}))$.

For each $E \in \mathsf{Cop}(F, (G, \mathbf{w}))$, we fix a $\chi_E \in \mathsf{Inj}(F, (G, \mathbf{w}))$ such that $\chi_E(E_F) = E$. Then we define $\alpha(\sigma, E) = \chi_E \circ \sigma$.

First notice that $\alpha(\sigma, E)(E_F) = \chi_E(\sigma(E_F)) = \chi_E(E_F) = E$. Thus if $\alpha(\sigma, E) = \alpha(\sigma', E')$, then $E = E'$. But since χ_E is injective, for any $\sigma \neq \sigma'$, we have $\chi_E \circ \sigma \neq \chi_E \circ \sigma'$. Thus α is one-to-one.

To show that α is onto, take any $\chi \in \mathsf{Inj}(F, (G, \mathbf{w}))$. Let $E = \chi(E_F)$. As F is label-connected and $\chi_E(E_F) = \chi(E_F)$, we have $\chi_E(V_F) = \chi(V_F)$. We may now define $\sigma \in \mathsf{Aut}(F)$ by $\sigma(u) = \chi_E^{-1}(\chi(u))$ for each $u \in V_F$. Clearly, $\alpha(\sigma, E) = \chi$, and so α is onto.

Thus α is a bijection, and the lemma follows. ■

Note that Lemma 3.2.2 and Lemma 3.3.1 follow formally from the above lemma.

Let $\mathrm{K}_1(I)$ be the I-labelled graph with $|I| + 1$ vertices: $|I|$ labelled vertices and one isolated unlabelled vertex. The role of $\mathrm{K}_1(I)$ in the I-labelled theory is similar to the role of K_1 in the unlabelled case.

Definition 3.6.6 (Feasible frequency vectors) *We define the set of* feasible frequency vectors, $\mathsf{FFreq}(\tau, I, a)$ *to be the set of $f \in \mathbb{Z}_q^{\mathsf{Conn}_I^a}$ such that*

(a) for any $F \in \mathsf{Conn}_I^a$, we have $f_F \in \mathsf{aut}(F/\Pi_\tau) \cdot \mathbb{Z}_q$.

(b) for any $F, F' \in \mathsf{Conn}_I^a$ that are Π_τ-equivalent, we have $f_F = f_{F'}$.

Let $\mathsf{FFreq}_n(\tau, I, a)$ *be the set* $\{f \in \mathsf{FFreq}(\tau, I, a) : f_{K_1(I)} = n - |\Pi_\tau| \mod q\}$. *Note that if* $n = n' \mod q$, *then* $\mathsf{FFreq}_n(\tau, I, a) = \mathsf{FFreq}_{n'}(\tau, I, a)$.

Observe that for any $\mathbf{w} \in V_G^I$ with $\mathsf{type}_G(\mathbf{w}) = \tau$, the vector $\mathsf{freq}_G^a(\mathbf{w})$ is an element of $\mathsf{FFreq}(\tau, I, a)$. This follows from Lemma 3.6.4 and Lemma 3.6.5, which allow us to deduce (recall that $(\mathbf{w}/\Pi_\tau)_J$ are all distinct for $J \in \Pi_\tau$) that for any $F \in \mathsf{Conn}_I^a$,

$$[F](G, \mathbf{w}) = \mathsf{aut}(F/\Pi_\tau) \cdot \langle F/\Pi_\tau \rangle (G, \mathbf{w}/\Pi_\tau). \tag{3.3}$$

Observe also that if $|V_G| = n$, then $\mathsf{freq}_G^a(\mathbf{w}) \in \mathsf{FFreq}_n(\tau, I, a)$, since $[K_1(I)](G, \mathbf{w}) = |V_G \setminus \{w_1, \ldots, w_k\}| = n - |\Pi_{\mathsf{type}(\mathbf{w})}|$, as required by the definition.

Definition 3.6.7 (Extending) *Let* I *be a set and let* $J = I \cup \{i^*\}$. *Let* $a \geq b > 0$ *be positive integers. We say* $(\tau', f') \in \mathsf{Types}(J) \times \mathsf{FFreq}(\tau', J, b)$ extends $(\tau, f) \in \mathsf{Types}(I) \times \mathsf{FFreq}(\tau, I, a)$ *if* τ' *extends* τ, *and for every* $F \in \mathsf{Conn}_I^b$, *we have*

1. *if* $\{i^*\} \notin \Pi_{\tau'}$,

$$f_F = f'_{\widetilde{F}}, \tag{3.4}$$

where $\widetilde{F}$ *is the graph obtained from* F *by introducing an isolated vertex labelled* i^*.

2. *if* $\{i^*\} \in \Pi_{\tau'}$, *letting* $\delta_H : \mathbb{Z}_q^{\mathsf{Conn}_J^b} \to \mathbb{Z}_q$ *be the function given by Lemma 3.6.2,*

$$f_F = f'_{\widetilde{F}} + \sum_{u \in V_F \setminus \mathcal{L}(F)} c_u \delta_{F_u}(f'), \tag{3.5}$$

where

- $\widetilde{F}$ *is the graph obtained from* F *by introducing an isolated vertex labelled* i^*.
- c_u *equals* 1 *if for all* $i \in I$, *if* u *is adjacent to* $F(i)$, *then* $\{\Pi_{\tau'}(i^*), \Pi_{\tau'}(i)\} \in E_{\tau'}$. *Otherwise,* $c_u = 0$.
- F_u *is the graph obtained from* F *by labelling the vertex* u *by* i^* *and deleting all edges between* u *and the other labelled vertices of* F.

The crux of the above definition is captured in the following lemma.

Lemma 3.6.8 *Let G be a graph. Let $a \geq b > 0$ be integers. Let $\mathbf{w} \in V^k$ and $v \in V$. Let $\tau = \mathsf{type}_G(\mathbf{w})$, $\tau' = \mathsf{type}_G(\mathbf{w}, v)$, $f = \mathsf{freq}^a_G(\mathbf{w})$ and $f' = \mathsf{freq}^b_G(\mathbf{w}, v)$. Then (τ', f') extends (τ, f).*

Proof We keep the notation of the previous definition. First observe that τ' extends τ.

If $\{k+1\} \notin \Pi_{\tau'}$, then we need to show that $[F]_q(G, \mathbf{w}) = [\widetilde{F}]_q(G, \mathbf{w}, v)$ for each $F \in \mathsf{Conn}^b_k$. This is immediate from the definitions.

If $\{k+1\} \in \Pi_{\tau'}$, then we need to show that $[F]_q(G, \mathbf{w}) = [\widetilde{F}]_q(G, \mathbf{w}, v) + \sum_{u \in V_F \setminus \mathcal{L}(F)} c_u [F_u]_q(G, \mathbf{w}, v)$. We do this by counting the $\chi \in \mathsf{Inj}(F, (G, \mathbf{w}))$ based on its image $\chi(V_F)$ as follows:

1. Category 1: $v \notin \chi(V_F)$. There are precisely $[\widetilde{F}](G, \mathbf{w}, v)$ such χ.

2. Category 2: $v = \chi(u)$ (in this case u is uniquely specified). Note that $u \notin \mathcal{L}(F)$. Then it must be the case that for any $i \in [k]$ such that u is adjacent to $F(i)$, w_i is adjacent to v. Thus $\{\Pi_{\tau'}(i), \Pi_{\tau'}(k+1)\} \in E_{\tau'}$, and so $c_u = 1$. The number of such χ is $[F_u](G, \mathbf{w}, v)$.

This proves the desired relation. ■

We now state and prove two key uniqueness properties enjoyed by the notion of extension.

Lemma 3.6.9 *Let $a \geq b > 0$ be integers. Let $\mathbf{w} \in V_G^k$. Let $u \in V_G \setminus \{w_1, \ldots, w_k\}$. Let $\tau = \mathsf{type}_G(\mathbf{w})$ and $\tau' = \mathsf{type}_G(\mathbf{w}, u)$. Let $f = \mathsf{freq}^a_G(\mathbf{w})$. Then $\mathsf{freq}^b_G(\mathbf{w}, u)$ is the unique $f' \in \mathbb{Z}_q^{\mathsf{Conn}^b_{k+1}}$ such that:*

- *for each $H \in \mathsf{Conn}^b_{k+1}$ that is dependent on label $k+1$, we have $f'_H = [H]_q(G, \mathbf{w}, u)$.*

- *(τ', f') extends (τ, f).*

Proof By Lemma 3.6.8, the vector $\mathsf{freq}^b_G(\mathbf{w}, u)$ is such an f'.

To prove uniqueness, it suffices to show that any f' satisying these two properties equals $\mathsf{freq}^b_G(\mathbf{w}, u)$. Thus it suffices to show that for any $H \in \mathsf{Conn}^b_{k+1}$ not dependent on label $k+1$, $f'_H = (\mathsf{freq}^b_G(\mathbf{w}, u))_H$.

We prove this by induction on $|V_H \setminus \mathcal{L}(H)|$. Let $H \in \mathsf{Conn}^b_{k+1}$ not dependent on label $k+1$. Thus H is of the form $\widetilde{F}$ for some graph $F \in \mathsf{Conn}^b_k$ (as in the previous lemma, for a $[k]$-labelled graph F, we let $\widetilde{F}$ be the $[k+1]$-labelled graph obtained by adjoining an isolated vertex labelled $k+1$ to F). By Equation (3.5), we see that f'_H is *uniquely* determined by τ, τ', f_F and the numbers $(f'_{H'})_{H' \in \mathsf{Conn}^{|V_H \setminus \mathcal{L}(H)|-1}_{k+1}}$ (since each c_u is determined by τ' and each of the graphs F_u have $|F_u \setminus \mathcal{L}(F_u)| \leq |V_H \setminus \mathcal{L}(H)| - 1$). By induction hypothesis, all the $f'_{H'} = (\mathsf{freq}^b_G(\mathbf{w}, u))_{H'}$. Thus, since $\mathsf{freq}^b_G(\mathbf{w}, u)$ also satisfies Equation (3.5), we have $f'_H = (\mathsf{freq}^b_G(\mathbf{w}, u))_H$, as required. ■

Lemma 3.6.10 *Let $a \geq b > 0$ be integers. Let $(\tau, f) \in \mathsf{Types}(k) \times \mathsf{FFreq}(\tau, [k], a)$. Let $\tau' \in \mathsf{Types}(k+1)$ extend τ with $\{k+1\} \notin \Pi_{\tau'}$. Then there is at most one $f' \in \mathsf{FFreq}(\tau', [k+1], b)$ such that (τ', f') extends (τ, f).*

Proof As in the previous lemma, for a $[k]$-labelled graph F, we let $\widetilde{F}$ be the $[k+1]$-labelled graph obtained by adjoining an isolated vertex labelled $k+1$ to F. For any $F \in \mathsf{Conn}^b_k$, we must have $f'_{\widetilde{F}} = f_F$. Now we claim that any $H \in \mathsf{Conn}^b_k$ is Π-equivalent to some graph of the form $\widetilde{F}$. To prove this, let $j \in [k]$ be such that $\Pi_{\tau'}(j) = \Pi_{\tau'}(k+1)$. Let H^* be the graph obtained from H by adding, for each neighbor u of $H(k+1)$, an edge between u and the $H(j)$, and then removing (a) all edges incident on $H(k+1)$, and (b) any duplicate edges introduced. By construction, $H/\Pi_{\tau'} \cong H^*/\Pi_{\tau'}$, and so $f'_H = f'_{H^*}$ by Equation (3.3). In addition, the $H^*(k+1)$ is isolated, and hence H^* is of the form $\widetilde{F}$ for some $F \in \mathsf{Conn}^b_k$.

What we have shown is that for every $H \in \mathsf{Conn}^b_{k+1}$, f'_H is forced to equal f_F for some $F \in \mathsf{Conn}^b_k$. This implies that f' is specified uniquely. ■

Finally, we will need to deal with random graphs $G(n,p)$ with some of the edges already exposed. The next definition captures this object.

Definition 3.6.11 (Conditioned Random Graph) *Let $A = (V_A, E_A)$ be a graph with $V_A \subseteq [n]$. We define the* conditioned random graph *$G(n,p \mid V_A, E_A)$ to be the graph $G = (V_G, E_G)$ with $V_G = [n]$ and $E_G = E_A \cup E'$, where each $\{i,j\} \in \binom{[n]}{2} \setminus \binom{V_A}{2}$ is included in E' independently with probability p.*

We can now state the main technical theorem that describes the distribution of labelled subgraph frequencies, and will eventually be useful for eliminating $\exists$ quantifiers.

Theorem 3.6.12 *Let $a \geq b$ be positive integers. Let A be a graph with $V_A \subseteq [n]$ and $|V_A| \leq n' \leq n/2$. Let $G \in G(n,p \mid V_A, E_A)$. Let $\mathbf{w} = (w_1, \ldots, w_k) \in V_A^k$, and let $u_1, \ldots, u_s \in V_A \setminus \{w_1, \ldots, w_k\}$ be distinct. Let $\tau = \mathsf{type}_G(\mathbf{w})$ and let $\tau_i = \mathsf{type}_G(\mathbf{w}, u_i)$ (note that $\tau, \tau_1, \ldots, \tau_s$ are already determined by E_A). Let f denote the random variable $\mathsf{freq}_G^a(\mathbf{w})$. Let f_i denote the random variable $\mathsf{freq}_G^b(\mathbf{w}, u_i)$.*

Then, there exists a constant $\rho = \rho(a,q,p) > 0$, such that if $s \leq \rho \cdot n$, then the distribution of $(f, f_1, \ldots, f_s)$ over $\mathsf{FFreq}_n(\tau, [k], a) \times \prod_i \mathsf{FFreq}_n(\tau_i, [k+1], b)$ is $2^{-\Omega(n)}$-close to the distribution of $(h, h_1, \ldots, h_s)$ generated as follows:

1. *h is picked uniformly at random from $\mathsf{FFreq}_n(\tau, [k], a)$.*
2. *For each i, each h_i is picked independently and uniformly from the set of all $f' \in \mathsf{FFreq}_n(\tau_i, [k+1], b)$ such that (τ_i, f') extends (τ, h).*

3.6.3 Proof of Theorem 3.5.8

We now prove Theorem 3.5.8, where the main quantifier elimination step is carried out.

Theorem 3.5.8 (restated) *For every prime q and integers $k, t > 0$, there is a constant $c = c(k,t,q)$ such that for every $\mathsf{FO}[\mathsf{Mod}_q]$ formula $\varphi(\alpha_1, \ldots, \alpha_k)$ with quantifier depth t, there is a function $\psi : \mathsf{Types}(k) \times \mathbb{Z}_q^{\mathsf{Conn}_k^c} \to \{0,1\}$ such that for all $p \in (0,1)$, the quantity*

$$\Pr_{G \in G(n,p)} \left[\begin{array}{l} \forall w_1, \ldots, w_k \in V_G, \\ \quad (G \text{ satisfies } \varphi(w_1, \ldots, w_k)) \Leftrightarrow (\psi(\mathsf{type}_G(\mathbf{w}), \mathsf{freq}_G^c(\mathbf{w})) = 1) \end{array} \right] \geq 1 - 2^{-\Omega(n)}.$$

Proof The proof is by induction on the size of the formula. If $\varphi(w_1, \ldots, w_k)$ is an atomic formula, then trivially there exists a $\psi : \mathsf{Types}(k) \to \{0,1\}$ such that for every graph G and every $\mathbf{w} \in V_G^k$, the statement $\varphi(w_1, \ldots, w_k)$ holds if and only if $\psi(\mathsf{type}_G(\mathbf{w})) = 1$. Thus we may take $c(k,0,q) = 0$. We will show that one may take $c(k,t,q) = (q-1) \cdot c(k+1, t-1, q) \cdot 2^{c(k+1,t-1,q)^2} + 1$.

Now assume the result holds for all formulae smaller than φ.

Case $\wedge$: Suppose $\varphi(\alpha_1, \ldots, \alpha_k) = \varphi_1(\alpha_1, \ldots, \alpha_k) \wedge \varphi_2(\alpha_1, \ldots, \alpha_k)$. By induction hypothesis, we have functions ψ_1, ψ_2 and a constant c such that $\Pr_G[\forall w_1, \ldots, w_k \in V_G, (\varphi_1(w_1, \ldots, w_k) \Leftrightarrow \psi_1(\mathsf{type}_G(\mathbf{w}), \mathsf{freq}_G^c(\mathbf{w})) = 1)] \geq 1 - 2^{-\Omega(n)}$ and $\Pr_G[\forall w_1, \ldots, w_k \in V_G, (\varphi_2(w_1, \ldots, w_k) \Leftrightarrow \psi_2(\mathsf{type}_G(\mathbf{w}), \mathsf{freq}_G^c(\mathbf{w})) = 1)] \geq 1 - 2^{-\Omega(n)}$. Setting $\psi(\tau, f) = \psi_1(\tau, f) \cdot \psi_2(\tau, f)$, it follows from the union bound that

$$\Pr_G[\forall w_1, \ldots, w_k \in V_G, (\varphi(w_1, \ldots, w_k) \Leftrightarrow \psi(\mathsf{type}_G(\mathbf{w}), \mathsf{freq}_G^c(\mathbf{w})) = 1)] \geq 1 - 2^{-\Omega(n)}.$$

Case $\neg$: Suppose $\varphi(\alpha_1, \ldots, \alpha_k) = \neg\varphi'(\alpha_1, \ldots, \alpha_k)$. Let $\psi' : \mathsf{Types}(k) \times \mathbb{Z}_q^{\mathsf{Conn}_k^c} \to \{0,1\}$ be such that $\Pr_G[\forall w_1, \ldots, w_k \in V_G, (\varphi'(w_1, \ldots, w_k) \Leftrightarrow \psi'(\mathsf{type}_G(\mathbf{w}), \mathsf{freq}_G^c(\mathbf{w})) = 1)] \geq 1 - 2^{-\Omega(n)}$. Setting $\psi(\tau, f) = 1 - \psi'(\tau, f)$, we see that

$$\Pr_G[\forall w_1, \ldots, w_k \in V_G, (\varphi(w_1, \ldots, w_k) \Leftrightarrow \psi(\mathsf{type}_G(\mathbf{w}), \mathsf{freq}_G^c(\mathbf{w})) = 1)] \geq 1 - 2^{-\Omega(n)}.$$

Case Mod_q^i: Suppose $\varphi(\alpha_1, \ldots, \alpha_k) = \mathsf{Mod}_q^i\beta, \varphi'(\alpha_1, \ldots, \alpha_k, \beta)$. Let $c' = c(k+1, t-1, q)$ and let $\psi' : \mathsf{Types}(k+1) \times \mathbb{Z}_q^{\mathsf{Conn}_{k+1}^{c'}} \to \{0,1\}$ be given by the induction hypothesis, so that

$$\Pr_G[\forall w_1, \ldots, w_k, v \in V_G, (\varphi'(w_1, \ldots, w_k, v) \Leftrightarrow \psi'(\mathsf{type}_G(\mathbf{w}, v), \mathsf{freq}_G^{c'}(\mathbf{w}, v)) = 1)] \geq 1 - 2^{-\Omega(n)}.$$

Call G *good* if this event occurs, i.e., if

$$\forall w_1, \ldots, w_k, v \in V_G, (\varphi'(w_1, \ldots, w_k, v) \Leftrightarrow \psi'(\mathsf{type}_G(\mathbf{w}, v), \mathsf{freq}_G^{c'}(\mathbf{w}, v)) = 1).$$

Let $\gamma(w_1, \ldots, w_k)$ be the number (mod q) of v such that $\varphi'(w_1, \ldots, w_k, v)$ is true.

Then for any good G (doing arithmetic mod q),

$$\gamma(w_1,\ldots,w_k) = \sum_{v\in V_G} \psi'(\mathsf{type}_G(\mathbf{w},v), \mathsf{freq}^{c'}_G(\mathbf{w},v)).$$

Grouping terms, we have

$$\begin{aligned}\gamma(w_1,\ldots,w_k) &= \sum_{\tau'\in\mathsf{Types}(k+1)} \sum_{f'\in\mathbb{Z}_q^{\mathsf{Conn}^{c'}_{k+1}}} \psi'(\tau',f')\cdot|\{v\in V_G : \mathsf{type}_G(\mathbf{w},v)=\tau')\wedge \mathsf{freq}_G(\mathbf{w},v)=f'\}| \\ &= \sum_{\tau',f'} \psi'(\tau',f')\cdot\lambda(\tau',f',\mathsf{type}_G(\mathbf{w}),\mathsf{freq}^c_G(\mathbf{w})) \\ &\qquad \text{(applying Theorem 3.6.1, and taking } c=(q-1)c'2^{(c')^2}+1)\end{aligned}$$

which is solely a function of $\mathsf{type}_G(\mathbf{w})$ and $\mathsf{freq}^c_G(\mathbf{w})$. Thus, there is a function ψ : $\mathsf{Types}(k)\times\mathbb{Z}_q^{\mathsf{Conn}^c_k}\to\{0,1\}$ such that for all good G and for all $w_1,\ldots,w_k\in V_G$, $\psi(\mathsf{type}_G(\mathbf{w}),\mathsf{freq}^c_G(\mathbf{w}))=1$ if and only if $\gamma(\mathbf{w})\equiv i \mod q$. Thus,

$$\Pr_G[\forall w_1,\ldots,w_k,((\mathsf{Mod}^i_q v,\varphi'(w_1,\ldots,w_k,v))\Leftrightarrow\psi(\mathsf{type}_G(\mathbf{w}),\mathsf{freq}^c_G(\mathbf{w}))=1)]\geq 1-2^{-\Omega(n)},$$

as desired.

Case $\exists$: Suppose $\varphi(\alpha_1,\ldots,\alpha_k)=\exists\beta,\varphi'(\alpha_1,\ldots,\alpha_k,\beta)$. Let $c'=c(k+1,t-1,q)$ and let $\psi':\mathsf{Types}(k+1)\times\mathbb{Z}_q^{\mathsf{Conn}^{c'}_{k+1}}\to\{0,1\}$ be such that

$$\Pr_G[\forall w_1,\ldots,w_k,v\in V_G,(\varphi'(w_1,\ldots,w_k,v)\Leftrightarrow\psi'(\mathsf{type}_G(\mathbf{w},v),\mathsf{freq}_G(\mathbf{w},v))=1)]\geq 1-2^{-\Omega(n)}. \tag{3.6}$$

For this case, we may choose c to be any integer at least c'. Define $\psi:\mathsf{Types}(k)\times\mathbb{Z}_q^{\mathsf{Conn}^c_k}\to\{0,1\}$ by the rule: $\psi(\tau,f)=1$ if there is a $(\tau',f')\in\mathsf{Types}(k+1)\times\mathsf{FFreq}_n(\tau',[k+1],c')$ extending (τ,f) such that $\psi'(\tau',f')=1$.

Fix any $\mathbf{w}\in[n]^k$. We will show that

$$\Pr_G[(\exists v,\psi'(\mathsf{type}_G(\mathbf{w},v),\mathsf{freq}^{c'}_G(\mathbf{w},v))=1)\Leftrightarrow\psi(\mathsf{type}_G(\mathbf{w}),\mathsf{freq}^c_G(\mathbf{w}))=1]\geq 1-2^{-\Omega(n)}. \tag{3.7}$$

Taking a union bound of (3.7) over all $\mathbf{w} \in [n]^k$, and using Equation (3.6), we conclude that

$$\Pr_{G \in G(n,p)}[\forall w_1, \ldots, w_k \in V_G, (\varphi(w_1, \ldots, w_k) \Leftrightarrow \psi(\mathsf{type}_G(\mathbf{w}), \mathsf{freq}_G(\mathbf{w})) = 1)] \geq 1 - 2^{-\Omega(n)},$$

as desired.

It remains to show Equation (3.7). It will help to expose the edges of the random graph G in three stages.

In the first stage, we expose all the edges between the vertices in $\{w_1, \ldots, w_k\}$.

For the second stage, let $s = \rho(c, q, p) \cdot n$ (where ρ comes from Theorem 3.6.12) and pick distinct vertices $u_1, \ldots, u_s \in [n] \setminus \{w_1, \ldots, w_k\}$. In the second stage, we expose all the unexposed edges between the vertices in $\{w_1, \ldots, w_k, u_1, \ldots, u_s\}$ (i.e., the edges between u_is and w_js, as well as the edges between the u_is and u_js). Denote the resulting graph induced on $\{w_1, \ldots, w_k, u_1, \ldots, u_s\}$ after the second stage by A (so that $V_A = \{w_1, \ldots w_k, u_1, \ldots, u_s\}$).

In the third stage, we expose the rest of the edges in G. Thus G is sampled from $G(n, p \mid V_A, E_A)$.

Let τ denote the random variable $\mathsf{type}_G(\mathbf{w})$. Note that τ is determined after the first stage. Let $\tau_1, \ldots, \tau_s$ denote the random variables $\mathsf{type}_G(\mathbf{w}, u_1), \ldots, \mathsf{type}_G(\mathbf{w}, u_s)$. Note that $\tau_1, \ldots, \tau_s$ are all determined after the second stage. Let f denote the random variable $\mathsf{freq}_G(\mathbf{w})$. Let $f_1, \ldots, f_s$ denote the random variables $\mathsf{freq}_G(\mathbf{w}, u_1), \ldots, \mathsf{freq}_G(\mathbf{w}, u_s)$. The variables $f, f_1, \ldots, f_s$ are all determined after the third stage. Notice that the content of Theorem 3.6.12 is precisely a description of the distribution of $(f, f_1, \ldots, f_s)$.

We identify two bad events B_1 and B_2.

B_1 is defined to be the event: there exists $\sigma \in \mathsf{Types}(k+1)$ extending τ, with $\{k+1\} \in \Pi_\sigma$ (ie, types σ where vertex $k+1$ is distinct from the other vertices), such that

$$|\{i \in [s] : \tau_i = \sigma\}| \leq \frac{1}{2} s \min\{p^k, (1-p)^k\}.$$

(This can be interpreted as saying that the type σ appears abnormally infrequently amongst the τ_i). Note that for any σ extending τ, the events "$\tau_i = \sigma$", for $i \in [s]$,

are independent conditioned on the outcome of the first stage, since they depend on disjoint sets of edges of G. Also, for each i and each σ extending τ with $\{k+1\} \in \Pi_\sigma$, the probability that $\tau_i = \sigma$ is $\geq \min\{p^k, (1-p)^k\}$. Therefore, applying the Chernoff bound, and taking a union bound over all σ extending τ with $\{k+1\} \in \Pi_\sigma$, we see that

$$\Pr[B_1] \leq 2^k \exp(-s \min\{p^k, (1-p)^k\}) \leq 2^{-\Omega(n)}.$$

Now let

$$S = \{(\sigma, g) \in \mathsf{Types}(k+1) \times \mathsf{FFreq}_n(\sigma, [k+1], c') \mid \{k+1\} \in \Pi_\sigma \\ \text{AND } (\sigma, g) \text{ extends } (\tau, f) \text{ AND } \psi'(\sigma, g) = 1\}.$$

B_2 is defined to be the event: $S \neq \emptyset$ and for each $i \in [s]$, $(\tau_i, f_i) \notin S$. We study the probability of $\neg B_1 \wedge B_2$. Let U be the set of $(d, d_1, \ldots, d_s) \in \mathsf{FFreq}_n(\tau, [k], c) \times \prod_i \mathsf{FFreq}_n(\tau_i, [k+1], c')$ such that

1. The set $S(d)$ defined by

$$S(d) = \{(\sigma, g) \in \mathsf{Types}(k+1) \times \mathsf{FFreq}_n(\sigma, [k+1], c') \mid \{k+1\} \in \Pi_\sigma \\ \text{AND } (\sigma, g) \text{ extends } (\tau, d) \text{ AND } \psi'(\sigma, g) = 1\},$$

   is nonempty.

2. For each $i \in [s]$, $(\tau_i, d_i) \notin S(d)$.

By definition, the event B_2 occurs precisely when $(f, f_1, \ldots, f_s) \in U$.

By Theorem 3.6.12, for any fixing of E_A, the probability that $(f, f_1, \ldots, f_s) \in U$ is at most $2^{-\Omega(n)}$ more than the probability that $(h, h_1, \ldots, h_s) \in U$. As the event B_1 is solely a function of E_A, we conclude that $\Pr[\neg B_1 \wedge (f, f_1, \ldots, f_s) \in U)] \leq \Pr[\neg B_1 \wedge (h, h_1, \ldots, h_s) \in U] + 2^{-\Omega(n)}$.

It remains to bound $\Pr[\neg B_1 \wedge (h, h_1, \ldots, h_s) \in U]$. If $S(h) \neq \emptyset$, take a $(\sigma, g) \in S(h)$. In the absence of B_1, the number of $i \in [s]$ with $\tau_i = \sigma$ is at least $\frac{1}{2} s \min\{p^k, (1-p)^k\}$. For all these i, it must hold that $h_i \neq g$ in order for $(h, h_1, \ldots, h_s)$ to lie in U.

Therefore,

$$\Pr[\neg B_1 \wedge (h, h_1, \ldots, h_s) \in U] \leq \left(1 - \frac{1}{|\mathsf{FFreq}_n(\tau, k+1, c')|}\right)^{\frac{1}{2}s \min\{p^k, (1-p)^k\}}.$$

Notice that this last quantity is of the form $2^{-\Omega_{p,q,k,d}(s)}$.

Putting everything together,

$$\Pr[\neg B_1 \wedge B_2] \leq \Pr[\neg B_1 \wedge (h, h_1, \ldots, h_s) \in U] + 2^{-\Omega(n)} \leq 2^{-\Omega(s)} + 2^{-\Omega(n)} \leq 2^{-\Omega(n)}.$$

Therefore, with probability at least $1 - 2^{-\Omega(n)}$, the event B_2 does not occur. The next claim finishes the proof of Equation (3.7), and with that the proof of Theorem 3.5.8.

Claim 3.6.13 *If B_2 does not occur, then*

$$(\exists v, \psi'(\mathsf{type}_G(\mathbf{w}, v), \mathsf{freq}^{c'}_G(\mathbf{w}, v)) = 1) \Leftrightarrow (\psi(\mathsf{type}_G(\mathbf{w}), \mathsf{freq}^{c}_G(\mathbf{w})) = 1).$$

Proof Let $\tau = \mathsf{type}_G(\mathbf{w})$ and $f = \mathsf{freq}^{c}_G(\mathbf{w})$.

If $\psi(\tau, f) = 0$, then we know that for all $(\tau', f') \in \mathsf{Types}(k+1) \times \mathsf{FFreq}_n(\tau', k+1, c')$ extending (τ, f), we have $\psi'(\tau', f') = 0$. Thus by Lemma 3.6.8, for all $v \in V_G$, $\psi'(\mathsf{type}_G(\mathbf{w}, v), \mathsf{freq}^{c'}_G(\mathbf{w}, v)) = 0$, as required.

If $\psi(\tau, f) = 1$, then we consider two situations.

- **The self-fulfilling situation:** If there is a $(\tau', f') \in \mathsf{Types}(k+1) \times \mathsf{FFreq}_n(\tau', k+1, c')$ extending (τ, f) with $\{k+1\} \notin \Pi_{\tau'}$ and $\psi'(\tau', f') = 1$. In this case, take any $j \in [k]$ with $\Pi_{\tau'}(j) = \Pi_{\tau'}(k+1)$, and let $v = w_j$. Thus $\mathsf{type}_G(\mathbf{w}, v) = \tau'$. By Lemma 3.6.10, since (τ', f') extends (τ, f) with $\{k+1\} \notin \Pi_{\tau'}$, it follows that $\mathsf{freq}^{c'}_G(\mathbf{w}, v) = f'$. Therefore, with this choice of v, we have $\psi'(\mathsf{type}_G(\mathbf{w}, v), \mathsf{freq}^{c'}_G(\mathbf{w}, v)) = 1$, as required.

- **The default situation:** In this case, there is a $(\tau', f') \in \mathsf{Types}(k+1) \times \mathsf{FFreq}_n(\tau', k+1, c')$ extending (τ, f) with $\{k+1\} \in \Pi_{\tau'}$ and $\psi'(\tau', f') = 1$. This is precisely the statement that $S \neq \emptyset$. Therefore, by the absence of the event

B_2, there must be an $i \in [r]$ such that $(\tau_i, f_i) \in S$. Taking $v = u_i$, we see that $\psi'(\mathsf{type}_G(\mathbf{w}, v), \mathsf{freq}'^{c'}_G(\mathbf{w}, v)) = 1$, as required.

This completes the proof of the claim. ■■

3.7 Counting Extensions

In this section we prove Theorem 3.6.1.

3.7.1 Subgraph frequency arithmetic

We begin with a definition. A *partial matching* between two I-labelled graphs F_1, F_2 is a subset $\eta \subseteq (V_{F_1} \setminus \mathcal{L}(F_1)) \times (V_{F_2} \setminus \mathcal{L}(F_2))$ that is one-to-one. For two graphs F_1, F_2, let $\mathsf{PMatch}(F_1, F_2)$ be the set of all partial matchings between them.

Definition 3.7.1 (Gluing along a partial matching) *Let F_1 and F_2 be two I-labelled graphs, and let $\eta \in \mathsf{PMatch}(F_1, F_2)$. Define* the gluing of F_1 and F_2 along η, *denoted $F_1 \vee_\eta F_2$, to be the graph obtained by first taking the disjoint union of F_1 and F_2, identifying pairs of vertices with the same label, and then identifying the vertices in each pair of η (and removing duplicate edges). We omit the subscript when $\eta = \emptyset$.*

We have the following simple identity.

Lemma 3.7.2 *For any I-labelled graphs F_1, F_2, any graph G and any $\mathbf{w} \in V_G^I$:*

$$[F_1](G, \mathbf{w}) \cdot [F_2](G, \mathbf{w}) = \sum_{\eta \in \mathsf{PMatch}(F_1, F_2)} [F_1 \vee_\eta F_2](G, \mathbf{w}). \tag{3.8}$$

Proof We give a bijection

$$\alpha : \mathsf{Inj}(F_1, (G, \mathbf{w})) \times \mathsf{Inj}(F_2, (G, \mathbf{w})) \to \coprod_{\eta \in \mathsf{PMatch}(F_1, F_2)} \mathsf{Inj}(F_1 \vee_\eta F_2, (G, \mathbf{w})).$$

Define $\alpha(\chi_1, \chi_2)$ as follows. Let $\eta = \{(v_1, v_2) \in (V_{F_1} \setminus \mathcal{L}(F_1)) \times (V_{F_2} \setminus \mathcal{L}(F_2)) \mid \chi_1(v_1) = \chi_2(v_2)\}$. Let $\iota_1 \in \mathsf{Inj}(F_1, F_1 \vee_\eta F_2)$ and $\iota_2 \in \mathsf{Inj}(F_2, F_1 \vee_\eta F_2)$ be the natural inclusions.

Let $\chi \in \mathsf{Inj}(F_1 \vee_\eta F_2, (G, \mathbf{w}))$ be the unique homomorphism such that for all $v \in V_{F_1}$, $\chi \circ \iota_1(v) = \chi_1(v)$, and for all $v \in V_{F_2}$, $\chi \circ \iota_2(v) = \chi_2(v)$. We define $\alpha(\chi_1, \chi_2) := \chi$.

To see that α is a bijection, we give its inverse β. Let $\eta \in \mathsf{PMatch}(F_1, F_2)$ and $\chi \in \mathsf{Inj}(F_1 \vee_\eta F_2, (G, \mathbf{w}))$. Let $\iota_1 \in \mathsf{Inj}(F_1, F_1 \vee_\eta F_2)$ and $\iota_2 \in \mathsf{Inj}(F_2, F_1 \vee_\eta F_2)$ be the natural inclusions. Define $\beta(\chi) := (\chi \circ \iota_1, \chi \circ \iota_2)$.

Then β is the inverse of α. ■

We can now prove Lemma 3.6.2.

Lemma 3.6.2 (Label-connected subgraph frequencies determine all subgraph frequencies, restated) *For every k-labelled graph F' with $|V_{F'} \setminus \mathcal{L}(F')| \leq t$, there is a polynomial $\delta_{F'} \in \mathbb{Z}[(X_F)_{F \in \mathsf{Conn}_k^t}]$ such that for all graphs G and $\mathbf{w} \in V_G^k$,*

$$[F'](G, \mathbf{w}) = \delta_{F'}(x),$$

where $x \in \mathbb{Z}^{\mathsf{Conn}_k^t}$ is given by $x_F = [F](G, \mathbf{w})$.

Proof By induction on the number of connected components of $F' \setminus \mathcal{L}(F')$. If F' is label-connected , then we take $\delta_{F'}(\mathbf{X}) = X_{F'}$.

Now suppose F' is label-disconnected. Write $F' = F_1 \vee F_2$ where F_1 and F_2 are both k-labelled graphs, and $F_1 \setminus \mathcal{L}(F_1)$ and $F_2 \setminus \mathcal{L}(F_2)$ have fewer connected components.

By equation (3.8), for all G and $\mathbf{w}$,

$$[F_1 \vee F_2](G, \mathbf{w}) = [F_1](G, \mathbf{w}) \cdot [F_2](G, \mathbf{w}) - \sum_{\emptyset \neq \eta \in \mathsf{PMatch}(F_1, F_2)} [F_1 \vee_\eta F_2](G, \mathbf{w}).$$

Observe that for any $\eta \neq \emptyset$, each graph $F_1 \vee_\eta F_2$ has at least one fewer label-connected component than $F_1 \vee F_2 = F'$. Thus, by induction hypothesis, we may take

$$\delta_{F'}(\mathbf{X}) = \delta_{F_1}(\mathbf{X}) \cdot \delta_{F_2}(\mathbf{X}) - \sum_{\emptyset \neq \eta \in \mathsf{PMatch}(F_1, F_2)} \delta_{F_1 \vee_\eta F_2}(\mathbf{X}).$$

This completes the proof of the lemma. ■

3.7.2 Proof of Theorem 3.6.1

Theorem 3.6.1 (restated) *Let q be a prime, let $k, b > 0$ be integers and let $a \geq (q-1) \cdot b \cdot |\mathsf{Conn}^b_{k+1}| + 1$. There is a function*

$$\lambda : \mathsf{Types}(k+1) \times \mathbb{Z}_q^{\mathsf{Conn}^b_{k+1}} \times \mathsf{Types}(k) \times \mathbb{Z}_q^{\mathsf{Conn}^a_k} \to \mathbb{Z}_q$$

such that for all $\tau' \in \mathsf{Types}(k+1)$, $f' \in \mathbb{Z}_q^{\mathsf{Conn}^b_{k+1}}$, $\tau \in \mathsf{Types}(k)$, $f \in \mathbb{Z}_q^{\mathsf{Conn}^a_k}$, it holds that for every graph G, and every $w_1, \ldots, w_k \in V_G$ with $\mathsf{type}_G(\mathbf{w}) = \tau$ and $\mathsf{freq}^a_G(\mathbf{w}) = f$, the cardinality of the set

$$\{v \in V_G : \mathsf{type}_G(\mathbf{w}, v) = \tau' \wedge \mathsf{freq}^b_G(\mathbf{w}, v) = f'\}$$

is congruent to $\lambda(\tau', f', \tau, f) \mod q$.

Proof We describe the function $\lambda(\tau', f', \tau, f)$ explicitly. If τ' does not extend τ, then we set $\lambda(\tau', f', \tau, f) = 0$.

Now assume τ' extends τ. We take cases on whether $k+1$ is a singleton in $\Pi_{\tau'}$ or not.

Case 1: $\{k+1\} \in \Pi_{\tau'}$. In this case, there is an $I \subseteq [k]$ such that $\mathsf{type}_G(w_1, \ldots, w_k, v) = \tau'$ if and only if $v \notin \{w_1, \ldots, w_k\}$ and $(v, w_i) \in E_G \Leftrightarrow i \in I$ (explicitly, $I = \{i \in [k] \mid \{\{k+1\}, \Pi_{\tau'}(i)\} \in E_{\tau'}\}$).

For each $u, v \in V_G$, let $x_{uv} \in \{0, 1\}$, where $x_{uv} = 1$ if and only if u is adjacent to v in G.

Then, using the fact that q is prime, the number (mod q) of v with $\mathsf{type}_G(\mathbf{w}, v) = \tau'$ and $\mathsf{freq}_G(\mathbf{w}, v) = f'$ can be compactly expressed as (doing arithmetic mod q):

$$\sum_{v \in V_G \setminus \{w_1, \ldots, w_k\}} \prod_{i \in I} x_{vw_i} \prod_{j \in [k] \setminus I} (1 - x_{vw_j}) \prod_{F \in \mathsf{Conn}^b_{k+1}} \left(1 - ([F]_q(G, \mathbf{w}, v) - f'_F)^{q-1}\right)$$

Expanding, the expression $\prod_{i \in I} x_{vw_i} \prod_{j \in [k] \setminus I} (1 - x_{vw_j})$ may be expressed in the form $\sum_{S \subseteq [k]} b_S \prod_{i \in S} x_{vw_i}$. Using Lemma 3.7.2, the expression $\prod_{F \in \mathsf{Conn}^b_{k+1}} \left(1 - ([F]_q(G, \mathbf{w}, v) - f'_F)^{q-1}\right)$ may be expressed in the form $\sum_j c_j [F_j]_q(G, \mathbf{w}, v)$, where each F_j is a $k+1$-labelled

graph with at most $|\mathsf{Conn}^b_{k+1}| \cdot b \cdot (q-1)$ unlabelled vertices.

Thus we may rewrite the expression for $\lambda(\tau', f', \tau, f)$ as:

$$\begin{aligned}
&\sum_{v \in [n] \setminus \{w_1, \ldots, w_k\}} \left(\sum_S b_S \prod_{i \in S} x_{vw_i} \right) \left(\sum_j c_j [F_j]_q(G, \mathbf{w}, v) \right) \\
&= \sum_{S,j} b_S c_j \sum_{v \in [n] \setminus \{w_1, \ldots, w_k\}} \left(\left(\prod_{i \in S} x_{vw_i} \right) [F_j]_q(G, \mathbf{w}, v) \right) \\
&= \sum_{S,j} b_S c_j [F'_{S,j}]_q(G, \mathbf{w}),
\end{aligned}$$

where $F'_{S,j}$ is the k-labelled graph obtained from F_j by

(a) For each $i \in S$, adding an edge between the vertex labelled $k+1$ and the vertex labelled i, and

(b) Removing the label from the vertex labelled $k+1$.

Note that $F'_{S,j}$ has at most $|\mathsf{Conn}^b_{k+1}| \cdot b \cdot (q-1) + 1 \leq a$ unlabelled vertices. Thus, by Lemma 3.6.2, $[F'_{S,j}]_q(G, \mathbf{w})$ is determined by $\mathsf{freq}^a_G(\mathbf{w})$. This completes the definition of λ in this case.

Case 2: $\{k+1\} \notin \Pi_{\tau'}$. This case is much easier to handle. Pick any $j \in [k]$ such that $\Pi_{\tau'}(j) = \Pi_{\tau'}(k+1)$. Then there is only one $v \in V_G$ such that $\mathsf{type}_G(\mathbf{w}, v) = \tau'$ (namely, w_j).

Then $\lambda(\tau', f', \tau, f) = 1$ if and only if for all $F' \in \mathsf{Conn}^b_{k+1}$, $f'_{F'} = f_F$, where $F \in \mathsf{Conn}^b_k$ is the graph obtained by identifying the vertex labelled $k+1$ with the vertex labelled j, and labelling this new vertex j. Otherwise $\lambda(\tau', f', \tau, f) = 0$.

This completes the definition of our desired function λ. ■

3.8 The Distribution of Labelled Subgraph Frequencies mod q

In this section, we prove Theorem 3.6.12. As in Section 3.3, the proof will be via an intermediate theorem (Theorem 3.8.2) that proves the equidistribution of the number

of *copies* of labelled subgraphs in $G(n,p)$.

3.8.1 Equidistribution of labelled subgraph copies

First, we gather some simple observations about injective homomorphisms from label-connected graphs for later use (the proofs are simple and are omitted).

Proposition 3.8.1 (Simple but delicate observations about label-connected graphs) *Let $F, F' \in \mathsf{Conn}_I^t$. Let G be a graph and let $\mathbf{w} \in V_G^I$ with all $(w_i)_{i\in I}$ distinct.*

1. *If $E \in \mathsf{Cop}(F,(G,\mathbf{w}))$, the $|E| = |E_F|$.*

2. *If $F \not\cong F'$, we have $\mathsf{Cop}(F,(G,\mathbf{w})) \cap \mathsf{Cop}(F',(G,\mathbf{w})) = \emptyset$.*

3. *Let $\chi_1,\ldots,\chi_r \in \mathsf{Inj}(F,(G,\mathbf{w}))$ be such that for any distinct $j,j' \in [r]$, $\chi_j(V_F \setminus \mathcal{L}(F)) \cap \chi_{j'}(V_F \setminus \mathcal{L}(F)) = \emptyset$. Let $\chi \in \mathsf{Inj}(F',(G,\mathbf{w}))$. Suppose $\chi(E_{F'}) \subseteq (\cup_j \chi_j(E_F))$. Then there is a $j \in [r]$ such that $\chi(E_{F'}) \subseteq \chi_j(E_F)$.*

We can now state and prove an equidistribution theorem for the number of copies of labelled subgraphs in a conditioned random graph. Theorem 3.6.12 will follow from this.

Theorem 3.8.2 *Let A be a graph with $V_A \subseteq [n]$ and $|V_A| \leq n'$. Let $\mathbf{w} = (w_1,\ldots,w_k) \in V_A^k$ with $w_1,\ldots,w_k$ distinct. Let $u_1,\ldots u_s \in V_A \setminus \{w_i : i \in I\}$ be distinct. Let $F_1,\ldots,F_\ell$ be distinct k-labelled label-connected graphs, with $1 \leq |E_{F_i}| \leq d$. Let $H_1,\ldots,H_{\ell'}$ be distinct $k+1$-labelled label-connected graphs dependent on label $k+1$, with $1 \leq |E_{H_i}| \leq d$.*

Let $G \in G(n,p \mid V_A, E_A)$. Then the distribution of

$$((\langle F_i\rangle_q(G,\mathbf{w}))_{i\in[\ell]}, (\langle H_{i'}\rangle_q(G,\mathbf{w},u_{j'}))_{i'\in[\ell'],j'\in[s]})$$

on $\mathbb{Z}_q^{\ell+s\ell'}$ is $2^{-\Omega_{q,p,d}(n-n')+(\ell+\ell's)\log q}$-close to uniform in statistical distance.

Proof By the Vazirani XOR lemma (Lemma 3.3.4), it suffices to show that for any nonzero $(c, c') \in \mathbb{Z}_q^{\ell} \times \mathbb{Z}_q^{\ell' \times s}$, we have $\left|\mathbb{E}\left[\omega^R\right]\right| \leq 2^{-\Omega_{q,p,d}(n-n')}$, where

$$R := \sum_{i \in [\ell]} c_i \langle F_i \rangle_q(G, \mathbf{w}) + \sum_{i' \in [\ell']} \sum_{j' \in [s]} c'_{i'j'} \langle H_{i'} \rangle_q(G, \mathbf{w}, u_{j'})$$

and $\omega \in \mathbb{C}$ is a primitive q^{th}-root of unity.

We will show this by appealing to Lemma 3.3.3. Let $m = \binom{n}{2} - \binom{a}{2}$. Let $\mathbf{z} \in \{0,1\}^{\binom{[n]}{2}}$ be the random variable where, for each $e \in \binom{[n]}{2}$, $z_e = 1$ if and only if edge e is present in G. Thus, independently for each $e \in \binom{[n]}{2} \setminus \binom{V_A}{2}$, $\Pr[z_e = 1] = p$, while for $e \in \binom{V_A}{2}$, the value of z_e is either identically 1 or identically 0 (depending on whether $e \in E_A$ or not).

We may now express R in terms of the z_e. We have,

$$\begin{aligned} R &= \sum_{i \in [\ell]} c_i \langle F_i \rangle_q(G, \mathbf{w}) + \sum_{i' \in [\ell']} \sum_{j' \in [s]} c'_{i'j'} \langle H_{i'} \rangle_q(G, \mathbf{w}, u_{j'}) \\ &= \sum_{i \in [\ell]} c_i \sum_{E \in \mathsf{Cop}(F_i, (\mathrm{K}_n, \mathbf{w}))} \prod_{e \in E} z_e + \sum_{i' \in [\ell']} \sum_{j' \in [s]} c'_{i'j'} \sum_{E \in \mathsf{Cop}(H_{i'}, (\mathrm{K}_n, \mathbf{w}, u_{j'}))} \prod_{e \in E} z_e \\ &= \sum_{E \in \mathcal{F}_1} c_E \prod_{e \in E} z_e + \sum_{E \in \mathcal{F}_2} c'_E \prod_{e \in E} z_e, \end{aligned}$$

where $\mathcal{F}_1 \subseteq 2^{\binom{[n]}{2}}$ is the set $\bigcup_{i \in [\ell] : c_i \neq 0} \mathsf{Cop}(F_i, (\mathrm{K}_n, \mathbf{w}))$, $\mathcal{F}_2$ is the set $\bigcup_{i' \in [\ell'], j' \in [s] : c'_{i'j'} \neq 0} \mathsf{Cop}(H_{i'}, (\mathrm{K}_n, \mathbf{w},$ for each $E \in \mathcal{F}_1$, $c_E = c_i$ where $i \in [\ell]$ is such that $E \in \mathsf{Cop}(F_i, (\mathrm{K}_n, \mathbf{w}))$ (note that by Proposition 3.8.1 there is exactly one such i), and similarly, for $E \in \mathcal{F}_2$, $c'_E = \sum_{i' \in [\ell'], j' \in [s] : E \in \mathsf{Cop}(H_{i'}, (\mathrm{K}_n, \mathbf{w}, u_{j'}))} c'_{i'j'}$. Thus if E is such that there is a unique $(i', j') \in [\ell'] \times [s]$ for which $E \in \mathsf{Cop}(H_{i'}, (\mathrm{K}_n, \mathbf{w}, u_{j'}))$ and $c'_{i',j'} \neq 0$, then $c'_E \neq 0$.

Let $Q(\mathbf{Z}) \in \mathbb{Z}_q[\mathbf{Z}]$, where $\mathbf{Z} = (Z_e)_{e \in \binom{[n]}{2} \setminus \binom{V_A}{2}}$, be the polynomial

$$\sum_{E \in \mathcal{F}_1} c_E \prod_{e \in E \cap \binom{V_A}{2}} z_e \prod_{e \in E \setminus \binom{V_A}{2}} Z_e + \sum_{E \in \mathcal{F}_2} c'_E \prod_{e \in E \cap \binom{V_A}{2}} z_e \prod_{e \in E \setminus \binom{V_A}{2}} Z_e.$$

Let $\widehat{\mathbf{z}} \in \{0,1\}^{\binom{[n]}{2} \setminus \binom{V_A}{2}}$ be the random variable $\mathbf{z}$ restricted to the coordinates indexed by $\binom{[n]}{2} \setminus \binom{V_A}{2}$ (thus each coordinate of $\widehat{\mathbf{z}}$ independently equals 1 with probability

p). Then $R = Q(\widehat{\mathbf{z}})$. We wish to show that

$$\left|\mathbb{E}\left[\omega^{Q(\widehat{\mathbf{z}})}\right]\right| \leq 2^{-\Omega_{q,p,d}(n-n')}. \tag{3.9}$$

We do this by demonstrating that the polynomial $Q(\mathbf{Z})$ satisfies the hypotheses of Lemma 3.3.3.

Let $d_1^* = \max_{i:c_i \neq 0} |E_{F_i}|$. Let $d_2^* = \max_{i',j':c'_{i'j'} \neq 0} |E_{H_{i'}}|$. We take cases depending on whether $d_1^* < d_2^*$ or $d_1^* \geq d_2^*$.

Case 1: Suppose $d_1^* < d_2^*$. Let i'_0, j'_0 be such that $c'_{i'_0 j'_0} \neq 0$ and $|E_{H_{i'_0}}| = d_2^*$. Then $Q(\mathbf{Z})$ may be written as $\sum_{E \in \mathcal{F}} c'_E \prod_{e \in E} Z_e + Q'(\mathbf{Z})$, where $\mathcal{F} = \{E \in \mathcal{F}_2 : E \cap \binom{V_A}{2} = \emptyset\}$ and $\deg(Q') < d_2^*$.

Let $\chi_1, \chi_2, \ldots, \chi_r \in \mathsf{Inj}(H_{i'_0}, (\mathrm{K}_n, \mathbf{w}, u_{j'_0}))$ be a collection of homomorphisms such that:

1. For all $j \in [r]$, we have $\chi_j(V_{H_{i'_0}} \setminus \mathcal{L}(H_{i'_0})) \subseteq [n] \setminus V_A$.
2. For all distinct $j, j' \in [r]$, we have $\chi_j(V_{H_{i'_0}} \setminus \mathcal{L}(H_{i'_0})) \cap \chi_{j'}(V_{H_{i'_0}} \setminus \mathcal{L}(H_{i'_0})) = \emptyset$.

Such a collection can be chosen greedily so that $r = \Omega(\frac{n-n'}{d})$. Let $E_j \in \mathsf{Cop}(H_{i'_0}, (\mathrm{K}_n, \mathbf{w}, u_{j'_0}))$ be given by $\chi_j(E_{H_{i'_0}})$. Let $\mathcal{E}$ be the family of sets $\{E_1, \ldots, E_r\} \subseteq \mathcal{F}$. We observe the following properties of the E_j:

1. For each $j \in [r]$, $|E_j| = d_2^*$ (since χ_j is injective and $w_1, \ldots, w_k, u_{j'_0}$ are distinct).
2. For each $j \in [r]$, $c'_{E_j} \neq 0$. This is because there is a unique (i', j') (namely (i'_0, j'_0)) for which $c_{i'j'} \neq 0$ and $E_j \in \mathsf{Cop}(H_{i'}, (\mathrm{K}_n, \mathbf{w}, u_{j'}))$. Indeed, if $j' \neq j'_0$, then each $E^* \in \mathsf{Cop}(H_{i'}, (\mathrm{K}_n, \mathbf{w}, u_{j'}))$ has some element incident on $u_{j'}$ (while E_j does not). On the other hand, if $j' = j'_0$ and $i' \neq i'_0$, then Proposition 3.8.1 implies that $\mathsf{Cop}(H_{i'}, (\mathrm{K}_n, \mathbf{w}, u_{j'})) \cap \mathsf{Cop}(H_{i'_0}, (\mathrm{K}_n, \mathbf{w}, u_{j'_0})) = \emptyset$.
3. For distinct $j, j' \in [r]$, $E_j \cap E_{j'} = \emptyset$ (by choice of the χ_j).
4. For any $S \in \mathcal{F} \setminus \mathcal{E}$, $|S \cap (\cup_j E_j)| < d_2^*$. To see this, take any $S \in \mathcal{F} \setminus \mathcal{E}$ and suppose $|S \cap (\cup_j E_j)| \geq d_2^*$. Let $i' \in [\ell'], j' \in [s]$ be such that $S \in \mathsf{Cop}(H_{i'}, (\mathrm{K}_n, \mathbf{w}, u_{j'}))$.

Let $\chi \in \mathsf{Inj}(H_{i'}, (\mathrm{K}_n, \mathbf{w}, u_{j'}))$ with $\chi(E_{H_{i'}}) = S$. By choice of d_2^*, we know that $|S| \leq d_2^*$. Therefore, the only way that $|S \cap (\cup_j E_j)|$ can be $\geq d_2^*$ is if (a) $|S| = d_2^*$, and (b) $S \cap (\cup_j E_j) = S$, or in other words, $S \subseteq (\cup_j E_j)$. Since $H_{i'}$ is dependent on label $k+1$, we know that S has some element incident on vertex $u_{j'}$, and thus (b) forces $j' = j'_0$ (otherwise no E_j is incident on $u_{j'}$). Now by Proposition 3.8.1, this implies that $S \subseteq E_j$ for some j. But since $|E_j| = |S|$, we have $S = E_j$, contradicting our choice of S. Therefore, $|S \cap (\cup_j E_j)| < d_2^*$ for any $S \in \mathcal{F} \setminus \mathcal{E}$.

It now follows that $Q(\mathbf{Z}), \mathcal{F}$ and $\mathcal{E}$ satisfy the hypothesis of Lemma 3.3.3. Consequently, (noting that $d_2^* \leq d$) Equation (3.9) follows, completing the proof in Case 1.

Case 2: Suppose $d_1^* \geq d_2^*$. Let i_0 be such that $c_{i_0} \neq 0$ and $|E_{F_{i_0}}| = d_1^*$. Then $Q(\mathbf{Z})$ may be written as $\sum_{E \in \mathcal{F}} (c_E + c'_E) \prod_{e \in E} Z_e + Q'(\mathbf{Z})$, where $\mathcal{F} = \{E \in \mathcal{F}_1 \cup F_2 : E \cap \binom{V_A}{2} = \emptyset\}$ and $\deg(Q') < d_1^*$.

Let $\chi_1, \chi_2, \ldots, \chi_r \in \mathsf{Inj}(F_{i_0}, (\mathrm{K}_n, \mathbf{w}))$ be a collection of homomorphisms such that:

1. For all $j \in [r]$, we have $\chi_j(V_{F_{i_0}} \setminus \mathcal{L}(F_{i_0})) \subseteq [n] \setminus V_A$.

2. For all distinct $j, j' \in [r]$, we have $\chi_j(V_{F_{i_0}} \setminus \mathcal{L}(F_{i_0})) \cap \chi_{j'}(V_{F_{i_0}} \setminus \mathcal{L}(F_{i_0})) = \emptyset$.

Such a collection can be chosen greedily so that $r = \Omega(\frac{n-n'}{d})$. Let $E_j \in \mathsf{Cop}(F_{i_0}, (\mathrm{K}_n, \mathbf{w}))$ be given by $\chi_j(E_{F_{i_0}})$. Let $\mathcal{E}$ be the family of sets $\{E_1, \ldots, E_r\} \subseteq \mathcal{F}$. We observe the following properties of the E_j:

1. For each $j \in [r]$, $|E_j| = d_1^*$ (since χ_j is injective and $w_1, \ldots, w_k$ are distinct).

2. For each $j \in [r]$, $c_{E_j} + c'_{E_j} \neq 0$. This is because $c_{E_j} = c_{i_0} \neq 0$ and for any (i', j'), $E_j \notin \mathsf{Cop}(H_{i'}, (\mathrm{K}_n, \mathbf{w}, u_{j'}))$ (and so $c'_{E_j} = 0$). To see the latter claim, note that each $E^* \in \mathsf{Cop}(H_{i'}, (\mathrm{K}_n, \mathbf{w}, u_{j'}))$ has an element incident on $u_{j'}$ (which E_j does not).

3. For distinct $j, j' \in [r]$, $E_j \cap E_{j'} = \emptyset$ (by choice of the χ_j).

4. For any $S \in \mathcal{F} \setminus \mathcal{E}$, $|S \cap (\cup_j E_j)| < d_1^*$. To see this, take any $S \in \mathcal{F} \setminus \mathcal{E}$ and suppose $|S \cap (\cup_j E_j)| \geq d_1^*$.

(a) If $S \in \mathcal{F}_1$, then let $i \in [\ell]$ be such that $S \in \mathsf{Cop}(F_i, (\mathrm{K}_n, \mathbf{w}))$. Let $\chi \in \mathsf{Inj}(F_i, (\mathrm{K}_n, \mathbf{w}))$ with $\chi(E_{F_i}) = S$. We know that $|S| \leq d_1^*$. Therefore, the only way that $|S \cap (\cup_j E_j)|$ can be $\geq d_1^*$ is if (1) $|S| = d_1^*$, and (2) $S \cap (\cup_j E_j) = S$, or in other words, $S \subseteq (\cup_j E_j)$. However, by Proposition 3.8.1, this implies that $S \subseteq E_j$ for some j. But since $|E_j| = |S|$, we have $S = E_j$, contradicting our choice of S.

(b) If $S \in \mathcal{F}_2$, then let $i' \in [\ell'], j' \in [s]$ be such that $S \in \mathsf{Cop}(H_{i'}, (\mathrm{K}_n, \mathbf{w}, u_{j'}))$. Let $\chi \in \mathsf{Inj}(H_{i'}, (\mathrm{K}_n, \mathbf{w}, u_{j'}))$ with $\chi(E_{H_{i'}}) = S$. We know that $|S| \leq d_2^* \leq d_1^*$. Now S has an element incident on $u_{j'}$. On the other hand none of the E_j have any edges incident on $u_{j'}$. Therefore $|S \cap (\cup_j E_j)| < |S| \leq d_1^*$.

Therefore, $|S \cap (\cup_j E_j)| < d_1^*$ for any $S \in \mathcal{F} \setminus \mathcal{E}$.

It now follows that $Q(\mathbf{Z}), \mathcal{F}$ and $\mathcal{E}$ satisfy the hypothesis of Lemma 3.3.3. Consequently, (noting that $d_1^* \leq d$) Equation (3.9) follows, completing the proof in Case 2. ■

3.8.2 Proof of Theorem 3.6.12

Theorem 3.6.12 (restated) *Let $a \geq b$ be positive integers. Let A be a graph with $V_A \subseteq [n]$ and $|V_A| \leq n' \leq n/2$. Let $G \in G(n, p \mid V_A, E_A)$. Let $\mathbf{w} = (w_1, \ldots, w_k) \in V_A^k$, and let $u_1, \ldots, u_s \in V_A \setminus \{w_1, \ldots, w_k\}$ be distinct. Let $\tau = \mathsf{type}_G(\mathbf{w})$ and let $\tau_i = \mathsf{type}_G(\mathbf{w}, u_i)$ (note that $\tau, \tau_1, \ldots, \tau_s$ are already determined by E_A). Let f denote the random variable $\mathsf{freq}_G^a(\mathbf{w})$. Let f_i denote the random variable $\mathsf{freq}_G^b(\mathbf{w}, u_i)$.*

Then, there exists a constant $\rho = \rho(a, q, p) > 0$, such that if $s \leq \rho \cdot n$, then the distribution of $(f, f_1, \ldots, f_s)$ over $\mathsf{FFreq}_n(\tau, [k], a) \times \prod_i \mathsf{FFreq}_n(\tau_i, [k+1], b)$ is $2^{-\Omega(n)}$-close to the distribution of $(h, h_1, \ldots, h_s)$ generated as follows:

1. *h is picked uniformly at random from $\mathsf{FFreq}_n(\tau, [k], a)$.*

2. *For each i, each h_i is picked independently and uniformly from the set of all $f' \in \mathsf{FFreq}_n(\tau_i, [k+1], b)$ such that (τ_i, f') extends (τ, h).*

Proof Let $\mathbf{v} = \mathbf{w}/\Pi_\tau$. Let $F_1, \ldots, F_\ell$ be an enumeration of the elements of $\mathsf{Conn}^a_{\Pi_\tau}$.

Let $\Pi' \in \mathsf{Partitions}([k+1])$ equal $\Pi_\tau \cup \{\{k+1\}\}$. Notice that for each $i \in [s]$, $\Pi_{\tau_i} = \Pi'$. Let $H_1, \ldots, H_{\ell'}$ be an enumeration of those elements of $\mathsf{Conn}^b_{\Pi'}$ that are dependent on label i^*.

By Theorem 3.8.2 and the hypothesis on s for a suitable constant ρ, the distribution of

$$(g, g^1, \ldots, g^s) = ((\langle F_i \rangle_q(G, \mathbf{v}))_{i \in [\ell]}, (\langle H_{i'} \rangle(G, \mathbf{v}, u_{j'}))_{i' \in [\ell'], j' \in [s]})$$

is $2^{-\Omega(n)}$ close to uniform over $\mathbb{Z}_q^{\ell + \ell' s}$. Given the vector $(g, g^1, \ldots, g^s)$, we may compute the vector $(f, f_1, \ldots, f_s)$ as follows:

1. For $F = \mathrm{K}_1([k])$, we have $f_F = n - |\Pi_\tau|$.
2. For all other $F \in \mathsf{Conn}^a_k$, let $i \in [\ell]$ be such that $F/\Pi_\tau \cong F_i$. Then $f_F = g_i \cdot \mathsf{aut}(F_i)$.
3. For $H \in \mathsf{Conn}^b_{k+1}$ dependent on label $k+1$, let $i' \in [\ell']$ be such that $H/(\Pi') \cong H'_i$. Then for each $j' \in [s]$, $(f_{j'})_H = g^{j'}_{i'} \cdot \mathsf{aut}(H_{i'})$.
4. For $H \in \mathsf{Conn}^b_{k+1}$ not dependent on label $k+1$ and for any $j' \in [s]$, there is a *unique* setting of $(f_{j'})_H$ (given the settings above) that is consistent with the fact that (τ_j, f_j) extends (τ, f). This follows from Lemma 3.6.9.

This implies the desired claim about the distribution of $(f, f_1, \ldots, f_s)$. ■

3.9 Open problems

We conclude with some open problems:

1. What is the complexity of computing the numbers $a_0, \ldots, a_{q-1}$ in Theorem 3.2.1? We know that it is PSPACE-hard to compute these numbers (it is already PSPACE-hard to tell if the asymptotic probability of a FO sentence is 0 or 1). Our proof shows that they may be computed in time $2^{2^{2^{\cdots}}}$ of height proportional to the quantifier depth of the formula.

2. Is there a modular convergence law for $\mathsf{FO}[\mathsf{Mod}_m]$ for arbitrary m? This encompasses the study of logics which include multiple modular-counting quantifiers, such as $\mathsf{FO}[\mathsf{Mod}_2, \mathsf{Mod}_3]$. Our methods face the same obstacles that prevent the Razborov-Smolensky methods for circuit complexity from generalizing to $\mathsf{AC}^0[\mathsf{Mod}_m]$. Perhaps an answer to the above question will give some hints for $\mathsf{AC}^0[\mathsf{Mod}_m]$?

3. In the spirit of the Shelah-Spencer 0-1 law [SS87, SS88], is there a modular convergence law for $\mathsf{FO}[\oplus]$ on $G(n, n^{-\alpha})$, for irrational α? Even the behavior of subgraph frequencies mod 2 in this setting seems quite intriguing.

Chapter 4

Affine Dispersers from Subspace Polynomials

4.1 Introduction

A *one-output-bit seedless disperser* (often called a deterministic disperser) for a family $\mathcal{F}$ of subsets of $\{0,1\}^m$ is a function $\text{Disp} : \{0,1\}^m \to \{0,1\}$ satisfying the property that on any subset $X \in \mathcal{F}$, $X \subset \{0,1\}^m$ (the set X is often called a "source") the function Disp takes more than one value, i.e., $|\{\text{Disp}(x) : x \in X\}| > 1$. An *extractor* for $\mathcal{F}$ is a function $\text{Extr} : \{0,1\}^m \to \{0,1\}$ satisfying the stronger requirement that for every $X \in \mathcal{F}$, if x is picked uniformly from X, then $\text{Extr}(x)$ is nearly-uniformly distributed. We think of dispersers and extractors as behaving *pseudorandomly* on sources $X \in \mathcal{F}$ because in typical settings where the size of $\mathcal{F}$ is not too large, a random function is indeed an extractor and hence also a disperser. Extractors and dispersers have been intensively studied in recent years in the context of extracting randomness from imperfect sources of randomness. The goal of these studies has been to obtain extractors and dispersers computable in polynomial time, and today several constructions of seedless dispersers for various structured families of subsets are known, including for "bit-fixing" and "samplable" sources [CGH$^+$85, GRS06, TV00, KZ07]. We refer the reader to [BKS$^+$05] for more information on seedless dispersers and extractors.

A particularly interesting family of structured subsets that has been considered in this context, and is also the focus of our work, is the family of affine subspaces over a fixed finite field $\mathbb{F}_p$ of size p (think of $p = 2$). Extractors and dispersers for this family of sources are known as affine extractors and dispersers. Affine extractors for spaces of dimension greater than $m/2$ are relatively easy to construct [BHRV01]. However, for spaces of dimension smaller than $m/2$ the problem becomes much harder, and to date, only two explicit pseudorandom constructions are known [BKS+05, Bou07][1]. Both these works give constructions that are shown to behave pseudorandomly on all affine spaces of dimension $\geq \epsilon m$, where $\epsilon > 0$ is any fixed constant. The work of [BKS+05] constructs affine dispersers and that of [Bou07] constructs affine extractors. Both constructions use recent sum-product theorems over finite fields [BKT04, BGK06] and related results from additive combinatorics, along with several other non-trivial ideas.

4.1.1 Results

Our main result (Theorem 4.2.2) is the explicit construction of an affine disperser for spaces of dimension $o(m)$. Specifically, our disperser works for spaces of dimension at least $6m^{4/5}$. The structure of our main affine disperser is as follows. The m coordinates are grouped into r blocks, each with an equal number k of coordinates, and each block is interpreted as specifying an element of the finite field $\mathbb{F}_{p^k}$. The r elements thus obtained in $\mathbb{F}_{p^k}$ are now substituted into a certain polynomial over $\mathbb{F}_{p^k}$, and its output, which is an element of $\mathbb{F}_{p^k}$, is projected onto $\mathbb{F}_p$ via a nontrivial $\mathbb{F}_p$-linear mapping of $\mathbb{F}_{p^k}$ to $\mathbb{F}_p$.

The techniques we use allow for a host of results with a similar flavor. The simplest-to-state result is a "univariate" affine extractor below the $m/2$ barrier. By "univariate" we mean that the function we use to compute the extractor is naturally viewed as a univariate polynomial. Let $\phi : \mathbb{F}_p^m \to \mathbb{F}_{p^m}$ be any $\mathbb{F}_p$-linear isomorphism

[1] A related, though incomparable, result of Gabizon and Raz [GR05] constructs extractors for affine sources over "large" finite fields, where "large" means $p > m^2$, see also [DG09] for recent improvements along this line of research.

and $\pi : \mathbb{F}_{p^m} \to \mathbb{F}_p$ be any nontrivial $\mathbb{F}_p$-linear map[2]. We show that the function $f : \mathbb{F}_p^m \to \mathbb{F}_p$ defined by

$$f(x) = \pi\left((\phi(x))^{1+p+p^2}\right) \tag{4.1}$$

is an extractor for dimension at least $2m/5 + O(1)$, as long as m is odd. Another pseudorandom univariate construction appearing here is

$$f(x) = \pi\left((\phi(x))^{1+p+p^2+p^3}\right) \tag{4.2}$$

which we prove is an affine disperser for dimension greater than $n/3 + O(1)$. We conjecture that this construction is in fact an extractor and that both univariate constructions are merely the first two members of a larger family of univariate extractors (see Conjecture 4.2.6).

We point out that if m is even, then $\mathbb{F}_{p^m}$ has a subfield $\mathbb{F}_{p^{m/2}}$ which is also a $m/2$-dimensional subspace of $\mathbb{F}_{p^m}$ for which the above mentioned constructions will not be a disperser. Indeed, when x belongs to $\mathbb{F}_{p^{m/2}}$ then so does every power of x, hence some nontrivial $\mathbb{F}_p$-linear map π will be constant on both $\left\{x^{1+p+p^2} \;\middle|\; x \in \mathbb{F}_{p^{m/2}}\right\}$ and $\left\{x^{1+p+p^2+p^3} \;\middle|\; x \in \mathbb{F}_{p^{m/2}}\right\}$. In the next section we comment on the role that the oddness of m, and more generally, the absence of subfields, plays in our proofs.

On subspaces and polynomials Our analysis makes use of a class of polynomials called *subspace polynomials*. These polynomials were first systematically studied by Ore in the 1930's [Ore33, Ore34]. They have numerous applications in the study of finite fields and in the theory of error correcting codes (See [Ber68, Chapter 11] and [LN97, Chapter 3, Section 4]). More recently, they have been used within computational complexity to construct short PCPs [BGH+04, BS05, BGH+05] and to study limitations on the list-decodability of the Reed-Solomon code [BKR06].

The polynomials studied in this last line of works are what we call the *kernel-*

[2]Explicitly, ϕ is a bijection between the vector space $\mathbb{F}_p^m$ and the finite field $\mathbb{F}_{p^m}$ and π is a nonzero linear map from $\mathbb{F}_{p^m}$ to $\mathbb{F}_p$. Both mappings are $\mathbb{F}_p$-linear, i.e., they respect addition and multiplication by scalars in $\mathbb{F}_p$.

subspace[3] polynomial associated to a linear subspace $L \subseteq \mathbb{F}_{p^m}$, which is a polynomial whose set of roots equals L. In this work we analyze our dispersers using elementary properties of the *image-subspace polynomial* of a linear subspace L. These polynomials have the property that their image, i.e., the set of values they take over $\mathbb{F}_{p^m}$, equals L. Our proofs begin by first reformulating the property of being an affine disperser in terms of these polynomials. We then use a simple-to-prove, yet extremely powerful, structural lemma about these polynomials, to get our main results.

Pseudorandomness from the absence of subfields It was recently realized, starting with the work of [BIW04] and further developed in [Zuc06, KRVZ06, BRSW06, Bou07], that finite fields without large subfields are the source of many pseudorandom phenomena, and that this can be put to good use in the construction of extractors and dispersers. The above mentioned works all harnessed this pseudorandomness via recent results from additive combinatorics such as the sum-product theorem of Bourgain, Katz and Tao [BKT04] and the related multilinear exponential sum estimates of Bourgain, Glibichuk and Konyagin [BGK06].

In our work, we offer a different algebraic incarnation of this phenomenon. Specifically, we show that the absence of large subfields directly affects the structure of the image-subspace polynomials of the field. Image-subspace polynomials are *linearized*, which means that they are of the form $\sum_{i=0}^{m-1} a_i X^{p^i}$. Roughly speaking, our main structural lemma (Lemma 4.4.3) says that the image-subspace polynomial of a subspace $\mathcal{A}$ of dimension d cannot have d consecutive coefficients a_i that are all zero. Moreover, and this is the crucial part, if $\mathcal{A}$ is not contained in a constant multiple of a subfield of $\mathbb{F}_{p^n}$, then the polynomial cannot have even $d-1$ consecutive coefficients that are all zero. This lemma has a short proof (appearing in Section 4.4.1), yet is extremely powerful. Surprisingly, reducing the maximal length of a sequence of zero-coefficients by 1 (from d to $d-1$) for spaces that are not contained in subfields is all it takes for the underlying pseudorandomness to get exposed.

[3]The terms "kernel-" and "image-subspace polynomials" were suggested by Prahladh Harsha and we thank him for introducing this nomenclature.

4.1.2 Proof strategy for affine dispersers

We now give a brief description of the basic proof strategy that we use to prove that a function is an affine disperser. We demonstrate the steps involved in the special case of the function f defined in (4.1) for the case of $p = 2$ and $\pi(y) = \mathrm{Tr}(y)$ (where $\mathrm{Tr} : \mathbb{F}_{2^m} \to \mathbb{F}_2$ is the Trace map). We will show in Theorem 4.2.4 that if m is odd (so that $\mathbb{F}_{2^m}$ has no proper subfields of size $2^{m/2}$), then for any affine space $\mathcal{A} \subseteq \mathbb{F}_{2^m}$ of dimension $\geq 2m/5 + \Omega(1)$, we have $\{\mathrm{Tr}(a^7) \mid a \in \mathcal{A}\} = \mathbb{F}_2$.

1. **Reduce to showing that a certain polynomial h is not a constant polynomial:** We first parameterize the affine space $\mathcal{A}$ using subspace polynomials. Let $Q(X)$ be the image-subspace of $\mathcal{A}$, so that $\mathcal{A} = \{Q(x) : x \in \mathbb{F}_{2^m}\}$. In terms of the polynomial $Q(X)$, we want to show that the composed map $f \circ Q : \mathbb{F}_{2^m} \to \mathbb{F}_2$ is non-constant. Let $h(X)$ be the polynomial $\mathrm{Tr}(Q(X)^7)$ mod $\langle X^{2^m} - X \rangle$, so that $h(x) = f(Q(x))$ for each $x \in \mathbb{F}_{2^m}$ (cf. Proposition 4.3.1). Thus to show that h is a nonconstant map, it suffices to show that $h(X)$ is a nonconstant polynomial. We do this in the next two steps of our proof strategy, by finding a monomial of positive degree that appears in $h(X)$ with a nonzero coefficient.

2. **Express the coefficients of h in terms of the coefficients of the subspace polynomials:** To show that h has a monomial of positive degree with a nonzero coefficient, it will be convenient to get an explicit expression for the coefficients themselves. Such an explicit expression can be obtained by direct substitution. In all the cases we consider, there is a good deal of structure in the resulting formulae. For example, for the polynomial we obtained while studying $f(x) = \mathrm{Tr}(x^7)$, we have the following lemma.

Lemma 4.1.1 *Let $Q(X) = \sum_{i=0}^{m-1} a_i X^{2^i}$. Let $h(X) = \mathrm{Tr}(Q(X)^7)$ mod $\langle X^{2^m} - X \rangle$. Then for distinct j, k, l, the coefficient of $X^{2^j+2^k+2^l}$ in $h(X)$ is given by the*

expression:

$$\sum_{r=0}^{m-1} \mathrm{Perm} \begin{pmatrix} a_{j-r} & a_{k-r} & a_{l-r} \\ a_{j-r-1}^2 & a_{k-r-1}^2 & a_{l-r-1}^2 \\ a_{j-r-2}^4 & a_{k-r-2}^4 & a_{l-r-2}^4 \end{pmatrix}^{2^r}, \tag{4.3}$$

where Perm *is the matrix permanent, and the subscripts of the a's are taken mod m.*

3. **Argue combinatorially that some coefficient of h must be nonzero:** Finally, we show that some positive degree monomial of h has a nonzero coefficient. Using the regular form of the coefficients of the polynomial h, for example as given in Lemma 4.1.1, and the structural results about the coefficients of subspace polynomials, this part of the argument reduces to the combinatorics of cyclic shifts on $\mathbb{Z}_m$. More to the point, we use our main structural lemma (Lemma 4.4.3) to prove that there is a choice of j, k, l such that *(i)* the matrix appearing in the first summand (corresponding to $r = 0$) in equation (4.3) is lower triangular with nonzero entries on its diagonal, hence its permanent is nonzero, whereas *(ii)* the matrices appearing in all other summands in equation (4.3) (corresponding to $r = 1, \ldots, m-1$) contain a zero column, hence have a zero permanent.

4.1.3 From affine dispersers to extractors

We believe that all constructions provided here are affine extractors, not merely dispersers. We can prove this only for our simplest construction, that described in (4.1). This proof goes via a general theorem saying that *every* function of $\mathbb{F}_p$-degree 3 that is an affine disperser for dimension d, is also an affine extractor for dimension slightly higher than d (with the bias decreasing to 0 as the dimension increases). The function described in (4.1) is of this form (cf. Proposition 4.3.2) whereas that appearing in (4.2) is already of $\mathbb{F}_p$-degree 4 and the other dispersers analyzed here have even higher degree.

4.2 Main results

In this section, we state our main results. We start by formally defining affine dispersers and extractors.

Definition 4.2.1 ($\mathbb{F}_p$-affine dispersers and extractors) *A function $f : \mathbb{F}_p^m \to \mathbb{F}_p$ is an* **$\mathbb{F}_p$-affine disperser** *for dimension d if for every affine subspace $S \subseteq \mathbb{F}_p^m$ of dimension at least d, we have $|f(A)| > 1$.*

A function $f : \mathbb{F}_p^m \to \mathbb{F}_p$ is an **$\mathbb{F}_p$-affine ϵ-extractor** *for dimension d if for every affine subspace $S \subseteq \mathbb{F}_p^m$, if x is picked uniformly at random from S, the statistical distance of $f(x)$ from the uniform distribution on $\mathbb{F}_p$ is at most ϵ.*

We briefly indicate the relation between this definition and the more general setting. Following the derandomization literature, we will refer to a distribution over a domain $\mathcal{D}$ as a "source". A function $f : \mathcal{D} \to \mathcal{R}$ is said to be an **ϵ-extractor** for a set of sources $\mathcal{S}$ if, for every $S \in \mathcal{S}$, if x is picked according S, then the statistical distance of $f(x)$ from the uniform distribution on $\mathcal{R}$ is at most ϵ (ϵ is called the error-parameter of the extractor). The function f is a **disperser** for $\mathcal{S}$ if it is an ϵ-extractor for some $\epsilon < 1$. (This is equivalent to saying that f is nonconstant on the support of S for each source $S \in \mathcal{S}$).

A **d-dimensional affine source in $\mathbb{F}_p^m$** is the uniform distribution over some d-dimensional affine space. In this language, we see that a function $f : \mathbb{F}_p^m \to \mathbb{F}_p$ is an *$\mathbb{F}_p$-affine disperser (ϵ-extractor, respectively) for dimension d* if and only if it is a disperser (ϵ-extractor, respectively) for the set of d-dimensional affine sources in $\mathbb{F}_p^m$.

A more standard definition of a disperser, as appearing in, say, [Sha02], requires that for every d-dimensional affine source S, $f(\mathrm{supp}(S))$ equals the full range $\mathbb{F}_p$. Notice that for the case of $p = 2$ the two definitions match. All our constructions give $\mathbb{F}_p$-affine dispersers according to Definition 4.2.1. When p is clear from the context, we simply refer to them as *affine dispersers*.

4.2.1 Disperser for affine spaces of sublinear dimension

We begin by describing the function $f : \mathbb{F}_p^m \to \mathbb{F}_p$ which will prove to be a disperser over $\mathbb{F}_p$ for affine sources of sublinear dimension. The integers n, r and t are parameters of the construction to be specified later. As in [Bou07], we partition the m coordinates of an input x into r blocks $(x_1, \ldots, x_r)$ of n coordinates each (we assume n divides m by discarding a few field-elements, if necessary). We will pick n to be prime, so that $\mathbb{F}_{p^n}$ has no nontrivial subfields. Each block x_i is interpreted as an element of $\mathbb{F}_{p^n}$ by using an $\mathbb{F}_p$-linear isomorphism from $\mathbb{F}_p^n$ to $\mathbb{F}_{p^n}$. We then raise each x_i to a suitable distinct power and let y_i denote the result of this powering. Next, we apply the t^{th} symmetric polynomial to $y_1, \ldots, y_r$, and get $z \in \mathbb{F}_{p^n}$, where this polynomial is defined by

$$\mathsf{Sym}_r^t(Y_1, \ldots, Y_r) = \sum_{I \subseteq [r], |I|=t} \prod_{i \in I} Y_i.$$

Finally, we take a nontrivial $\mathbb{F}_p$-linear map $\pi : \mathbb{F}_{p^n} \to \mathbb{F}_p$, and output $\pi(z)$. We now formally state our main result.

Theorem 4.2.2 (Affine dispersers for sublinear dimension) *Given integer m fix parameters n, r, t as follows. Let n be the smallest prime bigger than $2 \cdot m^{3/5}$. Let $r = \lceil m/n \rceil$ and let $t = \lceil \sqrt{r} \rceil$. (We have $n \approx m^{3/5}, r \approx m^{2/5}$ and $t \approx m^{1/5}$.) Let $\phi : \mathbb{F}_p^m \to (\mathbb{F}_{p^n})^r$ be an injective $\mathbb{F}_p$-linear map, where $\phi(y) = (\phi_1(y), \ldots, \phi_r(y))$ and $\phi_i(y) \in \mathbb{F}_{p^n}$. Let $\pi : \mathbb{F}_{p^n} \to \mathbb{F}_p$ be a nontrivial $\mathbb{F}_p$-linear map. Then the function $f : \mathbb{F}_p^m \to \mathbb{F}_p$ defined by*

$$f(x) = \pi\left(\mathsf{Sym}_r^t\left((\phi_1(x))^{1+p}, (\phi_2(x))^{1+p+p^2}, \ldots, (\phi_r(x))^{1+p+p^2+\ldots+p^r}\right)\right) \tag{4.4}$$

is an affine disperser for dimension greater than $6m^{4/5}$, i.e., for all affine $\mathcal{A} \subseteq \mathbb{F}_p^m$ with $\dim(\mathcal{A}) > 6m^{4/5}$ we have $|f(\mathcal{A})| > 1$.

Notice f can be computed in polynomial time in p and m because Sym_r^t can be computed efficiently in the said time (using the Newton-Girard identities). From a

computational viewpoint our construction is more efficient than that of [Bou07] which for spaces of dimension ϵm required a running time of $m^{2^{\Omega(1/\epsilon)}}$.

The method by which we prove Theorem 4.2.2 is quite general and in the following subsections we show that a few natural variants of it can also be shown to be good affine dispersers and extractors in various settings.

4.2.2 Disperser for independent affine sources

Informally, we say a function $f : (\mathbb{F}_p^n)^t \to \mathbb{F}_p$ is a *disperser for independent affine sources* if on every set of affine spaces $\mathcal{A}_1, \ldots, \mathcal{A}_t \subseteq \mathbb{F}_p^n$ of sufficiently large dimensions, we have $|f(\mathcal{A}_1 \times \cdots \times \mathcal{A}_t)| > 1$. The following theorem presents an affine disperser for independent sources that works as long as the sum of dimensions is greater than n. The analysis of this independent source affine disperser turns out to play a crucial role in our proof Theorem 4.2.2.

In what follows, a *proper subfield* of $\mathbb{F}_{p^n}$ is a subfield $\mathbb{K}$ of size $< p^n$ and an *affine shift* of $\mathbb{K}$ is a set of the form $\{a \cdot s + b \mid s \in \mathbb{K}\}$ for some fixed $a, b \in \mathbb{F}_{p^n}$. (Notice that every one-dimensional $\mathbb{F}_p$-affine subspace of $\mathbb{F}_{p^n}, n > 1$ is an affine shift of the proper subfield $\mathbb{F}_p$.)

Theorem 4.2.3 (Disperser for independent affine sources) *Let $\pi : \mathbb{F}_{p^n} \to \mathbb{F}_p$ be a nontrivial $\mathbb{F}_p$-linear map. Consider the function $f : \mathbb{F}_{p^n}^t \to \mathbb{F}_p$ given by*

$$f(x_1, \ldots, x_t) = \pi\left(\prod_{i=1}^{t} x_i^{1+p}\right). \tag{4.5}$$

Let $\mathcal{A}_1, \cdots, \mathcal{A}_t \subseteq \mathbb{F}_{p^n}$ be $\mathbb{F}_p$-affine spaces of dimensions $d_1, \ldots, d_t$ respectively, where $\sum_{i=1}^{t}(d_i - 2) > n$. Suppose furthermore that no $\mathcal{A}_i$ is contained in an affine shift of a proper subfield of $\mathbb{F}_{p^n}$. Then $|f(\mathcal{A}_1 \times \cdots \times \mathcal{A}_t)| > 1$.

Remark The assumption that $\mathcal{A}_i$ is not contained in an affine shift of a proper subfield is necessary. Without it we could set $\mathcal{A}_i = \mathbb{K}$ for a proper subfield $\mathbb{K}$, and select some nontrivial π such that the resulting function f is constant on $\mathcal{A}_1 \times \ldots \times \mathcal{A}_t$.

Remark A result of Hou, Leung and Xiang [HLX02] implies the following statement (cf. [DG09]). Let $\mathcal{A}_1, \ldots, \mathcal{A}_t \subseteq \mathbb{F}_{p^n}$ be affine spaces of dimensions $d_1, \ldots, d_t$ respectively and none are contained in affine shifts of proper subfields. Then

$$\dim\left(\operatorname{span}\left\{\prod_i x_i \;\middle|\; x_i \in \mathcal{A}_i\right\}\right) \geq \min\left\{n, \sum_i (d_i - 1)\right\}.$$

So if $\sum(d_i - 1) \geq n$ then $\pi\left(\prod_{i=1}^t x_i\right)$ is nonconstant on $\mathcal{A}_1 \times \cdots \times \mathcal{A}_t$. The proof technique of [HLX02] differs significantly from ours and it is not clear how to derive one result from the other.

4.2.3 Univariate dispersers

Our next set of results is a pair of constructions based on univariate polynomials. We treat our input $x \in \mathbb{F}_p^n$ as a single element of the field $\mathbb{F}_{p^n}$ by using any $\mathbb{F}_p$-linear isomorphism between $\mathbb{F}_p^n$ and $\mathbb{F}_{p^n}$. We raise x to a suitable power and map the result to $\mathbb{F}_p$ using any nontrivial $\mathbb{F}_p$-linear map. The first construction will be shown in the next subsection to be an extractor for dimension greater than $2n/5$ and the second works for lower dimension $(n/3)$ but we cannot show that it is an extractor (cf. Conjecture 4.2.6). We call the next construction "cubic", and the one that follows "quartic", because the relevant functions f, when viewed as having domain $(\mathbb{F}_p)^n$, are computed by polynomials of degree 3 and 4 respectively (cf. the first bullet of Proposition 4.3.2).

Theorem 4.2.4 (Univariate cubic affine disperser) *Let $\pi : \mathbb{F}_{p^n} \to \mathbb{F}_p$ be a nontrivial $\mathbb{F}_p$-linear map. The function $f : \mathbb{F}_{p^n} \to \mathbb{F}_p$ given by*

$$f(x) = \pi\left(x^{1+p+p^2}\right)$$

is a disperser for the set of affine spaces of dimension greater than $\frac{2n}{5} + 10$ that are not contained in an affine shift of a proper subfield of $\mathbb{F}_{p^n}$.

In particular, if n is odd, then f is an affine disperser for dimension $\frac{2n}{5} + 10$.

Theorem 4.2.5 (Univariate quartic affine disperser) *Let $\pi : \mathbb{F}_{p^n} \to \mathbb{F}_p$ be a nontrivial $\mathbb{F}_p$-linear homomorphism. The function $f : \mathbb{F}_{p^n} \to \mathbb{F}_p$ given by*

$$f(x) = \pi\left(x^{1+p+p^2+p^3}\right)$$

is a disperser for the set of affine spaces of dimension greater than $\frac{n}{3} + 10$ that are not contained in an affine shift of a proper subfield of $\mathbb{F}_{p^n}$.

In particular, if n is odd, then f is an affine disperser for dimension $\frac{n}{3} + 10$.

We believe that the dimension bound in the above pair of theorems is not tight. In particular, we think the cubic construction of Theorem 4.2.4 should be a disperser for dimension $> n/3$ and the quartic construction of Theorem 4.2.5 should work for dimension $> n/4$. In fact, we believe in the stronger conjecture stated next.

Conjecture 4.2.6 (Univariate extractors) *For every prime p and integer k there exists an integer $c = c(p,k)$ and constant $\epsilon = \epsilon(p,k) > 0$ such that the following holds for all sufficiently large n. Let $\pi : \mathbb{F}_{p^n} \to \mathbb{F}_p$ be a nontrivial $\mathbb{F}_p$-linear map. Let $s_k = \sum_{i=0}^{k} p^i$. The function $f_k : \mathbb{F}_{p^n} \to \mathbb{F}_p$ given by*

$$f_k(x) = \pi\left(x^{s_k}\right)$$

is an $\exp(-\epsilon d)$-extractor for the set of affine spaces of dimension greater than $\left(\frac{n}{k} + c\right) + d$ that are not contained in an affine shift of a proper subfield of $\mathbb{F}_{p^n}$.

4.2.4 A cubic affine disperser is an affine extractor

Our final set of results shows that any *cubic* function that is a disperser for dimension d, is an $\epsilon(d')$-extractor for dimension $d + d'$, where $\epsilon(d')$ goes to 0 as d' increases.

Theorem 4.2.7 (Cubic affine dispersers are affine extractors) *There exists a universal constant $\epsilon > 0$ such that the following holds. Let $f : \mathbb{F}_p^m \to \mathbb{F}_p$ be computed by a cubic polynomial. If f is an affine disperser for dimension d_0 then f is an affine $O(d^{-\epsilon})$-extractor for dimension $d_0 + \hat{d}$.*

Using the cubic construction from Theorem 4.2.4, the previous theorem implies the following affine extractor.

Corollary 4.2.8 (Univariate cubic affine extractor) *There exists a universal constant $\epsilon > 0$ such that the affine disperser f defined in Theorem 4.2.4 is an affine $O(d^{-\epsilon})$-extractor for dimension $\left(\frac{2n}{5} + 10\right) + d$.*

The method of proof of Theorem 4.2.7, restated next, is very different from what we use in the rest of this paper. It relies on an *energy-increment* argument and Fourier analysis. Because the methods are unrelated to the main theme of this thesis, we omit the proof and refer the reader to the paper.

Remark Recent work of [HS09] gives a better bound on the error-parameter of f stated in 4.2.7. They show a bound of $\exp(-d^{\epsilon})$ on the error-parameter for some universal constant $\epsilon > 0$.

Counting arguments show that there exist cubic functions that are dispersers for affine spaces of dimension as small as $O(\sqrt{n})$. Given Theorem 4.2.7, this implies that one way to get affine extractors for sublinear dimension is to find an explicit cubic affine disperser that works for the same dimension bound.

Unfortunately, *quartic* affine dispersers are not necessarily affine extractors for comparable dimension. So, although we believe the quartic construction of Theorem 4.2.5 is an affine extractor (cf. Conjecture 4.2.6), a proof of this conjecture will have to rely on the particular algebraic structure of this quartic function.

Organization of this chapter The next section introduces some preliminaries. In Section 4.4 we establish some results about subspace polynomials that we will need for the following sections. Of particular importance are *(i)* the Main Structural Lemma 4.4.3 which connects the fact that a subspace is not an affine shift of a subfield to the zero-nonzero pattern of the coefficients of its image-subspace polynomial, and *(ii)* Lemma 4.4.6 which is used to show that our constructions, when restricted to a subspace of sufficiently large dimension, are polynomials of positive degree.

The proofs of our main results go in increasing order of complexity. In Section 4.5 we discuss our univariate constructions, proving Theorems 4.2.4 and 4.2.5. In Section 4.6 we analyze the disperser for independent sources and prove Theorem 4.2.3. In Section 4.7 we analyze our construction for sublinear dimension and prove Theorem 4.2.2. Together, Sections 2.4, Sections 4.4 4.6, and 4.7 contain a complete proof of Theorem 4.2.2. We conclude with some open problems.

4.3 Preliminaries

In this section we build up some preliminaries on polynomials, and recall some notions related to $\mathbb{F}_p$-degree.

We will use capital letters such as X_i are used for formal variables, and small letters such as x_i are used for field-elements.

For a polynomial $h(X_1, \ldots, X_r) \in \mathbb{F}_{p^n}[X_1, \ldots, X_r]$, abusing notation we define

$$h(X_1, \ldots, X_r) \mod \langle (X_i^{p^n} - X_i)_{i \in [r]} \rangle$$

to be the unique polynomial congruent to $h(X_1, \ldots, X_n) \mod \langle (X_i^{p^n} - X_i)_{i \in [r]} \rangle$ of degree $< p^n$ in each variable. Equivalently, h' is the polynomial obtained by starting with h and repeatedly replacing, for each i, every occurrence of $X_i^{p^n}$ by X_i. The following proposition, stated without proof, will be used repeatedly in our arguments.

Proposition 4.3.1 *Let $h(X_1, \ldots, X_r) \in \mathbb{F}_{p^n}[X_1, \ldots, X_r]$. Let*

$$h'(X_1, \ldots, X_r) = h(X_1, \ldots, X_r) \mod \langle (X_i^{p^n} - X_i)_{i \in [r]} \rangle.$$

Then for any $x \in \mathbb{F}_{p^n}^r$ we have $h(x) = h'(x)$.

Consequently, $|h(\mathbb{F}_{p^n}^r)| > 1$ if and only if $h'(X_1, \ldots, X_r)$ is a polynomial of degree greater than 0.

We recall some notions from Chapter 2, instantiated in a slightly simpler form for our purposes. For a nonnegative integer i, let $\mathrm{wt}_p(i)$ denote the sum of the digits of

i in the base-p representation. If $m(X_1, \ldots, X_t) \in \mathbb{F}_{p^n}[X_1, \ldots, X_t]$ is the monomial $\prod_{i=1}^{t} X_i^{\beta_i}$, we define the $\mathbb{F}_p$-*degree* of m in the variable X_i to be $\mathrm{wt}_p(\beta_i)$. We define the $\mathbb{F}_p$-*degree* of the monomial $\mathcal{M}$ to be the sum of the $\mathbb{F}_p$-degrees of $\mathcal{M}$ in each variable X_i. We then define the $\mathbb{F}_p$-degree of a polynomial to be the maximum $\mathbb{F}_p$-degree of any of its monomials.

Proposition 4.3.2 *Let $P(X_1, \ldots, X_t), Q(X_1, \ldots, X_t)$ be polynomials in $\mathbb{F}_{p^n}[X_1, \ldots, X_t]$ with $\mathbb{F}_p$-degrees $d_1, d_2 < n$ respectively. Let $\phi = (\phi_1, \ldots, \phi_t) : \mathbb{F}_p^{n \cdot t} \to \mathbb{F}_{p^n}^t$ be an $\mathbb{F}_p$-linear isomorphism. and $\pi : \mathbb{F}_{p^n} \to \mathbb{F}_p$ be an $\mathbb{F}_p$-linear map. Then*

- *Let $f = (f_1, \ldots, f_n) : \mathbb{F}_p^{n \cdot t} \to \mathbb{F}_p$ be given by $f(x) = \pi\left(P(\phi_1(x), \ldots, \phi_t(x))\right)$. Then f is computed by a polynomial $P' \in \mathbb{F}_p[Y_1, \ldots, Y_{n \cdot t}]$ of total degree at most d_1.*

- *The $\mathbb{F}_p$-degree of $P(X_1, \ldots, X_t) \cdot Q(X_1, \ldots, X_t)$ is at most $d_1 + d_2$.*

- *The $\mathbb{F}_p$-degree of $P(X_1, \ldots, X_t)^{p^r} \mod \langle (X_i^{p^n} - X_i)_{i \in [t]} \rangle$ equals d_1.*

We recall one final basic fact about finite field extensions — that $\mathbb{F}_p$-linear maps from $\mathbb{F}_{p^n}$ to $\mathbb{F}_p$ are computed by trace maps.

Proposition 4.3.3 *Let* $\mathrm{Tr}(Y) = \sum_{i=0}^{n-1} Y^{p^i}$ *be the* trace map *from $\mathbb{F}_{p^n}$ to $\mathbb{F}_p$. For every $\mathbb{F}_p$-linear map $\pi : \mathbb{F}_{p^n} \to \mathbb{F}_p$ there exists $\mu = \mu_\pi \in \mathbb{F}_{p^n}$ such that for all $x \in \mathbb{F}_{p^n}$ we have*

$$\pi(x) = \mathrm{Tr}(\mu \cdot x).$$

Furthermore, π is trivial if and only if $\mu = 0$.

4.4 Results on subspace polynomials

Preliminary material on subspace polynomials can be found in Section 2.4.

4.4.1 The main structural lemma

In this section, we prove our main structural lemma on the coefficients of subspace polynomials.

We begin with some basic facts about finite fields. In what follows, let $\overline{\mathbb{F}}_p$ denote the algebraic closure of $\mathbb{F}_p$.

Claim 4.4.1 *Let $k > 1$, and suppose $a, c \in \mathbb{F}_{p^n}$ are such that $a^{p^k} - ca = 0$. Then, letting b be any $(p^k - 1)$-th root of c in $\overline{\mathbb{F}}_p$, we have $a \in b \cdot \mathbb{F}_{p^k}$.*

Proof If $a = 0$ then the claim is trivial. Otherwise, we have $a^{p^k} = ca$, and hence $a^{p^k - 1} = c$. Thus $(a/b)^{p^k - 1} = 1$, which implies that $a/b \in \mathbb{F}_{p^k}$. ■

Claim 4.4.2 *For linearized polynomial $Q(X) = \sum_{j=0}^{n-1} a_j X^{p^j} + \hat{a} \in \mathbb{F}_{p^n}[X]$ and integer t, we have*

$$(Q(X))^{p^t} \pmod{X^{p^n} - X} \equiv \sum_{j=0}^{n-1} \left(a_{((j-t) \mod n)}\right)^{p^t} X^{p^j} + \hat{a}^{p^t}.$$

The proof follows by direct expansion, using the $\mathbb{F}_p$-linearity of the map $Z \longmapsto Z^{p^t}$.

We now state and prove our main structural lemma about the zero/nonzero pattern of consecutive coefficients of subspace polynomials.

Lemma 4.4.3 (Main structural lemma for subspace polynomials) *Let L be a d-dimensional linear subspace in $\mathbb{F}_{p^n}$. Let $Q_L(X) = \sum_{j=0}^{n-1} a_j X^{p^j}$ be the image-subspace polynomial of L.*

1. *For any integer r and set $J = \{(r + j) \mod n \mid j = 0, \ldots, d-1\}$ of d consecutive indices in $\mathbb{Z}_n$, there is some $j \in J$ with $a_j \neq 0$. In particular, a_0 and a_{n-d} are nonzero.*

2. *Suppose that L is not contained in any constant multiple of a proper subfield of $\mathbb{F}_{p^n}$, i.e. $L \not\subseteq \beta \cdot \mathbb{F}_{p^k}$ for any $\beta \in \mathbb{F}_{p^n}$ and any $\mathbb{F}_{p^k} \subset \mathbb{F}_{p^n}$. Then for any integer $r \neq n - d + 1$ and set $J = \{(r + j) \mod n \mid j = 0, \ldots, d-2\}$ of $d - 1$ consecutive indices in $\mathbb{Z}_n$, there is some $j \in J$ with $a_j \neq 0$.*

Proof For the first part, suppose $a_j = 0$ for all $j \in J$. Note that by Lemma 2.4.4, Q_L has p^{n-d} distinct roots in $\mathbb{F}_{p^n}$. Let $Q'(X) := Q_L(X)^{p^{n-(r+d)}} \mod X^{p^n} - X$. Then, by Claim 4.4.2 we conclude $Q'(X) = \sum_{j=0}^{n-1} a_{j+r+d-n}^{p^{n-(r+d)}} X^{p^j}$. Now for any $j \in [n-d, n-1]$, we have $a_{j+r+d-n} = 0$ by assumption, and thus $Q'(X)$ is of degree at most p^{n-d-1}. In addition, by Proposition 4.3.1, $Q'(\alpha) = Q_L(\alpha)^{p^{n-(r+d)}} = 0$ for every $\alpha \in \mathbb{F}_{p^n}$ satisfying $Q_L(\alpha) = 0$, and hence Q' has at least p^{n-d} roots. This is a contradiction.

In particular, since by definition $a_{n-d+1}, \ldots, a_{n-1}$ forms a sequence of $d-1$ consecutive coefficients that are all zero, we conclude both a_{n-d} and a_0 must be nonzero.

For the second part, suppose $a_j = 0$ for all $j \in J$. Again, by Lemma 2.4.4, Q_L has p^{n-d} distinct roots in $\mathbb{F}_{p^n}$. Let $k = n - (r+d) + 1$ (note that $0 < k < n$). Then as above the polynomial $Q'(X) := Q_L(X)^{p^k} \mod X^{p^n} - X$ is nonzero of degree at most p^{n-d}. In addition, $Q'(\alpha) = Q_L(\alpha)^{p^k} = 0$ for every $\alpha \in \mathbb{F}_{p^n}$ for which $Q_L(\alpha) = 0$. As Q' and Q_L are of the same degree p^{n-d}, there is a constant $c \in \mathbb{F}_{p^n}$ such that $Q'(X) - cQ_L(X)$ is of degree at most p^{n-d-1} and vanishes on the p^{n-d} roots of $Q_L(X)$. Thus the polynomial $Q'(X) - cQ_L(X)$ is identically zero. Recalling the definition of $Q'(X)$, have just showed that $Q_L(X)^{p^k} - cQ_L(X) = 0 \mod X^{p^n} - X$. Thus for each $\alpha \in \mathbb{F}_{p^n}$, we have $Q_L(\alpha)^{p^k} - cQ_L(\alpha) = 0$. Now, since the image of Q_L is L, by Claim 4.4.1 we conclude that $L \subseteq b \cdot \mathbb{F}_{p^k}$ (where $b \in \overline{\mathbb{F}}_p$ is a $p^k - 1$-th root of c). This almost gives the desired contradiction, but for the possibility that $b \notin \mathbb{F}_{p^n}$, and that $\mathbb{F}_{p^k}$ may not be a subfield of $\mathbb{F}_{p^n}$.

Let $\beta \in L \setminus \{0\}$. For any $\alpha \in L$, we have $\alpha/\beta \in (b \cdot \mathbb{F}_{p^k})/(b \cdot \mathbb{F}_{p^k})$, and hence $\alpha/\beta \in \mathbb{F}_{p^k}$. Thus $\beta^{-1} \cdot L \subseteq \mathbb{F}_{p^k} \cap \mathbb{F}_{p^n} = \mathbb{F}_{p^{(k,n)}}$, where $(k,n) = \gcd(k,n)$. Thus $L \subseteq \beta \cdot \mathbb{F}_{p^{(k,n)}}$, contradicting the hypothesis on L. ■

4.4.2 Coefficients of products of subspace polynomials

In our subsequent arguments, we will need time and again to prove that a certain polynomial P, which is the trace of products of linearized polynomials reduced mod $\langle (X_i^{p^n} - X_i)_{i \in [r]} \rangle$, is not a constant. In this subsection we describe a lemma that will allow us to argue such statements by showing that a well-chosen monomial

of P has a nonzero coefficient. We start with a definition.

Definition 4.4.4 (Associated matrix and its zero-one indicator matrix) *For a linearized polynomial $Q(X) = \sum_{i=0}^{n-1} a_i X^{p^i}$ over $\mathbb{F}_{p^n}$, we define its* associated matrix *$M_Q \in \mathbb{F}_{p^n}^{\{0,\ldots,n-1\}\times\{0,\ldots,n-1\}}$ by setting the (i,j)-entry of M_Q to be $(a_{j-i})^{p^i}$, where both rows and columns are indexed by $\{0,1,\ldots,n-1\}$ and index arithmetic, as well as powers of p are computed modulo n. Explicitly, M_Q is the following matrix*

$$\begin{pmatrix} a_0 & a_1 & a_2 & \ldots & \ldots & a_{n-1} \\ (a_{n-1})^p & (a_0)^p & (a_1)^p & \ldots & \ldots & (a_{n-2})^p \\ (a_{n-2})^{p^2} & (a_{n-1})^{p^2} & (a_0)^{p^2} & \ldots & \ldots & (a_{n-3})^{p^2} \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ (a_1)^{p^{n-1}} & (a_2)^{p^{n-1}} & (a_3)^{p^{n-1}} & \ldots & \ldots & (a_0)^{p^{n-1}} \end{pmatrix}.$$

For $a_i \in \mathbb{F}_{p^n}$ let a'_i indicate whether a_i is zero, i.e., $a'_i = 0$ if $a_i = 0$ and otherwise $a'_i = 1$. Similarly, let $M' = M'_Q$ denote the zero-one indicator matrix of M_Q. The (i,j)-entry of this matrix is a'_{j-i}, or, in other words, the (i,j)-entry of M' indicates whether the (i,j)-entry of M is nonzero.

The use of the associated matrix is captured by the following claim. The proof of the claim (which is omitted) follows immediately from Claim 4.4.2.

Claim 4.4.5 *The (i,j)-entry of M_Q is the coefficient of X^{p^j} in the linearized polynomial $(Q(X))^{p^i} \mod X^{p^n} - X$.*

To state the main lemma of this subsection we need the following notation. For A, B nonempty subsets of $\{0,\ldots,n-1\}$ let $M[A,B]$ be the minor corresponding to rows A and columns B. For an integer r, let $B + r = \{s + r \mod n \mid s \in B\}$.

Lemma 4.4.6 *Let $\mu \in \mathbb{F}_{p^n} \setminus \{0\}$. Let $A_1,\ldots,A_t,B_1,\ldots,B_t \subseteq \{0,\ldots,n-1\}$ satisfy $|A_i| = |B_i| > 0$ for $i = 1,\ldots,t$. Let $\alpha_i = \sum_{j\in A_i} p^j, \beta_i = \sum_{k\in B_i} p^k$. Let $Q_1(X_1),\ldots,Q_t(X_t)$ be linearized polynomials with associated matrices $M_1,\ldots,M_t$ and zero-one indicator matrices $M'_1,\ldots,M'_t$ respectively.*

The coefficient $c_{\mathcal{M}}$ *of the monomial* $\mathcal{M} = \prod_{i=1}^{t} X_i^{\beta_i}$ *in*

$$R(X_1, \ldots, X_t) = \mathrm{Tr}\left(\mu \prod_{i=1}^{t} (Q_i(X_i))^{\alpha_i}\right) \mod \langle (X_i^{p^n} - X_i)_{i\in[t]} \rangle \tag{4.6}$$

is given by the expression

$$c_{\mathcal{M}} = \sum_{r=0}^{n-1} \mu^{p^r} \cdot \prod_{i=1}^{t} \mathrm{Perm}\left(M_i[A_i + r, B_i]\right) = \sum_{r=0}^{n-1} \mu^{p^r} \cdot \prod_{i=1}^{t} \mathrm{Perm}\left(M_i[A_i, B_i - r]\right)^{p^r}. \tag{4.7}$$

Proof Notice

$$\mathrm{Tr}\left(\mu \prod_{i=1}^{t} (Q_i(X_i))^{\alpha_i}\right) = \sum_{r=0}^{n-1} \mu^{p^r} \cdot \prod_{i=1}^{t} (Q_i(X_i))^{\alpha_i \cdot p^r}.$$

Thus, $c_{\mathcal{M}}$ is a sum of n elements, where the rth element, denoted $c_{\mathcal{M}}^{(r)}$, is the coefficient of m in the rth summand in the right hand side above. We can break $c_{\mathcal{M}}^{(r)}$ further into μ^{p^r} times a product of t terms, where the ith term is the coefficient of $X_i^{\beta_i}$ in $(Q_i(X_i))^{\alpha_i \cdot p^r}$. So to prove the lemma it suffices to show that the coefficient of $X_i^{\beta_i}$ in $(Q_i(X_i))^{\alpha_i \cdot p^r}$ is $\mathrm{Perm}\left(M_i[A_i + r, B_i]\right)$.

Expand $(Q_i(X_i))^{\alpha_i \cdot p^r}$ as

$$\prod_{j\in A_i} (Q_i(X_i))^{p^{j+r}} = \prod_{j\in A_i + r} (Q_i(X_i))^{p^j}.$$

By assumption $|A_i| = |B_i|$ and expanding $X_i^{\beta_i}$ as $\prod_{k\in B_i} X_i^{p^k}$ we see that for every one-to-one mapping $h : B_i \to A_i$ we get a contribution to the coefficient of $X_i^{\beta_i}$ by picking $X_i^{p^k}$ from $(Q_i(X_i))^{p^{h(k)+r}}$, i.e., the coefficient of $X_i^{\beta_i}$ is (using Claim 4.4.5):

$$\sum_{h:B_i\to A_i, h \text{ one-to-one}} \prod_{k\in B_i} a_{i,k-(h(k)+r)}^{p^{h(k)+r}} = \mathrm{Perm}(M_i[A_i + r, B_i]).$$

This completes the proof of the lemma. ■

The above lemma gives us an explicit formula for the coefficients of a certain polynomials. The following remark describes the exact way in which this lemma gets

used to show that such a polynomial is nonzero.

Remark 4.4.7 *Keep the notation of the previous lemma. Suppose that the following two conditions hold:*

1. *$M'_1[A_1, B_1], \ldots, M'_t[A_t, B_t]$ are each, up to reordering of rows and columns, upper triangular with every diagonal entry nonzero.*

2. *For every $r \in \{1, \ldots, n-1\}$ there exists $i_r \in \{1, \ldots, t\}$ such that $M'_{i_r}[A_{i_r}, B_{i_r} - r]$ contains an all-zero column.*

Then the coefficient $c_{\mathcal{M}}$ of the monomial $\mathcal{M}$ in $R(X_1, \ldots, X_t)$ is nonzero.

Indeed, assumption 1 implies that the first summand on the right hand side of (4.7) is nonzero, because it is a product of permanents of upper triangular matrices with nonzero diagonal. Assumption 2 implies that all other summands are zero, because one matrix in the product has a zero permanent on account of its all-zero column.

4.5 Univariate constructions

In this section we prove our results about univariate dispersers. We start with the cubic affine disperser (in the next section, we will show that it is even an affine extractor).

4.5.1 Cubic affine disperser

Theorem 4.2.4 (Univariate cubic affine disperser, restated) *Let $\pi : \mathbb{F}_{p^n} \to \mathbb{F}_p$ be a nontrivial $\mathbb{F}_p$-linear map. The function $f : \mathbb{F}_{p^n} \to \mathbb{F}_p$ given by*

$$f(x) = \pi\left(x^{1+p+p^2}\right)$$

is a disperser for the set of affine spaces of dimension greater than $\frac{2n}{5} + 10$ that are not contained in an affine shift of a proper subfield of $\mathbb{F}_{p^n}$.

In particular, if n is odd, then f is an affine disperser for dimension $\frac{2n}{5} + 10$.

Proof We assume without loss of generality that $\dim(\mathcal{A}) = d = \lceil \frac{2n}{5} \rceil + 10$ (by replacing $\mathcal{A}$ with an arbitrary subspace of $\mathcal{A}$ of this dimension). By Proposition 4.3.3, we know that $\pi(x)$ is of the form $\mathrm{Tr}(\mu x)$ for some $\mu \in \mathbb{F}_{p^n} \setminus \{0\}$. Let $Q(X)$ be the image-subspace polynomial of $\mathcal{A}$, so that $\mathcal{A} = Q(\mathbb{F}_{p^n})$. Let

$$R(X) = \mathrm{Tr}(\mu \cdot Q(X)^{1+p+p^2}) \mod \langle X^{p^n} - X \rangle,$$

so that by Proposition 4.3.1, $R(x) = f(Q(x))$ for each $x \in \mathbb{F}_{p^n}$ and hence $R(\mathbb{F}_{p^n}) = f(\mathcal{A})$. The same proposition implies that to prove Theorem 4.2.5, it suffices to show that $R(X)$ has a monomial of positive degree, and this is what we shall do.

To find the desired monomial we start by invoking Lemma 4.4.6. Applying this lemma to our case we have $t = 1$ and we get a single linearized polynomial $Q_1(X_1) = Q(X)$. The set $A = A_i$ is $\{0, 1, 2\}$, which corresponds to the exponent $\alpha = \alpha_1 = p^0 + p^1 + p^2$. Thus, Lemma 4.4.6 reads in our case as follows.

Claim 4.5.1 *For $B = \{i, j, k\} \subseteq \{0, \ldots, n-1\}$ let $\beta = p^i + p^j + p^k$. The coefficient $c_{\mathcal{M}}$ of the monomial $\mathcal{M} = X^\beta$ in*

$$\mathrm{Tr}\left(\mu \cdot Q(X)^{p^0+p^1+p^2}\right) \pmod{X^{p^n} - X}$$

is given by

$$c_{\mathcal{M}} = \sum_{r=0}^{n-1} \mu^r \mathrm{Perm}\left(M[A, B-r]\right)^{p^r}. \tag{4.8}$$

By Remark 4.4.7, the above claim implies that in order to show that $R(X)$ is nonconstant, letting $M' = M'_Q$ be the zero-one indicator matrix of M_Q as defined in Definition 4.4.4, it suffices to find a $B \subseteq \{0, \ldots, n-1\}$ with $|B| = 3$, such that:

1. The matrix $M'[\{0, 1, 2\}, B]$ is, up to reordering of rows and columns, upper triangular with each diagonal entry nonzero.

2. For every $r \in \{1, \ldots, n-1\}$ the matrix $M'[\{0, 1, 2\}, B - r]$ contains an all-zero column.

We proceed to find such a B. Thus all the action is in the first 3 rows of the matrix M'.

To this end, we state a few useful properties of the coefficients of Q that all follow immediately from Lemma 4.4.3 and will be used later on in the proof. Notice that *(iv)* below follows via the second part of Lemma 4.4.3 from our assumption that $\mathcal{A}$ is not contained in an affine shift of a proper subfield of $\mathbb{F}_{p^n}$.

Claim 4.5.2 *Let $Q(X) = \sum_{i=0}^{n-1} a_i X^{p^i} + \hat{a}$ be the image-subspace polynomial of $\mathcal{A}$. Letting $d = \dim(\mathcal{A})$ we have (i) $d \geq \frac{2n}{5} + 10$, (ii) $a_0, a_{n-d} \neq 0$, (iii) $a_{n-d+1} = \ldots = a_{n-1} = 0$ and (iv) for every $0 \leq j \leq n-d$ there is at least one nonzero coefficient amongst $a_j, a_{j+1}, \ldots, a_{j+d-2}$.*

To further simplify notation, for $r_1 < r_2$ let $[r_1, r_2]$ denote the set of integers in the interval $[r_1, r_2]$. Let $I_0 = \{i \in [0, n-1] : a_i = 0\}$ denote the set of indices of the zero coefficients of Q and let $I_1 = [0, n-1] \setminus I_0$ be the set of indices of nonzero ones.

We show the existence of a set B satisfying properties 1 and 2 and break the proof into three cases according to the structure of I_0, I_1.

Case I — $I_1 \cap [n/5 - 15, 2n/5 + 7] \neq \emptyset$**:** Let $j \in I_1 \cap [n/5 - 15, 2n/5 + 7]$. We claim the set $B = \{0, j+1, n-d+2\}$ satisfies our pair of properties. Property 1 holds because

$$M'[\{0,1,2\}, \{0, j+1, n-d+2\}] = \begin{pmatrix} a'_0 & a'_{j+1} & a'_{n-d+2} \\ a'_{n-1} & a'_j & a'_{n-d+1} \\ a'_{n-2} & a'_{j-1} & a'_{n-d} \end{pmatrix} = \begin{pmatrix} a'_0 & a'_{j+1} & 0 \\ 0 & a'_j & 0 \\ 0 & a'_{j-1} & a'_{n-d} \end{pmatrix} = \begin{pmatrix} 1 & * & 0 \\ 0 & 1 & 0 \\ 0 & * & 1 \end{pmatrix}$$

The second equality holds because of Claim 4.5.2 *(i)* , *(ii)* .

We now argue property 2. We have

$$M'[A, B-r] = M'[\{0,1,2\}, \{n-r, j+1-r, n-d+2-r\}] = \begin{pmatrix} a'_{n-r} & a'_{j+1-r} & a'_{n-d+2-r} \\ a'_{n-r-1} & a'_{j-r} & a'_{n-d+1-r} \\ a'_{n-r-2} & a'_{j-1-r} & a'_{n-d-r} \end{pmatrix}$$

For $r \in [1, d-3]$ the first column of $M'[A, B-r]$ is seen to be zero because the set of indices appearing there is

$$\{n-r, n-r-1, n-r-2\} \subseteq [n-d+1, n-1] \subseteq I_0,$$

(the last inclusion follows from Claim 4.5.2 *(ii)*). Similarly, for $r \in [n-(d-3), n-1]$ the last column of $M'[A, B-r]$ is zero, since the set of indices appearing there,

$$\{n-d+2-r, n-d+1-r, n-d-r\} \subseteq [n-d+1, n-1] \subseteq I_0.$$

Finally, for the remaining $r \in [d-2, n-(d-2)] \subseteq [2n/5+8, 3n/5-8]$, using the fact that $j \in [n/5-15, 2n/5+7]$, we see that the middle column of $M'[A, B-r]$ is zero, since the set of indices appearing there,

$$\begin{aligned} \{j+1-r, j-r, j-1-r\} &\subseteq \left[\left(\frac{n}{5}-15\right)-1-\left(\frac{3n}{5}-8\right), \left(\frac{2n}{5}+7\right)+1-\left(\frac{2n}{5}+8\right)\right] \\ &\subseteq \left[\frac{3n}{5}-8, n-1\right] \subseteq I_0, \end{aligned}$$

where the last inclusion uses the bound on d which implies $3n/5-8 > n-d$. We conclude that property 2 also holds and the proof of the first case is complete.

Case II — $I_1 \cap [n/5-15, 2n/5+7] = \emptyset$**:** Let j_1 be the largest element in $[0, n/5-15] \cap I_1$ and let j_2 be the minimal element in $[2n/5+7, n-d] \cap I_1$. By Claim 4.5.2 *(iii)* we cannot have both $j_1 = 0$ and $j_2 = n-d$. Consider the following four intervals: $[0, j_1]$ (whose end points are in I_1), $[j_1+1, j_2-1]$ (which is contained int I_0), $[j_2, n-d]$ (whose end points are in I_1), and $[n-d+1, n-1]$ (which is contained in I_0). Denote the length of these intervals by $\alpha_1, \ldots, \alpha_4$ respectively. Notice the length of each of the zero intervals (α_2, α_4) is strictly greater than the length of the other two "nonzero" intervals. Moreover, by the assumption that $\mathcal{A}$ is not contained in an affine shift of a proper subfield, part 2 of Lemma 4.4.3 implies that $\alpha_4 > \alpha_2$. By assumption

$\alpha_1, \alpha_3 < n/5 - 10$. We summarize this for future reference by

$$d - 2 = \alpha_4 > \alpha_2 > \max\{\alpha_1, \alpha_3\} + 10. \tag{4.9}$$

There are two subcases,

Case II.a — $\alpha_1 \neq \alpha_3$: Assume without loss of generality $\alpha_1 > \alpha_3$. We claim that $B = \{0, n - d + 1, j_1 + 2\}$ satisfies our pair of properties. (The case of $\alpha_1 < \alpha_3$ can be seen to be identical by using the argument below to show $B = \{j_2, j_1 + 1, n - d + 2\}$ satisfies the said pair of properties.) Property 1 holds because

$$M'[\{0,1,2\}, \{0, n-d+1, j_1+2\}] = \begin{pmatrix} a'_0 & a'_{n-d+1} & a'_{j_1+2} \\ a'_{n-1} & a'_{n-d} & a'_{j_1+1} \\ a'_{n-2} & a'_{n-d-1} & a'_{j_1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & * & 1 \end{pmatrix}.$$

Regarding property 2, the key observation is that any nonzero shift will force either the first or the last column to be all zero. Indeed, the difference between the top left and bottom right indices of $M'[A, B - r = \{n - r, n - d + 1 - r, j_1 + 2 - r\}]$ is $j_1 = \alpha_1 - 1$, and the difference between any other pair of indices chosen one from each of the first and third columns is between j_1 and $j_1 + 4$. Thus, by (4.9) the only value of r such that both these columns are not entirely zero is $r = 0$. We conclude property 2 holds and the proof of this case is complete.

Case II.b — $\alpha_1 = \alpha_3$: In this case we claim that $B = \{0, j_2 + 1, j_1 + 2\}$ satisfies our pair of properties. Property 1 can be verified by inspection as in the previous two cases. Furthermore, since the first and last column in this case are identical to the first and last column in the previous case, the same argument as there shows that the only nonzero shift that has both these columns nonzero must be $r = n - j_2$. We

get the following matrix

$$M'[\{0,1,2\}+(n-j_2), B=\{0, j_2+1, j_1+2\}] = \begin{pmatrix} a'_{j_2} & a'_{2j_2+1} & a'_{n-d+2} \\ a'_{j_2-1} & a'_{2j_2} & a'_{n-d+1} \\ a'_{j_2-2} & a'_{2j_2-1} & a'_{n-d} \end{pmatrix} = \begin{pmatrix} 1 & * & 0 \\ 0 & a'_{2j_2} & 0 \\ 0 & * & 1 \end{pmatrix}.$$

The only way the matrix above can be nonzero is if $a'_{2j_2} \neq 0$. Since $j_2 > 2n/5$ the only way this can happen is to have $j_2 \geq n/2$. But in this case we get $\alpha_1 + \alpha_2 \geq \alpha_3 + \alpha_4$ which contradicts (4.9). We conclude the above matrix has permanent 0 and property 2 holds. This completes the proof for the final case and Theorem 4.2.4 follows. ■

4.5.2 Quartic affine disperser

Theorem 4.2.5 (Univariate quartic affine disperser, restated) *Let $\pi : \mathbb{F}_{p^n} \to \mathbb{F}_p$ be a nontrivial $\mathbb{F}_p$-linear homomorphism. The function $f : \mathbb{F}_{p^n} \to \mathbb{F}_p$ given by*

$$f(x) = \pi\left(x^{1+p+p^2+p^3}\right)$$

is a disperser for the set of affine spaces of dimension greater than $\frac{n}{3} + 10$ that are not contained in an affine shift of a proper subfield of $\mathbb{F}_{p^n}$.

In particular, if n is odd, then f is an affine disperser for dimension $\frac{n}{3} + 10$.

Proof The proof is similar to that of Theorem 4.2.4. Let μ be as before. Let $Q(X)$ be the image-subspace polynomial of $\mathcal{A}$, so that $\mathcal{A} = Q(\mathbb{F}_{p^n})$. Let

$$R(X) = \mathrm{Tr}(\mu \cdot Q(X)^{1+p+p^2+p^3}) \mod \langle X^{p^n} - X \rangle.$$

As in the case of the previous proof, it is sufficient to prove the existence of a quadruple $B \subseteq \{0, \ldots, n-1\}$ which satisfies the conditions of Remark 4.4.7. We get started by adapting Lemma 4.4.6 to our present situation.

Claim 4.5.3 *For $B = \{i_1, \ldots, i_4\} \subseteq \{0, \ldots, n-1\}$ let $\beta = p^{i_1} + p^{i_2} + p^{i_3} + p^{i_4}$. The*

coefficient $c_{\mathcal{M}}$ *of the monomial* $\mathcal{M} = X^{\beta}$ *in*

$$\mathrm{Tr}\left(\mu \cdot Q(X)^{p^0+p^1+p^2+p^3}\right) \pmod{X^{p^n} - X}$$

is given by

$$c_{\mathcal{M}} = \sum_{r=0}^{n-1} \mu^{p^r} \mathrm{Perm}\left(M[A, B-r]\right)^{p^r}. \tag{4.10}$$

As in Remark 4.4.7, letting $M' = M'_Q$ (as defined in Definition 4.4.4), we seek $B \subseteq \{0, \ldots, n-1\}$ with $|B| = 4$ such that:

1. The matrix $M'[\{0,1,2,3\}, B]$ is, up to reordering of rows and columns, upper triangular with a nonzero diagonal.

2. For every $r \in \{1, \ldots, n-1\}$ the matrix $M'[\{0,1,2,3\}, B-r]$ contains an all-zero column.

Having found such a B, the above claim lets us conclude that the polynomial $R(X)$ defined above is nonconstant.

As in the analysis of the cubic affine disperser, we begin by stating a few useful properties of the coefficients of Q that all follow immediately from Lemma 4.4.3 and will be used later on in the proof. Notice that *(iv)* below follows via the second part of Lemma 4.4.3 from our assumption that $\mathcal{A}$ is not contained in an affine shift of a proper subfield of $\mathbb{F}_{p^n}$.

Claim 4.5.4 *Let* $Q(X) = \sum_{i=0}^{n-1} a_i X^{p^i} + \hat{a}$ *be the image-subspace polynomial of* $\mathcal{A}$. *Letting* $d = \dim(\mathcal{A})$ *we have (i)* $d \geq \frac{n}{3} + 10$, *(ii)* $a_0, a_{n-d} \neq 0$, *(iii)* $a_{n-d+1} = \ldots = a_{n-1} = 0$ *and (iv) for every* $0 \leq j \leq n-d$ *there is at least one nonzero coefficient amongst* $a_j, a_{j+1}, \ldots, a_{j+d-1}$.

We use the notation introduced in the proof of Theorem 4.2.4 in the previous subsection. Recalling the definition of I_0, I_1, notice that *(ii)* implies $\{0, n-d\} \in I_1$, *(iii)* implies $[n-d+1, n-1] \subseteq I_0$ and *(iv)* implies $I_1 \cap [j, j+d-1] \neq \emptyset$.

As stated earlier, we show that a set B satisfying part 2 of Claim 4.5.3 exists, thereby proving Theorem 4.2.5. Our proof is divided into three cases according to the structure of I_0, I_1.

Case I — $I_1 \cap [n/3-14, n/3+4] \neq \emptyset$: Let $j \in I_1 \cap [n/3-14, n/3+4]$. We claim $B = \{0, 1, j+2, n-d+3\}$ satisfies the two properties. Property 1 holds because

$$M'[A,B] = \begin{pmatrix} a'_0 & a'_1 & a'_{j+2} & a'_{n-d+3} \\ a'_{n-1} & a'_0 & a'_{j+1} & a'_{n-d+2} \\ a'_{n-2} & a'_{n-1} & a'_j & a'_{n-d+1} \\ a'_{n-3} & a'_{n-2} & a'_{j-1} & a'_{n-d} \end{pmatrix} = \begin{pmatrix} 1 & * & * & 0 \\ 0 & 1 & * & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & * & 1 \end{pmatrix}, \quad \text{where } a'_i = \begin{cases} 0 & a_i = 0 \\ 1 & a_i \neq 0 \end{cases}.$$

Regarding property 2, consider

$$M'[A, B-r] = \begin{pmatrix} a'_{n-r} & a'_{n+1-r} & a'_{j+2-r} & a'_{n-d+3-r} \\ a'_{n-1-r} & a'_{n-r} & a'_{j+1-r} & a'_{n-d+2-r} \\ a'_{n-2-r} & a'_{n-1-r} & a'_{j-r} & a'_{n-d+1-r} \\ a'_{n-3-r} & a'_{n-2-r} & a'_{j-1-r} & a'_{n-d-r} \end{pmatrix}.$$

For $r \in [1, d-4]$, the first column of $M'[A, B-r]$ is zero, since the set of indices appearing there

$$[n-3-r, n-r] \subseteq [n-d+1, n-1] \subseteq I_0$$

The last inclusion follows from Claim 4.5.4 *(iii)* . Similarly, for $r \in [n-(d-4), n-1]$ the last column of $M'[A, B-r]$ is zero, since

$$[n-d+3-r, n-d-r] \subseteq [n-d+1, n-1] \subseteq I_0.$$

Finally, for the remaining $r \in [d-3, n-(d-3)] \subseteq [n/3+7, 2n/3-7]$ the third column of $M'[A, B-r]$ is zero by selection of $j \in [n/3-14, n/3+4]$, since

$$[j-1-r, j+2-r] \subseteq [(n/3-15)-(2n/3-7), (n/3+6)-(n/3+7)] \subseteq [2n/3-8, n-1] \subseteq I_0,$$

where the last inclusion uses Claim 4.5.4 *(i)* . We conclude property 2 also holds and the proof of the first case is complete.

Case II — $I_1 \cap [n/3 - 14, n/3 + 4] = \emptyset$**:** As in the proof of Theorem 4.2.4, let j_1 be the largest element in $[0, n/3 - 15] \cap I_1$ and let j_2 be the minimal element in $[n/3 + 5, n - d] \cap I_1$. By Claim 4.5.4 *(iv)* we cannot have both $j_1 = 0$ and $j_2 = n - d$. Consider the following four intervals: $[0, j_1]$ (whose end points are in I_1), $[j_1 + 1, j_2 - 1]$ (which is contained int I_0), $[j_2, n - d]$ (whose end points are in I_1), and $[n - d + 1, n - 1]$ (which is contained in I_0). Denote the length of these intervals by $\alpha_1, \ldots, \alpha_4$ respectively. Notice $\alpha_4 = d - 2 > \alpha_1 + 10, \alpha_3 + 10$. There are two subcases.

Case II.a — $\alpha_1 \neq \alpha_3$**:** Assume without loss of generality $\alpha_1 > \alpha_3$. We claim that the set $B = \{0, j_2 + 1, n - d + 2, j_1 + 3\}$ satisfies both properties. (The case of $\alpha_1 < \alpha_3$ can be seen to be identical by using the argument below to show that $B = \{j_2, 1, j_1 + 2, n - d + 3\}$ satisfies our properties.) Property 1 holds because

$$M'[A, B] = \begin{pmatrix} a'_0 & a'_{j_2+1} & a'_{n-d+2} & a'_{j_1+3} \\ a'_{n-1} & a'_{j_2} & a'_{n-d+1} & a'_{j_1+2} \\ a'_{n-2} & a'_{j_2-1} & a'_{n-d} & a'_{j_1+1} \\ a'_{n-3} & a'_{j_2-2} & a'_{n-d-1} & a'_{j_1} \end{pmatrix} = \begin{pmatrix} 1 & * & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & * & 1 \end{pmatrix}.$$

Regarding property 2, the key observation is that any the difference between the top left and bottom right indices of $M'[A, B - r]$ — and this difference is independent of r — equals $j_1 = \alpha_1 - 1$, and similarly the difference between any other pair of indices chosen one from each of the first and third columns is between j_1 and $j_1 + 6$. Since $\alpha_4 - 10 > \alpha_1 > \alpha_3$ and $\alpha_2 \geq 20$ the only shifts r that make both the first and the last columns nonzero must satisfy $r \in [n - j_1, n - 10]$. This implies that the third column is zero (since this choice of r puts the indices of its entries in I_0), hence property 2 holds and the proof of this case is complete.

Case II.b — $\alpha_1 = \alpha_3$: We claim that $B = \{0, j_2+1, n-d+2, j_1+3\}$ satisfies both our properties. Indeed, property 1 holds by the reasoning of case II.a. By the same reasoning as in that case, the only nonzero shift r for which both the first and last columns are nonzero is the shift $r = n - j_2$ which gives

$$M'[A, B-(n-j_2)] = \begin{pmatrix} a'_{j_2} & a'_{2j_2+1} & a'_{n-d+j_2+2} & a'_{j_2+j_1+3} \\ a'_{j_2-1} & a'_{2j_2} & a'_{n-d+j_2+1} & a'_{j_2+j_1+2} \\ a'_{j_2-2} & a'_{2j_2-1} & a'_{n-d+j_2} & a'_{j_2+j_1+1} \\ a'_{j_2-3} & a'_{2j_2-2} & a'_{n-d+j_2-1} & a'_{j_2+j_1} \end{pmatrix} = \begin{pmatrix} 1 & * & * & 0 \\ 0 & a'_{2j_2} & a'_{n-d+j_2+1} & 0 \\ 0 & a'_{2j_2-1} & a'_{n-d+j_2} & 0 \\ 0 & * & * & 1 \end{pmatrix}.$$

The last column is calculated using $\alpha_1 = \alpha_3$ which implies $j_2 + j_1 = n - d$. Consider the middle 2×2 matrix on the right hand side above. The difference between the upper left and bottom right indices is $n-d+j_2-2j_2 = n-d-j_2 = j_1$ and that between the bottom left and upper right is $j_1 + 2$. Thus, we conclude $a'_{n-d+j_2+1} = a'_{2j_2-1} = 0$. Claim 4.5.4 *(iv)* , which relies on the fact that $\mathcal{A}$ is not contained in an affine shift of a proper subfield, implies that $\alpha_2 < \alpha_4$. Together with the assumption $\alpha_1 = \alpha_3$ we conclude $\alpha_1+\alpha_2 < \alpha_3+\alpha_4$. This implies $j_2 < n/2$ which, together with the assumption $j_2 > n/3$ gives us $n - d < 2j_2 < n$. This implies, via part *(i)* of Claim 4.5.4, that $a_{2j_2} = 0$ and the third column is all zero. This shows property 2.

Summing up, in each of the three cases above, we have shown the existence of a set B that satisfies both properties of part 2 of Claim 4.5.3. This implies $R(X)$ is nonzero and Theorem 4.2.5 follows. ■

4.6 Disperser for independent affine sources

In this section we prove Theorem 4.2.3, restated below. Although the analysis is simpler than what is involved in the proof of our main disperser for sublinear dimension (Theorem 4.2.2), the proof of the following theorem lies at the heart of the more complicated case which is discussed in the next section.

Theorem 4.2.3 (Disperser for independent affine sources, restated) *Let $\pi : \mathbb{F}_{p^n} \to \mathbb{F}_p$ be a nontrivial $\mathbb{F}_p$-linear map. Consider the function $f : \mathbb{F}_{p^n}^t \to \mathbb{F}_p$ given by*

$$f(x_1, \ldots, x_t) = \pi\left(\prod_{i=1}^{t} x_i^{1+p}\right). \tag{4.11}$$

Let $\mathcal{A}_1, \ldots, \mathcal{A}_t \subseteq \mathbb{F}_{p^n}$ be $\mathbb{F}_p$-affine spaces of dimensions $d_1, \ldots, d_t$ respectively, where each $\mathcal{A}_i$ is not contained in an affine shift of a proper subfield of $\mathbb{F}_{p^n}$. If $\sum_{i=1}^{t}(d_i - 2) > n$, then $|f(\mathcal{A}_1 \times \cdots \times \mathcal{A}_t)| > 1$.

Proof We follow the steps outlined in our strategy described in Section 4.1.2. First, we notice that

$$f(\mathcal{A}_1, \ldots, \mathcal{A}_t) = f(Q_1(\mathbb{F}_{p^n}), \ldots, Q_t(\mathbb{F}_{p^n}))$$

where $Q_i(X_i)$ is the image-subspace polynomial of $\mathcal{A}_i$. By Propositions 4.3.1, 4.3.3, in order to show $|f(\mathcal{A}_1 \times \cdots \times \mathcal{A}_t)| > 1$ it suffices to show that for any $\mu \in \mathbb{F}_{p^n} \setminus \{0\}$ the polynomial

$$R(X_1, \ldots, X_t) \stackrel{\text{def}}{=} \mathrm{Tr}\left(\mu \prod_{i=1}^{t} (Q_i(X_i))^{\alpha_i}\right) \mod \langle (X_i^{p^n} - X_i)_{i \in [t]} \rangle$$

contains a monomial of positive degree with nonzero coefficient. We use Lemma 4.4.6 to prove the existence of such a monomial and in the proof we rely on the structural properties of image-subspace polynomials given in Lemma 4.4.3.

The key step in our proof is given by the following theorem. We state a somewhat more general form than needed for the proof of Theorem 4.2.3. The added generality will be useful in the proof of Theorem 4.2.2. (The general form we refer to deals with large powers α_i whereas for Theorem 4.2.3 setting all α_i to $1 + p$ would be sufficient.)

Theorem 4.6.1 (Disperser for independent affine sources — Algebraic version) *Assume that $\mathcal{A}_1, \ldots, \mathcal{A}_t \subseteq \mathbb{F}_{p^n}$ are affine subspaces of dimensions $d_1, \ldots, d_t > 1$, none of which are contained in an affine shift of a proper subfield of $\mathbb{F}_{p^n}$. Let $Q_i(X_i) \in \mathbb{F}_{p^n}[X_i]$ be the image-subspace polynomial of $\mathcal{A}_i$. Let $\mu \in \mathbb{F}_{p^n} \setminus \{0\}$. Let*

$e_1, \ldots, e_t$ satisfy $1 \leq e_i < d_i - 1$ and let $\alpha_i = \sum_{j=0}^{e_i} p^j$. Let

$$R(X_1, \ldots, X_t) \stackrel{\text{def}}{=} \mathrm{Tr}\left(\mu \prod_{i=1}^{t} (Q_i(X_i))^{\alpha_i}\right) \mod \langle (X_i^{p^n} - X_i)_{i \in [t]} \rangle \tag{4.12}$$

If $\sum_{i=1}^{t}(d_i - (e_i + 1)) > n - \max d_i + 1$, then $R(X_1, \ldots, X_t)$ has a monomial $\prod_{i=1}^{t} X_i^{\beta_i}$ with $\mathrm{wt}_p(\beta_i) = e_i + 1$, which has a nonzero coefficient. In particular, $|R(\mathbb{F}_{p^n}^t)| > 1$.

Before giving the proof of Theorem 4.6.1, let us first show how to use it to complete the proof of Theorem 4.2.3. We may assume without loss of generality that $d_i > 2$ by fixing nonzero elements of those spaces that have dimension 2. Next, in Theorem 4.6.1 we set $\mu = 1$ and $e_1 = \ldots = e_t = 1$, which gives $\alpha_1 = \ldots = \alpha_t = 1 + p$. Using Proposition 4.3.1, the polynomial R defined in (4.12) satisfies $R(\mathbb{F}_{p^n}^t) = f(\mathcal{A}_1 \times \cdots \times \mathcal{A}_t)$. Since $\sum(d_i - 2) = \sum(d_i - (e_i + 1)) > n$ we conclude from Theorem 4.6.1 that $|f(\mathcal{A}_1, \ldots, \mathcal{A}_t)| > 1$ and this completes the proof of Theorem 4.2.3. ■

Proof of Theorem 4.6.1: Let $A_i = \{0, \ldots, e_i\}$. By the first part of Lemma 4.4.6, if $B_1, \ldots, B_t \subset \{0, \ldots, n-1\}, |B_i| = e_i + 1$ and $\beta_i = \sum_{k \in B_i} p^i$, then the coefficient of $\mathcal{M} = \prod_{i=1}^{t} X_i^{\beta_i}$ in R, which is denoted henceforth by $c_{\mathcal{M}}$, equals

$$\sum_{r=0}^{n-1} \mu^{p^r} \cdot \prod_{i=1}^{t} \mathrm{Perm}\left(M_i[A_i, B_i - r]\right)^{p^r}, \tag{4.13}$$

where M_i is the matrix associated with Q_i (cf. Definition 4.4.4). We will find suitable powers β_i with $\mathrm{wt}_p(\beta_i) = e_i + 1$ such that $c_{\mathcal{M}} \neq 0$. We define β_i by specifying B_i with $|B_i| = e_i + 1$ and setting $\beta_i = \sum_{k \in B_i} p^k$.

Assume without loss of generality $d_1 = \max d_i$. To define B_i let $\ell_1 = 0$ and for $1 \leq i < t$ let $\ell_{i+1} = \ell_i + d_i - (e_i + 1) \mod n$. In other words, $\ell_i = \sum_{i' < i}(d_i - (e_i + 1)) \pmod n$ where $\ell_1 = 0$. Let $Q_i(X_i) = \sum_{j=0}^{n-1} a_{i,j} X^{p^j} + \hat{a}_i$. Our definition of B_i splits into two cases, depending on whether a_{i,ℓ_i} is nonzero or zero. In the first case we set B_i to be the set $\{\ell_i, n - d_i + 1, n - d_i + 2, \ldots, n - d_i + e_i\}$. In the second case let j_i be the smallest index j' greater than ℓ_i such that $a_{i,j'}$ is nonzero. Similarly, let $\widehat{j}_i$ be the largest index $\widehat{j}'$ smaller than ℓ_i such that $a_{i,\widehat{j}'}$ is nonzero. Let $q_i = j_i - \widehat{j}_i - 1$

be the length of the interval of zero-coefficients of Q_i between indices $\widehat{j}_i$ and j_i. Let $s_i = \min\{q_i, e_i\}$. We set B_i to be the set

$$\{j_i\} \cup \left\{\widehat{j}_i + 1, \ldots, \widehat{j}_i + s_i\right\} \cup \{n - d_i + s_i + 1, \ldots, n - d_i + e_i\}.$$

The last set might be empty in case $s_i = e_i$.

Our proof again employs the strategy of Remark 4.4.7 via the next two claims, proved below. We point out that the noncontainment of $\mathcal{A}_i$ in a proper subfield and the implication this has on the structure of coefficients of Q_i (cf. Lemma 4.4.3) will be crucially used in the proof of Claim 4.6.3 below. Let M_i' denote the zero-one indicator matrix of M_i as given in Definition 4.4.4.

Claim 4.6.2 *For all $i = 1, \ldots, t$ the matrix $M_i'[A_i, B_i]$ is lower triangular with nonzero diagonal entries.*

Claim 4.6.3 *For all $r \in \{1, \ldots, n-1\}$ there exists $i \in \{1, \ldots, t\}$ such that $M_i'[A_i, B_i - r]$ contains an all-zero column.*

Assuming these two claims, Lemma 4.4.6 and Remark 4.4.7 imply Theorem 4.6.1. ■

Proof of Claim 4.6.2: Notice that, by definition, $M_i'[A_i, B_i]$ is a $(e_i + 1) \times (e_i + 1)$ matrix constructed by taking the minor corresponding to the first $e_i + 1$ rows of M_i' and the columns indexed by B_i. To see that $M_i'[A_i, B_i]$ is lower diagonal with nonzero diagonal entries, consider B_i. To simplify notation in this proof let $a_j = a_{i,j}$ be the coefficient of X^{p^j} in $Q(X_i)$ and let a_j' be its zero-one indicator (cf. Definition 4.4.4). There are two cases.

$\mathbf{a}'_{\ell_i} = 1$: We have $B_i = \{\ell_i, n - d_i + 1, \ldots, n - d_i + e_i\}$. Consider the indices j of the coefficients a_j' residing in the various entries of $M_i'[A_i, B_i]$. By assumption $e_i < d_i$ so the entries above the diagonal of $M_i'[A_i, B_i]$ have indices belonging to

$$\{n - d_i + 1, \ldots, n - d_i + e_i\} \subseteq \{n - d_i + 1, \ldots, n - 1\}$$

and this proves $M'_i[A_i, B_i]$ is lower triangular. Regarding the diagonal, at the topmost left entry we have $a'_{\ell_i} = 1$ and in all subsequent positions we have a'_{n-d_i}, which is nonzero by Lemma 4.4.3. This completes the proof of this case.

$\mathbf{a'_{\ell_i} = 0}$**:** In this case we have $B_i = \{j_i\} \cup \left\{\widehat{j}_i + 1, \ldots, \widehat{j}_i + s_i\right\} \cup \{n - d_i + s_i + 1, \ldots, n - d_i + e_i\}$ where

$$j_i = \min\{j > \ell_i \mid a_{i,j} \neq 0\} \text{ and } \widehat{j}_i = \max\{j < \ell_i \mid a_{i,j} \neq 0\}.$$

The uppermost left $(s_i + 1) \times (s_i + 1)$ submatrix of $M'_i[A_i, B_i]$ in this case is

$$\begin{pmatrix} a'_{j_i} & a'_{\widehat{j}_i+1} & \cdots & a'_{\widehat{j}_i+s_i} \\ a'_{j_i-1} & a'_{\widehat{j}_i} & \cdots & a'_{\widehat{j}_i+s_i-1} \\ \vdots & \vdots & \vdots & \vdots \\ a'_{j_i-s_i} & a'_{(\widehat{j}_i+1)-s_i} & \cdots & a'_{\widehat{j}_i} \end{pmatrix}$$

which is lower triangular because the $a'_{\widehat{j}_i+1} = \ldots = a_{\widehat{j}_i+s_i} = 0$, and the diagonal entries of this submatrix are nonzero because $a'_{j_i}, a'_{\widehat{j}_i}$ are nonzero. The last $e_i - s_i$ columns of the matrix — if they exist — are identical to the same last columns of the previous case and this shows that $M'_i[A_i, B_i]$ is lower triangular with nonzero diagonal entries. This completes the proof of Claim 4.6.2. ■

Proof of Claim 4.6.3: In what follows we denote for $c < d$ by $[c, d]$ the set of integers in the interval $[c, d]$ and by $[c, d] \bmod n$ the set $\{i \bmod n \mid i \in [c, d]\}$. We start by observing that

$$M'_i[A_i, \{k\} - r] = \begin{pmatrix} a'_{i,k-r} \\ a'_{i,k-(r+1)} \\ \vdots \\ a'_{i,k-(r+e_i)} \end{pmatrix}.$$

Thus for any $k \in B_i$, if r is such that

$$[k - (r + e_i), k - r] \bmod n \subseteq [n - d_i + 1, n - 1], \tag{4.14}$$

then the matrix $M_i'[A_i, B_i - r]$ contains a zero column. So we get the following proposition.

Proposition 4.6.4 *Whenever $k \in B_i$ and*

$$r \in [k+1, k+d_i-(e_i+1)] \bmod n$$

then $M_i'[A_i, B_i - r]$ contains an all-zero column.

Thus, to prove the claim it suffices to show

$$[1, n-1] \subseteq \bigcup_{i=1}^{t} \cup_{k \in B_i} [k+1, k+(d_i-e_i)-1]. \tag{4.15}$$

(Notice that Claim 4.6.2 implies the containment in the previous equation is in fact an equality.)

Indeed, since $\ell_1 = 0$ we have $B_1 = \{0\} \cup [n-d_1+1, n-d_1+e_1]$, which implies by Proposition 4.6.4 that $M_1'[A_1, B_1 - r]$ contains a zero column for r belonging to

$$[1, d_1-(e_1+1)] \cup [n-d_1+2, n-1] = [\ell_1+1, \ell_2] \cup [n-d_1+2, n-1]. \tag{4.16}$$

Let t' be the minimal i such that $\sum_{i' \le i}(d_{i'}-(e_{i'}+1)) \ge n-d_1+1$, noticing such t' exists by assumption. In this case we have $\sum_{i' \le t'}(d_{i'}-(e_{i'}+1)) < n$ and so $\ell_{t'+1} = \sum_{i' \le t'}(d_{i'}-(e_{i'}+1))$.

We claim that for $1 < i \le t'$ we have

$$\bigcup_{k \in B_i} [k+1, k+d_i-(e_i+1)] \supseteq [\ell_i+1, \ell_{i+1}]. \tag{4.17}$$

which, together with (4.16), proves (4.15) and completes the proof of our claim. There are two cases to consider when proving (4.17).

$\mathbf{a_{i,\ell_i} \neq 0}$: In this case $\ell_i \in B_i$ so the claim follows from Proposition 4.6.4 by recalling that $\ell_{i+1} = \ell_i + d_i - (e_i+1)$.

$\mathbf{a_{i,\ell_i}} = \mathbf{0}$: There are two subcases to consider.

Case 1: $q_i < e_i$. In this case

$$B_i = \{j_i\} \cup [\widehat{j}_i + 1, \widehat{j}_i + q_i] \cup [n - d_i + q_i + 1, n - d_i + e_i].$$

Substituting $\widehat{j}_i + (q_i + 1)$ for j_i and reordering elements of B_i we get

$$B_i = [\widehat{j}_i + 1, j_i = \widehat{j}_i + q_i + 1] \cup [n - d_i + q_i + 1, n - d_i + e_i].$$

We conclude $\ell_i \in B_i$ so by Proposition 4.6.4 our proof is complete, as in the case of $a_{i,\ell_i} \neq 0$ above.

Case 2: $q_i \geq e_i$. In this case we have

$$B_i = \{j_i\} \cup [\widehat{j}_i + 1, \widehat{j}_i + e_i].$$

Substituting $j_i = \widehat{j}_i + q_i + 1$ we get

$$B_i = [\widehat{j}_i + 1, \widehat{j}_i + e_i] \cup \left\{\widehat{j}_i + q_i + 1\right\}$$

Now we use the fact that $\mathcal{A}_i$ is not contained in an affine shift of a proper subfield. We notice that since $i \leq t'$ we have by maximality of d_1 that

$$\widehat{j}_i < \ell_i \leq n - d_1 \leq n - d_i$$

which implies (using the maximality of d_1 again) that $\widehat{j}_i + 1 \neq n - d_i + 1$. As $\mathcal{A}_i$ is not contained in an affine shift of a proper subfield and $\widehat{j}_i + 1 \neq n - d_i + 1$, our Structural Lemma 4.4.3 implies that $j_i - \widehat{j}_i \leq d_i - 1$, or, equivalently, $j_i \leq \widehat{j}_i + d_i - 1$.

Taking all but the last element of B_i in the previous equation notice

$$\bigcup_{k\in[\widehat{j}_i+1,\widehat{j}_i+e_i]} [k+1, k+d_i-(e_i+1)] \supseteq [\widehat{j}_i+2, \widehat{j}_i+d_i-1],$$

which contains j_i. Now, since $\widehat{j}_i < \ell_i < j_i$ when we reinsert j_i into B_i we conclude

$$\begin{aligned} \cup_{k\in B_i}[k+1, k+d_i-(e_i+1)] &\supseteq [\widehat{j}_1+2, j_i+d_i-(e_i+1)] \\ &\supseteq [\ell_i+1, \ell_{i+1}]. \end{aligned}$$

This completes the last case and with it the proof of Claim 4.6.3 is complete. ■

4.7 Disperser for affine spaces of sublinear dimension

In this section we prove Theorem 4.2.2. We start by examining what happens to $\mathcal{A} \subset \mathbb{F}_p^{nr}$ when it is partitioned into r blocks of size n. Then we prove the main theorem, by essentially reducing it to the case of independent affine sources described in Theorem 4.6.1.

4.7.1 Preparatory lemmata

Our first lemma, already used by Bourgain [Bou07] in his construction of affine extractors, gives a certain kind of direct sum decomposition of $\mathbb{F}_p$-affine subspaces of $\mathbb{F}_{p^n}^r$.

Lemma 4.7.1 (Bourgain's decomposition) *Let $\mathcal{A} \subseteq (\mathbb{F}_p^n)^r$ be an $\mathbb{F}_p$-affine subspace. Let $\gamma \in \mathcal{A}$. Then there exist linear spaces $Y_1, \ldots, Y_r \subseteq \mathbb{F}_p^n$ and linear maps $\sigma_{ij} : Y_j \to \mathbb{F}_p^n$ such that:*

$$\mathcal{A} = \{(x_1, \ldots, x_r) \mid \exists y_i \in Y_i \text{ such that } x_i = \gamma_i + y_i + \sum_{j<i} \sigma_{ij}(y_j)\}$$

and $\dim \mathcal{A} = \sum_{i\in[r]} \dim Y_i$.

This lemma amounts to taking the echelon-form of a matrix whose rows form a basis for the linear subspace underlying the affine subspace $\mathcal{A}$.

The next lemma should be thought of as a complement to Theorem 4.6.1. It expands the class of sources on which the function R given in that theorem is non-constant. This expanded class is what we will use in the proof of our main theorem.

Lemma 4.7.2 *For each* $i \in [r]$, *let* $P_i(X_i) \in \mathbb{F}_{p^n}[X_i]$ *be a linearized polynomial. For each* $j < i$, *let* $P_{ij}(X_j) \in \mathbb{F}_{p^n}[X_j]$ *be a linearized polynomial. Let* $\gamma \in \mathbb{F}_{p^n}^r$. *Let* $I_0 \subseteq [r]$ *with* $I_0 = \{i_1 < i_2 < \ldots < i_t\}$. *Let* $e_{i_1}, \ldots, e_{i_t} > 1$ *be integers and let* $\alpha_i = \sum_{k=0}^{e_i} p^i$, *for* $i \in I_0$. *Let* $\mu \in \mathbb{F}_{p^n} \setminus \{0\}$. *Let* $g(X_1, \ldots, X_r) \in \mathbb{F}_{p^n}[X_1, \ldots, X_r]$ *be the polynomial*

$$\mathrm{Tr}\left(\mu \prod_{i\in I_0} \left(P_i(X_i) + \sum_{j<i} P_{ij}(X_j) + \gamma_i\right)^{\alpha_i}\right) \bmod \langle X_i^q - X_i\rangle_{i\in[r]}.$$

Let $g'(X_1, \ldots, X_r) \in \mathbb{F}_{p^n}[X_1, \ldots, X_r]$ *be the polynomial*

$$\mathrm{Tr}\left(\mu \prod_{i\in I_0} P_i(X_i)^{\alpha_i}\right) \bmod \langle X_i^q - X_i\rangle_{i\in[r]}.$$

Then for any $(\beta_{i_1}, \ldots, \beta_{i_t})$ *where* $\mathrm{wt}_p(\beta_{i_k}) = e_{i_k} + 1$, *the coefficients of the monomial* $\prod_{i\in I_0} X_i^{\beta_i}$ *in* g *and in* g' *are equal.*

Proof We want to show that the coefficient of $\prod_{i\in I_0} X_i^{\beta_i}$ in $g(X_1, \ldots, X_r)$ is the same as in $g'(X_1, \ldots, X_r)$. We do this by expanding out the expressions for $g(X_1, \ldots, X_r)$ and $g'(X_1, \ldots, X_r)$ and keeping track of the monomials.

Let $X_{<i}$ denote the tuple of variables $(X_1, \ldots, X_{i-1})$. Let $\hat{P}_i(X_{<i})$ be the polynomial $\sum_{j<i} P_{ij}(X_j) + \gamma_i$.

Expanding $g(X_1, \ldots, X_r)$ we get

$$g(X_1, \ldots, X_r) = \sum_{r=0}^{n-1} \left(\mu \prod_{i\in I_0} \left(P_i(X_i) + \hat{P}_i(X_{<i})\right)^{\alpha_i}\right)^{p^r} \bmod \langle (X_i^{p^n} - X_i)_{i\in[r]}\rangle \quad (4.18)$$

$$= \sum_{r=0}^{n-1} \mu^{p^r} \left(\prod_{i\in I_0} \left(P_i(X_i) + \hat{P}_i(X_{<i})\right)^{\sum_{l=0}^{e_i} p^{r+l}}\right) \bmod \langle (X_i^{p^n} - X_i)_{i\in[r]}\rangle \quad (4.19)$$

We now expand the term $\left(P_i(X_i) + \hat{P}_i(X_{<i})\right)^{\sum_{l=0}^{e_i} p^{r+l}}$ to obtain

$$\prod_{l=0}^{e_i} \left(P_i(X_i) + \hat{P}_i(X_{<i})\right)^{p^{r+l}} \tag{4.20}$$

$$= \prod_{l=0}^{e_i} \left(P_i(X_i)^{p^{r+l}} + \hat{P}_i(X_{<i})^{p^{r+l}}\right) \tag{4.21}$$

$$= \prod_{l=0}^{e_i} (P_i(X_i))^{p^{r+l}} + \sum_{\substack{L \subseteq \{0,1,\ldots,e_i\} \\ L \neq \emptyset}} \left(\prod_{l \notin L} P_i(X_i)^{p^{r+l}}\right) \left(\prod_{l' \in L} \hat{P}_i(X_{<i})^{p^{r+l'}}\right) \tag{4.22}$$

Now the first term has $\mathbb{F}_p$-degree in X_i equal to $e_i + 1$, while all the other terms have $\mathbb{F}_p$-degree in X_i strictly less than e_i+1. The reason for this is that the polynomial $\hat{P}_i(X_{<i})$ does not mention the variable X_i, and each $P_i(X_i))^{p^{r+l}}$ and $\hat{P}_i(X_{<i})^{p^{r+l'}}$ are linearized polynomials (and hence of $\mathbb{F}_p$-degree 1). Let us summarize this by writing (4.22) as

$$(P_i(X_i))^{\sum_{l=0}^{e_i} p^{r+l}} + G_i(X_{\leq i})$$

and noting that the $\mathbb{F}_p$-degree of G_i in X_i is at most e_i.

Now let us go back to (4.19) and consider the rth summand within the parenthesis.

$$\begin{aligned}
\prod_{i \in I_0} \left(P_i(X_i) + \hat{P}_i(X_{<i})\right)^{\sum_{l=0}^{e_i} p^{r+l}} &= \prod_{i \in I_0} \left((P_i(X_i))^{\sum_{l=0}^{e_i} p^{r+l}} + G_i(X_{\leq i})\right) \\
&= \left((P_{i_t}(X_{i_t}))^{\sum_{l=0}^{e_{i_t}} p^{r+l}} + G_{i_t}(X_{\leq i_t})\right) \\
&\qquad \cdot \prod_{i \in I_0, i < i_t} \left((P_i(X_i))^{\sum_{l=0}^{e_i} p^{r+l}} + G_i(X_{\leq i})\right) \\
&= \left((P_{i_t}(X_{i_t}))^{\sum_{l=0}^{e_{i_t}} p^{r+l}} + G_{i_t}(X_{\leq i_t})\right) \cdot H_t(X_{\leq i_{t-1}})
\end{aligned}$$

The rightmost term above, denoted H_t, does not mention X_{i_t}. Furthermore, as stated above G_t has $\mathbb{F}_p$-degree at most e_{i_t} in X_{i_t}. But m_ξ has $\mathbb{F}_p$-degree $e_{i_t} + 1$ in X_{i_t}, so to contribute to the coefficient of m_ξ we must select terms *only* from $(P_{i_t}(X_t))^{\sum_{l=0}^{e_{i_t}} p^{r+l}}$ and multiply them by the appropriate terms in H_t. Next, consider the terms inside

H_t,

$$H_t(X_{\leq i_{t-1}}) = \left(\left(P_{i_{t-1}}(X_{i_{t-1}})\right)^{\sum_{l=0}^{e_{i_{t-1}}} p^{r+l}} + G_{i_{t-1}}(X_{\leq i_{t-1}}) \right) \cdot H_{t-1}(X_{\leq i_{t-2}})$$

where,

$$H_{t-1}(X_{\leq i_{t-2}}) = \prod_{i \in I_0, i < i_{t-1}} \left((P_i(X_i))^{\sum_{l=0}^{e_i} p^{r+l}} + G_i(X_{\leq i}) \right)$$

As before, we notice that $H_{t-1}(X_{\leq i_{t-2}})$ does not mention $X_{i_{t-1}}$ and H_{t-1} has $\mathbb{F}_p$-degree $e_{i_{t-1}}$ in $X_{i_{t-1}}$. But m_ξ has $\mathbb{F}_p$-degree $e_{i_{t-1}} + 1$ in $X_{i_{t-1}}$, implying that we must select terms *only* from $\left(P_{i_{t-1}}(X_{i_{t-1}})\right)^{\sum_{l=0}^{e_{i_{t-1}}} p^{r+l}}$. Continuing in this manner for $i = i_{t-2}, \ldots, i_1$ we conclude that the only contributions to the coefficient of m_ξ come from $\left(\prod_{i \in I_0} (P_i(X_i))^{\sum_{l=0}^{e_i} p^l}\right)^{p^r}$. Summing up over all r, the lemma follows. ■

4.7.2 Proof of Theorem 4.2.2

We can now analyze our main affine disperser construction. Theorem 4.2.2 will follow by setting the proper parameters into the following theorem.

Theorem 4.7.3 (Affine disperser — non-parameterized version) *Let $t < r$ be integers. Let n be prime with $n \geq r(r+1)/t$. For each $i \in [r]$, let $e_i = r + 1 - i$ and let $\alpha_i = \sum_{k=0}^{e_i} p^k$. Let $\mu \in \mathbb{F}_{p^n} \setminus \{0\}$. Let $f : \mathbb{F}_{p^n}^r \to \mathbb{F}_p$ be given by*

$$f(x_1, \ldots, x_r) = \mathrm{Tr}\left(\mu \sum_{I \subseteq [r], |I| = t} \prod_{i \in I} x_i^{\alpha_i} \right).$$

Let $\mathcal{A} \subseteq \mathbb{F}_{p^n}^r$ be any $\mathbb{F}_p$-affine space with $\dim(\mathcal{A}) > \frac{nr}{t} + nt + r(r+1)$. Then $|f(\mathcal{A})| > 1$.

Before proving this theorem let us show how it implies Theorem 4.2.2.

Proof of Theorem 4.2.2: For our selection of parameters n, t, r we notice the assumptions of Theorem 4.7.3 hold. Indeed, by Bertrand's postulate we can bound n

from above by $4m^{3/5}$, hence $r \geq m^{2/5}/4$. Notice that for our setting of parameters

$$r(r+1)/t \leq \sqrt{r}(r+1) \leq m^{3/5} < n$$

and if $d > 6m^{4/5}$ then we have

$$\frac{nr}{t} + nt + r(r+1) \leq \frac{4}{\sqrt{2}}m^{4/5} + \frac{4}{\sqrt{2}}m^{4/5} + \frac{1}{4}m^{4/5} + o(m^{4/5}) < d.$$

Thus, the function f in Theorem 4.7.3 has the property that for any $\mathcal{A}$ with $\dim(\mathcal{A}) > 6m^{4/5}$, we have $|f(\mathcal{A})| > 1$. Finally notice that Proposition 4.3.3 implies that f as defined in Theorem 4.7.3 is identical to f defined in Theorem 4.2.2, up to renaming of the variables x_i. This completes the proof. ■

Proof of Theorem 4.7.3: Our proof strategy is again as outlined in the introduction. Our first goal is to find a polynomial mapping $H : \mathbb{F}_{p^n}^r \to \mathbb{F}_{p^n}^r$ such that $H(\mathbb{F}_{p^n}^r) = \mathcal{A}$. We will then show that the composed function $f \circ H$ is a non-constant map, by showing that in its representation as a polynomial, there is a positive degree monomial with a nonzero coefficient.

To define the mapping H, we first decompose the affine space $\mathcal{A}$ using Lemma 4.7.1. Let $\gamma \in \mathcal{A}$. Then by that lemma, we may find a collection of $\mathbb{F}_p$-linear subspaces $Y_1, \ldots, Y_r \subseteq \mathbb{F}_{p^n}$ and linear maps $\sigma_{ij} : \mathbb{F}_{p^n} \to \mathbb{F}_{p^n}$ for $i, j \in [r]$ with $i < j$ such that:

$$\mathcal{A} = \{(x_1, \ldots, x_r) \mid \exists y_i \in Y_i \text{ such that } x_i = \gamma_i + y_i + \sum_{j<i} \sigma_{ij}(y_j)\}$$

and $\dim \mathcal{A} = \sum_{i \in [r]} \dim Y_i$.

Let $Q_i(X) \in \mathbb{F}_{p^n}[X]$ be the image-subspace polynomial of Y_i. Let $Q_{ij}(X)$ be the linearized polynomial (guaranteed to exist by Lemma 2.4.4) such that $Q_{ij}(x) = \sigma_{ij}(Q_i(x))$ for each $x \in \mathbb{F}_p^n$. Let $R_i(X_1, \ldots, X_r) \in \mathbb{F}_{p^n}[X_1, \ldots, X_r]$ be the polynomial $Q_i(X_i) + \sum_{j<i} Q_{ij}(X_j) + \gamma_i$. Then by the above comments, the image of the function H mapping $x = (x_1, \ldots, x_r) \in \mathbb{F}_{p^n}^r$ to $(R_1(x), \ldots, R_r(x))$ is precisely $\mathcal{A}$.

Now let $h(X_1, \ldots, X_r) \in \mathbb{F}_{p^n}[X_1, \ldots, X_r]$ be the polynomial representing $f \circ H$, namely

$$h(X_1, \ldots, X_r) = f(R_1(X_1, \ldots, X_r), \ldots, R_r(X_1, \ldots, X_r)) \mod \langle X_i^q - X_i \rangle_{i \in [r]} \tag{4.23}$$

$$= \sum_{I \subseteq [r], |I| = t} \mathrm{Tr}\left(\mu \prod_{i \in I} R_i(X_1, \ldots, X_r)^{\alpha_i}\right) \mod \langle X_i^q - X_i \rangle_{i \in [r]} \tag{4.24}$$

By Proposition 4.3.1, we have $h(\mathbb{F}_{p^n}^r) = f(\mathcal{A})$.

Therefore, to show that $|f(\mathcal{A})| > 1$, it suffices to show that $|h(\mathbb{F}_{p^n}^r)| > 1$. We do this by showing that $h(X_1, \ldots, X_r)$ has a monomial of positive degree with a nonzero coefficient and invoking Proposition 4.3.1.

To find this monomial, we consider the representation (4.24) of the polynomial $h(X_1, \ldots, X_n)$. We will first find a set $I_0 \subseteq [r]$, with $|I_0| = t$, of "blocks with high entropy". Then via Theorem 4.6.1, we will argue that the summand in (4.24) corresponding to I_0 is a nonzero polynomial, with certain monomial $\mathcal{M}$ having a nonzero coefficient. We will then show that no other summand in the sum (4.24) can have the monomial $\mathcal{M}$ with a nonzero coefficient, thus establishing that $\mathcal{M}$ appears in h with a nonzero coefficient, as desired.

We proceed with implementing this plan. Let $d_i = \dim(Y_i)$ and let $d = \dim(\mathcal{A}) > \frac{nr}{t} + nt + r(r+1)$. We have $\sum_i d_i = d$. Let $S = \{i \in [r] \mid d_i > r + 1\}$. Then we get

- $|S| \geq t$ (since each $d_i \leq n$ and $\sum d_i > nt + r(r+1)$).
- $\sum_{i \in S} d_i \geq \sum_{i \in [r]} (d_i - r - 1) = d - r(r+1) \geq nr/t + nt$.

Thus there exists $I_0 \subseteq S$ (and hence each $i \in I_0$ has $d_i > r + 1$) with $|I_0| = t$ such that

$$\sum_{i \in I_0} (d_i - (r+1)) \geq \left(\sum_{i \in S} d_i\right) \frac{t}{r} - (r+1)t \geq n + nt^2/r - (r+1)t \geq n, \tag{4.25}$$

where the last inequality used the hypothesis that $n \geq r(r+1)/t$.

Let us focus on the term

$$g(X_1, \ldots, X_r) = \mathrm{Tr}\left(\mu \prod_{i \in I_0} R_i(X_1, \ldots, X_r)^{\alpha_i}\right) \bmod \langle X_i^q - X_i \rangle_{i \in [r]}$$

in the representation (4.24) of the polynomial $h(X_1, \ldots, X_r)$.

Putting $g'(X_1, \ldots, X_r) = \mathrm{Tr}\left(\mu \prod_{i \in I_0} Q_i(X_i)^{\alpha_i}\right) \bmod \langle X_i^q - X_i \rangle_{i \in [r]}$ and noting that each $e_i + 1 \leq r + 1$, Equation (4.25) and Theorem 4.6.1 imply that there is a monomial $\mathcal{M} = \prod_{i \in I_0} X_i^{\beta_i}$ with $\mathrm{wt}_p(\beta_i) = \mathrm{wt}_p(\alpha_i) = e_i + 1$, which has a nonzero coefficient in g'. Lemma 4.7.2 now implies that the coefficient of $\mathcal{M}$ in g is exactly the same as the coefficient of $\mathcal{M}$ in g', and hence nonzero.

We now show that in the representation (4.24) of the polynomial $h(X_1, \ldots, X_r)$, no summand other than g can have a nonzero coefficient for the monomial $\mathcal{M}$. First notice that each $R_i(X_1, \ldots, X_r)$ is a polynomial only in the variables $X_1, X_2, \ldots, X_i$, and is a sum of monomials of the form $aX_k^{p^b}$ plus possibly a constant term (i.e., monomials of total $\mathbb{F}_p$-degree at most 1).

Let $J \subseteq [r]$ with $|J| = t$, and consider the expression $\mathrm{Tr}(\mu \prod_{j \in J} R_j(X_1, \ldots, X_r)^{\alpha_j})$ mod $\langle (X_i^{p^n} - X_i)_{i \in [r]} \rangle$. By definition, it equals:

$$\mathrm{Tr}\left(\mu \prod_{j \in J} \prod_{l=0}^{e_j} R_j(X_1, \ldots, X_j)^{p^l}\right) \quad \bmod \langle (X_i^{p^n} - X_i)_{i \in [r]} \rangle.$$

Suppose the monomial $\mathcal{M}$ appeared in the above polynomial with a nonzero coefficient. Then, expanding the trace map, there is some $w \in [n-1]$ such that $\mathcal{M}$ appears in

$$\prod_{j \in J} \prod_{l=0}^{e_j} R_j(X_1, \ldots, X_j)^{p^{l+w}} \quad \bmod \langle (X_i^{p^n} - X_i)_{i \in [r]} \rangle$$

with a nonzero coefficient. Letting $R_{jl} = R_j^{p^{l+w}}$, we may rewrite the last polynomial as

$$\prod_{j \in J} \prod_{l=0}^{r-j+1} R_{jl}(X_1, \ldots, X_j) \quad \bmod \langle (X_i^{p^n} - X_i)_{i \in [r]} \rangle,$$

where each R_{jl} is a sum of monomials of total $\mathbb{F}_p$-degree at most 1. Each monomial $\mathcal{M}'$ that appears in this product is obtained by choosing, for each $j \in J$ and $l \in [0, r-j+1]$, a monomial from $R_{jl}(X_1, \ldots, X_j)$, and multiplying all these monomials out. Since we know that $\mathcal{M}$ appears in this product, let us focus on the choices made in order for $\mathcal{M}$ to appear. We set $\lambda_j(l) = i$ if for (j,l) we chose a monomial from $R_{jl}(X_1, \ldots, X_j)$ whose variable is indexed by i (i.e., we chose some aX_i^b). Observe that the $\mathbb{F}_p$-degree in X_i of $\mathcal{M}$ is at most the number of (j,l) pairs for which $\lambda_j(l) = i$ (which may be compactly written as $\sum_{j \in J} |\lambda_j^{-1}(i)|$). However, we know that for any $i \in I_0$, the $\mathbb{F}_p$-degree of $\mathcal{M}$ in the variable X_i is $e_i + 1$ (which equals $r + 2 - i$). The following combinatorial claim (whose proof appears next) now shows that J must be equal to I_0.

Claim 4.7.4 *Let $I_0 \subseteq [r]$ with $|I_0| = t$. Suppose $J \subseteq [r]$ with $|J| = t$, and that there exist functions $\lambda_j : \{0, 1, \ldots, r+1-j\} \to \{1, \ldots, j\}$ for $j \in J$, with the property that for each $i \in I_0$,*

$$\sum_{j \in J} |\lambda_j^{-1}(i)| \geq r + 2 - i. \tag{4.26}$$

Then $J = I_0$.

Therefore, we have shown that there is precisely one summand, namely the one corresponding to I_0, in the representation (4.24) of $h(X_1, \ldots, X_r)$ that has a nonzero coefficient for the monomial $\mathcal{M}$. Thus $\mathcal{M}$ appears in h with a nonzero coefficient, and thus $|h(\mathbb{F}_{p^n}^r)| > 1$, as desired. ■

Proof of Claim 4.7.4: Note that for any $j \in J$ we have

$$\sum_{i \in I_0} |\lambda_j^{-1}(i)| \leq |\{0, 1, \ldots, r+1-j\}| = r + 2 - j.$$

Thus

$$\sum_{i \in I_0} (r + 2 - i) \leq \sum_{j \in J} \sum_{i \in I_0} |\lambda_j^{-1}(i)| \leq \sum_{j \in J} (r + 2 - j).$$

As $|I_0| = |J|$, we have $\sum_{i \in I_0} i \geq \sum_{j \in J} j$.

Consider now the expression $\sum_{j\in J}\sum_{l=0}^{r+1-j}(j-\lambda_j(l))$, which is ≥ 0, because $\lambda_j(l) \leq j$. Thus,

$$\sum_{j\in J}(r+2-j)\cdot j \geq \sum_{j\in J}\sum_{l=0}^{r+1-j}\lambda_j(l) \geq \sum_{i\in I_0}(r+2-i)\cdot i.$$

The last inequality follows from the assumption (4.26). Rearranging, we get,

$$\sum_{i\in I_0} i^2 - \sum_{j\in J} j^2 \geq (r+2)\cdot(\sum_{i\in I_0} i - \sum_{j\in J} j) \geq 0.$$

Thus $\sum_{i\in I_0} i^2 \geq \sum_{j\in J} j^2$.

For general k, considering the expression $\sum_{j\in J}\sum_{l=0}^{r+1-j}(j^k-\lambda_j(l)^k)$, which is non-negative, we get

$$\sum_{i\in I_0} i^{k+1} - \sum_{j\in J} j^{k+1} \geq (r+2)\cdot(\sum_{i\in I_0} i^k - \sum_{j\in J} j^k),$$

which by induction on k is ≥ 0. Thus for all k,

$$\sum_{i\in I_0} i^k \geq \sum_{j\in J} j^k \tag{4.27}$$

This implies that $i_1 := \max(I_0) \geq \max(J) =: j_1$. However, $i_1 \leq j_1$, otherwise $\lambda_j^{-1}(i_1) = \emptyset$ for each j. Thus $i_1 = j_1$. This forces $\lambda_{j_1}(l) = i_1$ for each l.

Taking this information back to Equation (4.27), we now see that the second-largest element i_2 of $I_0 \geq$ the second-largest element j_2 of J. But we must have $i_2 \leq j_2$, otherwise $\lambda_j^{-1}(i_2) = \emptyset$ for all j (recall that $\lambda_{j_1}(l) = i_1$ for each l, and there is no other j for which $i_1 \leq j$). Thus $i_2 = j_2$.

Inducting now on s, and arguing about the s-th largest element of I_0 and J, we get that $I_0 = J$.

■

4.8 Open Problems

We conclude with some open problems.

1. Construct explicit affine dispersers from dimension n^δ for arbitrary $\delta > 0$.

2. Are our affine dispersers also affine extractors? We conjecture that they are.

3. Let n be even, and consider the cubic residue symbol $\chi : \mathbb{F}_{2^n}^* \to \{1, \omega, \omega^2\}$ (where ω is a cube-root of unity). Is χ an affine disperser/extractor from dimension δn for every $\delta > 0$?

Chapter 5

The Extended Method of Multiplicities

5.1 Introduction

In this chapter, our main goal is to present an improvement to 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 the subset K satisfying the algebraic conditions, one first constructs a non-zero 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 for its degree to derive a contradiction to the non-zeroness of the polynomial; this gives bounds on the combinatorial parameters of interest. In the form of a three word slogan: Interpolation, Extrapolation, Contradiction.

The polynomial method has seen utility in the computer science literature in works on "list-decoding" starting with Sudan [Sud97] and subsequent works. Recently the method has been applied to analyze "extractors" by Guruswami, Umans, and Vadhan [GUV07]. Most relevant to this work are its applications to lower bound the cardinality of "Kakeya sets" by Dvir [Dvi08], and the subsequent constructions of

"mergers" and "extractors" by Dvir and Wigderson [DW08]. (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 [GS99] and continuing through many subsequent works, most significantly in the works of Parvaresh and Vardy [PV05] and Guruswami and Rudra [GR06] 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 [SS08].

Our main contribution 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 nonzero 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 and arrive at a contradiction by showing that our polynomial has more high multiplicity zeroes than its degree allows it to have; this then gives better bounds on the combinatorial parameters of interest. In the form of a three word slogan: Interpolation-with-high-multiplicity, Extrapolation-with-high-multiplicity, Contradiction-with-high-multiplicity.

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).

5.1.1 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 [Wol99], 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 [SS08] 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). Till recently even the exponent of q was not known precisely (see [Dvi08] for details of work prior to 2008). This changed with the beautiful result of [Dvi08] (combined with an observation of Alon and Tao) who showed that for every n, $|K| \geq c_n q^n$, for some constant c_n depending only on n.

Subsequently the work [SS08] explored the growth of the constant c_n as a function of n. The result of [Dvi08] shows that $c_n \geq 1/n!$, and [SS08] improve 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 5.1.1 *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 $q \to \infty$.

5.1.2 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, possibly correlated, random variables $\mathsf{A}_1, \ldots, \mathsf{A}_\Lambda$, along with a short uniformly random seed B, which is independent of $\mathsf{A}_1, \ldots, \mathsf{A}_\Lambda$, and "merges" the randomness of $\mathsf{A}_1, \ldots, \mathsf{A}_\Lambda$. Specifically the output of the merger should be statistically close to a high-entropy-rate source of randomness provided at least one of the input variables $\mathsf{A}_1, \ldots, \mathsf{A}_\Lambda$ is uniform.

Mergers were first introduced by Ta-Shma [TS96a] in the context of explicit constructions of extractors. A general framework was given in [TS96a] that reduces the problem of constructing good extractors into that of constructing good mergers. Subsequently, in [LRVW03], mergers were used in a more complicated manner to create extractors which were optimal to within constant factors. The mergers of [LRVW03] 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 [LRVW03] merger analysis was improved in [DS07] using the connection to the finite field Kakeya problem and the (then) state of the art results on Kakeya sets.

The new technique in [Dvi08] inspired Dvir and Wigderson [DW08] to give a very simple, algebraic, construction of a merger which can be viewed as a derandomized version of the [LRVW03] 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 $\mathsf{A}_1, \ldots, \mathsf{A}_\Lambda$, they associate a curve $C : \mathbb{F}_q \to \mathbb{F}_q^n$ of degree $\leq \Lambda$ which 'passes' through all the points $\mathsf{A}_1, \ldots, \mathsf{A}_\Lambda$ (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 $q \geq \mathrm{poly}(\Lambda \cdot n)$ 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 $q \geq \mathrm{poly}(n)$ appears only in the analysis. In our

work we remove this restriction to show:

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 5.4.3) allows us to merge Λ sources using seed length which is only logarithmic in the number of sources and does not depend entirely on the length of each source. Earlier constructions of mergers required the seed to depend either linearly on the number of blocks [LRVW03, Zuc07] or to depend also on the *length* of each block [DW08]. [1]

One consequence of our improved merger construction is an improved construction 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 [Sha02] and the more recent [GUV07, DW08] for references) and many applications have been found for explicit extractor is various research areas spanning theoretical computer science. Yet all previous constructions lost a linear fraction of the min-entropy of the source (i.e., acheived $m = (1 - \epsilon)k$ for some constant $\epsilon > 0$) or used super-logarithmic 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 minentropy of the source, while still using $O(\log n)$-length seeds. We now state our extractor result in an informal way (see Theorem 5.5.3 for the formal statement).

Informal Theorem [Extractor]: *There exists an explicit (k, ϵ)-extractor for*

[1] The result we refer to in [Zuc07, Theorem 5.1] is actually a condenser (which is stronger than a merger).

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 polylog(n) *can be arbitrarily high constants.*

5.1.3 List-Decoding of Reed-Solomon Codes

The Reed-Solomon list-decoding problem 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. (Strictly speaking the problem requires α_i's to be distinct, but we will consider the more general problem here.) 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 set of distinct input points)?

A somewhat nonstandard, yet reasonable, interpretation of the list-decoding algorithms of [Sud97, GS99] 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 *happen* also 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 [CB04] 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 [GS99] 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 5.6.2 we recover the bound of [CB04] using our extended method of multiplicities.

5.1.4 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 non-zero 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 the 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 2.1.8). Using this lemma, we are able to derive much better benefits from the "polynomial method". Indeed this allows us to fully utilize the power of the polynomial ring $\mathbb{F}_q[\mathbf{X}]$ and are not limited by the power of the function space mapping $\mathbb{F}_q^n$ to $\mathbb{F}_q$.

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 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 such sets must also be large and this leads to the analysis of the

Dvir-Wigderson merger.

Organization of this chapter: In Section 5.2 we present our lower bounds for Kakeya sets. In Section 5.3 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 5.4. We then show how to use our mergers to construct the novel extractors in Section 5.5. Finally, in Section 5.6, we include the algebraic proof of the list-size bounds for the list-decoding of Reed-Solomon codes. We conclude with some open questions.

5.2 A lower bound on the size of Kakeya sets

We now give a lower bound on the size of Kakeya sets in $\mathbb{F}_q^n$. Preliminaries on derivatives and multiplicities appear in Chapter 2.

We implement the plan described in Section 5.1. Specifically, in Proposition 5.2.1 we show that we can find a somewhat low degree non-zero 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 5.2.3 we show that the homogenous 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 homogenous polynomial is identically zero if the Kakeya set is too small, leading to the desired contradiction. The resulting lower bound (slightly stronger than Theorem 5.1.1) is given in Theorem 5.2.2.

Proposition 5.2.1 *Given a set $K \subseteq \mathbb{F}^n$ and non-negative integers m, d such that*

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

there exists a non-zero 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 5.2.2 *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 zeros of P on K and ℓ will be the multiplicity of the zeros of the homogenous 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 $(m - \ell)q > d - \ell$. 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 5.2.1 there exists a non-zero polynomial $P(\mathbf{X}) \in \mathbb{F}[\mathbf{X}]$ of total degree exactly d^*, where $d^* \leq d$, 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 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 5.2.3 *For each* $\mathbf{b} \in \mathbb{F}_q^n$.

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

Proof Fix $\mathbf{i}$ with $\text{wt}(\mathbf{i}) = w \leq \ell - 1$. 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 2.1.4, $\text{mult}(Q, \mathbf{a} + t\mathbf{b}) \geq m - w$. Since $w \leq \ell - 1$ and $(m - \ell) \cdot q > d^* - \ell$, we get that $(m - w) \cdot q > d^* - w$.

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 2.1.6, 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 homogenous 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 2.1.3), 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 $\text{mult}(H_P, \mathbf{b}) \geq \ell$. ■

Applying Corollary 2.1.9, 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

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

■

5.3 Statistical Kakeya for curves

Next we extend the results of the previous section to a form conducive to analyze the mergers of Dvir and Wigderson [DW08]. The extension changes two aspects of the consideration in Kakeya sets, that 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 in 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 5.3.1 (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 such that $|S| = \lambda q^n$. 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 $\lambda \geq \eta$ *we get that* $|K| \geq \left(\frac{\eta q}{\Lambda+1}\right)^n$.

Proof Let ℓ be a large integer and let

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

$$m = \Lambda \frac{\lambda \ell q - 1 - (\ell - 1)}{\eta q} + \ell.$$

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 5.2.2, we will prove that

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

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

If possible, let $|K| < \frac{\binom{d+n}{n}}{\binom{m+n-1}{n}}$. As before, by Proposition 5.2.1 there exists a non-zero 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 2.1.9.

Claim 5.3.2 *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 2.1.4, 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 2.1.6, 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 $\eta q(m - w)$ zeros (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 zeros, $Q_{\mathbf{x}_0}(T)$ 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 $\text{wt}(\mathbf{i}) \leq \ell - 1$, we conclude that $\text{mult}(P, \mathbf{x}_0) \geq \ell$. ■

Thus P vanishes at every point in S with multiplicity ℓ. As $P(\mathbf{X})$ is a non-zero polynomial, Corollary 2.1.9 implies that $\ell|S| \leq d^* q^{n-1}$. Hence $\ell \lambda q^n \leq d q^{n-1}$, 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

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

■

5.4 Improved Mergers

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

5.4.1 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

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

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 the Λ-tuple is "somewhere-random" as defined below.

Definition 5.4.1 (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. A* (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 5.4.2 (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 $d = \Omega(\log N + \log \Lambda)$ or $d = \Omega(\Lambda)$. We only require $d = \Omega(\log \Lambda)$.

Theorem 5.4.3 *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*

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

5.4.2 The Curve Merger of [DW08] and its analysis

The merger that we consider is a very simple one proposed by Dvir and Wigderson [DW08], and we improve their analysis using our extended method of multiplicities. We note that they used the polynomial method in their analysis; and the basic method of multiplicities doesn't seem to improve their analysis.

The curve merger of [DW08], denoted f_{DW}, is obtained as follows. Let $q \geq \Lambda$ be a prime power, and let n be any integer. Let $\gamma_1, \ldots, \gamma_\Lambda \in \mathbb{F}_q$ be distinct, and 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 for any $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_\Lambda) \in (\mathbb{F}_q^n)^\Lambda$ and $u \in \mathbb{F}_q$, the curve merger f_{DW} maps $(\mathbb{F}_q^n)^\Lambda \times \mathbb{F}_q$ to $\mathbb{F}_q^n$ as follows:

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

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 5.4.4 *Let $q \geq \Lambda$ and A be 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 5.4.3 easily follows from the above. We note that [DW08] proved a similar theorem assuming $q \geq \mathrm{poly}(n, \Lambda)$, forcing their seed length to grow logarithmically with n as well.

Proof of Theorem 5.4.3: 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 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 5.4.4 shows that f has the required merger property. ■

We now prove Theorem 5.4.4.

Proof of Theorem 5.4.4: Without loss of generality, we may assume that A is 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

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

Suppose 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 $\mathbf{x} \in S$,

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

Fixing the values of A_{-i_0}, 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 $C_{\mathbf{x}}(T) = f(\mathbf{y}(\mathbf{x}), T) = \sum_{j=1}^{\Lambda} \mathbf{y}_j c_j(T)$. 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 5.3.1. 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 analog of Theorem 5.4.4 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 5.4.4 (in a black box manner). The reason why this works is that the merger commutes with the linear map (for details see [DW08]).

5.5 Extractors with sub-linear 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 sub linear entropy loss.

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

Definition 5.5.1 (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* $E(\mathsf{X}, \mathsf{U}_d)$ *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 sub-linear entropy loss.

Theorem 5.5.2 (Basic extractor with sub-linear 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*

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

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

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

The extractor of this theorem is constructed by composing several known explicit constructions of pseudorandom objects with the merger of Theorem 5.4.3. In Section 5.5.1 we describe the construction of our basic extractor. We then show, in Section 5.5.2 how to use the 'repeated extraction' technique of Wigderson and Zuckerman [WZ99] 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).

Theorem 5.5.3 (Final extractor with sub-linear 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*

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

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

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

5.5.1 Proof of Theorem 5.5.2

Note that we may equivalently view an extractor $E : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ as a randomized algorithm $E : \{0,1\}^n \to \{0,1\}^m$ which is allowed to use d uniformly random bits. We will present the extractor E as such an algorithm which takes 5 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 [GUV07] 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 [DW08]. 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 divisions of 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 [TS96b], after passing to a convex combination, one of these divisions is a (k', k_2) block source with k' being at least $k - O(k/\Lambda)$ and k_2 being at least poly-logarithmic in k. In the third step we use a block source extractor (from [RRS00]) on each one of the possible Λ divisions (using the same seed for each division) 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 sub-constant. 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 poly-logarithmic. 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 known techniques) with logarithmic seed and an additional entropy loss of $O(\delta \cdot k')$.

We now formally prove Theorem 5.5.2. We begin by reducing to the case where $n = O(k)$ using the lossless condensers of [GUV07].

Theorem 5.5.4 (Lossless condenser [GUV07]) *For all integers positive $k < n$ with $k = \omega(\log(n))$, there exists an explicit function $C_{\mathrm{GUV}} : \{0,1\}^n \times \{0,1\}^{d'} \mapsto \{0,1\}^{n'}$ with $n' = 2k$, $d' = O(\log(n))$, such that for every (n,k)-source X, $C(\mathsf{X}, \mathsf{U}_{d'})$ is*

$(1/n)$-*close to an* (n', k)-*source, where* $\mathsf{U}_{d'}$ *is distributed uniformly over* $\{0,1\}^{d'}$, *and* $\mathsf{X}, \mathsf{U}_{d'}$ *are independent.*

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

By the above theorem, X' is $(1/n)$-close to an (n', k)-source, where $n' = 2k$. Our next goal is to produce a *somewhere-block source*. We now define these formally.

Definition 5.5.5 (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 5.5.6 (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 (k_1, k_2)-somewhere-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 random sources.*

We now state a result of Ta-Shma [TS96b] which converts an arbitrary source into a somewhere-block source. This is the first step in the proof of Theorem 1 on Page 44 of [TS96b] (Theorem 1 shows how 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

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

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

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

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

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

Theorem 5.5.7 ([TS96b]) *Let Λ be an integer. Let $k = k_1 + k_2 + s$. Then the function $B^{\Lambda}_{\mathrm{TS}} : \{0,1\}^{n'} \to (\{0,1\}^{n'})^{\Lambda}$ is such that for any (n', k)-source X', letting $\mathsf{X}'' = B^{\Lambda}_{\mathrm{TS}}(\mathsf{X}')$, 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 $X'' = (\mathsf{X}''_1, \mathsf{X}''_2, \ldots, \mathsf{X}''_{\Lambda}) = B^{\Lambda}_{\mathrm{TS}}(\mathsf{X}')$.

Plugging $k_2 = O(\log^4(n')) = O(\log^4(k))$, $s = O(\log n)$ and $k_1 = k - k_2 - s$ in the above theorem, we conclude that X'' 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 [RRS00] to convert the above somewhere-block source to a somewhere-random source.

Theorem 5.5.8 ([RRS00]) *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 $E_{\mathrm{RSW}} : \{0,1\}^{n_1} \times \{0,1\}^{n_2} \times \{0,1\}^{d''} \mapsto \{0,1\}^{m''}$ with $m'' = k'$, $d'' = O(\log(n'))$, such that for any (k', k_2)-block source X, $E_{\mathrm{RSW}}(\mathsf{X}, \mathsf{U}_{d''})$ is $(n_1)^{-\Omega(1)}$-close to the uniform distribution over $\{0,1\}^{m''}$, where $\mathsf{U}_{d''}$ is distributed uniformly over $\{0,1\}^{d''}$, and $\mathsf{X}, \mathsf{U}_{d''}$ are independent.*

Set $d'' = O(log(n'))$ as in Theorem 5.5.8.

Step 3: Pick $\mathsf{U}_{d''}$ uniformly from $\{0,1\}^{d''}$.
For each $j \in [\Lambda]$, compute $\mathsf{X}'''_j = E_{\mathrm{RSW}}(\mathsf{X}''_j, U_{d''})$.

By the above theorem, X''' is $n'^{-\Omega(1)}$-close to a somewhere-random source. We are now ready to use the merger M from Theorem 5.4.3. 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

$$d''' = O(\frac{1}{\delta}\log\frac{\Lambda}{\epsilon}) = O(c_1\log(n)).$$

Step 4: Pick $\mathsf{U}_{d'''}$ uniformly from $\{0,1\}^{d'''}$.
Compute $\mathsf{X}'''' = M(\mathsf{X}''', \mathsf{U}_{d'''})$.

By Theorem 5.4.3, X'''' is $O(\frac{1}{\log^{c_1}(n)})$-close to a $(k', (1-\delta)k')$-source. Note that $\delta = o(1)$, and thus X'''' has nearly full entropy. We now apply an extractor for sources with extremely-high entropy rate, given by the following lemma.

Lemma 5.5.9 *For any k' and $\delta > 0$, there exists an explicit $(k'(1-\delta), k'^{-\Omega(1)})$-extractor $E_{\mathrm{HIGH}} : \{0,1\}^{k'} \times \{0,1\}^{d''''} \mapsto \{0,1\}^{(1-3\delta)k'}$ with $d'''' = O(\log(k'))$.*

The proof of this lemma follows easily from Theorem 5.5.8. Roughly speaking, the input is partitioned into blocks of length $k' - \delta k - \log^4 k'$ and $\delta k' + \log^4 k'$. It follows that this partition is close to a $(k'(1-2\delta) - \log^4 k', \log^4 k')$-block source. This block source is then passed through the block-source extractor of Theorem 5.5.8.

Step 5: Pick $\mathsf{U}_{d''''}$ uniformly from $\{0,1\}^{d''''}$.
Compute $\mathsf{X}''''' = E_{\mathrm{HIGH}}(\mathsf{X}'''', \mathsf{U}_{d''''})$.
Output X'''''.

This completes the description of the extractor E. It remains to note that d, the total number of random bits used, is at most $d' + d'' + d''' + d'''' = O(c_1 \log n)$, and that the output X''''' is $O(\frac{1}{\log^{c_1} n})$-close to uniformly distributed over

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

This completes the proof of Theorem 5.5.2.

We summarize the transformations in the following table:

Function	Seed length	Input-type	Output-type
C_{GUV}	$O(\log(n))$	(n,k)-source	$(2k,k)$-source
$B^{\wedge}_{\text{TS}}$	0	$(2k,k)$-source	somewhere-$(k',\log^4(k))$-block
E_{RSW}	$O(\log(k))$	somewhere-$(k',\log^4(k))$-block	$(k',O(\log(n)))$-somewhere-random
M	$O(\log(n))$	$(k',O(\log(n)))$-somewhere-random	$(k',k'-o(k))$-source
E_{HIGH}	$O(\log(k))$	$(k',k'-o(k))$-source	$\mathsf{U}_{k'-o(k)}$

5.5.2 Improving the output length by repeated extraction

We now use some ideas from [RRS00] and [WZ99] to extract an even larger fraction of the min-entropy out of the source. This will prove Theorem 5.5.3. 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 5.5.10 (Explicit extractor with improved sub-linear 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 5.5.2 into a *strong* extractor (defined below) via [RRS00, Theorem 8.2] (which gives a generic way of getting a strong extractor from any extractor). We then use a trick from [WZ99] that repeatedy uses the same extractor with independent seeds to extract the 'remaining entropy' from the source, thus improving the entropy loss.

Definition 5.5.11 *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* $(E(\mathsf{X},\mathsf{U}_d),\mathsf{U}_d)$ *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 5.5.12 ([RRS00, 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 5.5.13 ([WZ99, 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.

We can now prove Theorem 5.5.10. Let E be the (k,ϵ)-extractor with seed $O(c_1 \log n)$ of Theorem 5.5.2. By Theorem 5.5.12, 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 5.5.13 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 5.5.13 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 Theorem 5.5.2 and Theorem 5.5.12. Thus,

$$d_i = O(ic_1 \log n).$$

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

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

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

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

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

Remark In fact [GUV07] and [RRV99] show how to extract *all* the minentropy with polylogarithmic seed length. Combined with the lossless condenser of [GUV07] this gives an extractor that uses logarithmic seed to extract all the minentropy from sources that have minetropy rate at most $2^{O(\sqrt{\log n})}$.

Theorem 5.5.14 (Corollary of [GUV07, 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 5.5.10 gives an extractor with improved sublinear entropy loss that works for sources of all entropy rates, thus completing the proof of Theorem 5.5.3.

5.6 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^iY^j appears in $P(X, Y)$ with a nonzero coefficient. Let $N(k, d, \theta)$ be the number of monomials X^iY^j which have $(1, k)$-degree at most d and $j \leq \theta d/k$. We have the following simple fact.

Fact 5.6.1 *For any $k < d$ and $\theta \in [0, 1]$, $N(k, d, \theta) > \theta \cdot (2 - \theta) \cdot \frac{d^2}{2k}$.*

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 5.6.2 (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:

$$\frac{m(m+1)}{2} \cdot n < N(k,d,\theta). \tag{5.1}$$

This can be made to hold by picking m sufficiently large, since by Fact 5.6.1,

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

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 the degree of $R_j(X)$ is at most d. By Proposition 2.1.5 and Lemma 2.1.4, we have for every $i \in I$,

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

Thus

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

By Lemma 2.1.8, we conclude that $R_j(X) = 0$. Since this holds for every $j < m - d/A$, we conclude that $\text{mult}(Q, f(X)) \geq m - d/A$.

We now complete the proof of the theorem. By the above discussion, for each $j \in [L]$, we know that $\text{mult}(Q, f_j(X)) \geq m - \frac{d}{\gamma n}$. Thus, by Lemma 2.1.8 (applied to the nonzero polynomial $Q(Y) \in \mathbb{F}(X)[Y]$ and the set of evaluation points $S = \{f_j(X) : j \in [L]\}$)

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

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. ■

5.7 Open Problems

We conclude with some open problems.

1. Construct explicit extractors with logarithmic seed which extract all the entropy out of a weak random source.

2. Construct explicit extractors with seeds of length $(1 + o(1)) \log n$ which extract nearly all the entropy out of a weak random source.

3. Do there exist Kakeya sets of size $q^n/2^n(1+o(1))$? It would be very interesting to know that there are settings where the extended method of multiplicities can give the sharp answer.

4. More philosophically, when do multiplicities help? For which problems is the extended method of multiplicities likely to be effective?

Chapter 6

Explicit Functions Uncorrelated with Low-degree Polynomials

6.1 Introduction

The fourth and final result of this thesis is the explicit construction of functions $f : \mathbb{F}_2^n \to \mathbb{F}_2$ that have small correlation with low-degree multivariate polynomials. The main idea underlying this construction is to reinterpret the vector space $\mathbb{F}_2^n$ as the large finite field $\mathbb{F}_{2^n}$; under this reinterpretation, our explicit constructions have a simple description and admit a simple analysis. We start by informally describing this method.

For simplicity, let us work over the two element field $\mathbb{F}_2$ (all our results generalize to larger fields). Suppose we have a problem involving functions from $\mathbb{F}_2^n$ to $\mathbb{F}_2$ that can be represented as low-degree multivariate polynomials (henceforth called "low-degree" functions). The heart of the method is to view f as a univariate polynomial over the extension field $\mathbb{F}_{2^n}$ and to consider the problem in the larger field. More precisely, embed $\mathbb{F}_2^n$ in $\mathbb{F}_{2^n}$ using an addition-respecting isomorphism. Using this embedding, view $f : \mathbb{F}_{2^n} \to \mathbb{F}_2 \subset \mathbb{F}_{2^n}$ as a univariate polynomial in $\mathbb{F}_{2^n}[X]$.

At first sight this method may seem counter-intuitive because even simple functions f can have very high degree when viewed as univariate polynomials over $\mathbb{F}_{2^n}$. To wit, the degree of the simple function $f(x_1, \ldots, x_n) = x_1$, when viewed as a poly-

nomial over $\mathbb{F}_{2^n}$, jumps from 1 to 2^{n-1}, because f vanishes on 2^{n-1} inputs. However, what comes to our aid is the following observation. Although f may have very large degree, it is "nicely structured" in the sense that it is a very sparse polynomial and moreover the exact location of its nonzero coefficients can be easily specified. For instance, it is well-known since the work of Ore in the 1930's [Ore33, Ore34] that any degree 1 function, when represented over $\mathbb{F}_{2^n}$ is *linearized*, meaning it can be written as $f(X) = c + \sum_{i=0}^{n-1} a_i X^{2^i}$ where $c, a_1, \ldots, a_{n-1} \in \mathbb{F}_{2^n}$. Similarly, f is of degree d over $\mathbb{F}_2$ if and only if its representation over $\mathbb{F}_{2^n}$ is of the form

$$c + \sum_{i_1,\ldots,i_k \in \binom{n}{\leq d}} a_{i_1,\ldots,i_k} X^{2^{i_1}+\ldots+2^{i_k}}. \tag{6.1}$$

Thus, we have nontrivial information about the number of nonzero coefficients of f, and the set of possibly nonzero coefficients of f. It is precisely this extra structure that we use to perform our analysis, which we describe next.

We can now describe some examples of our explicit exponentially-hard functions. Let $p(X) = X^\ell \in \mathbb{F}_{2^n}[X]$ where ℓ is odd and the binary expansion of ℓ has more than d ones. Then, the function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ given by $f(x) = \mathrm{Tr}(p(x))$is the first bit of $p(X)$ (or any other nontrivial linear combination of the bits of $p(X)$) has correlation at most $(4\ell^2/2^n)^{1/2^{d+2}}$ with degree d polynomials. In particular, taking $\ell = 2^{d+1} - 1$ we conclude that the function that computes the first bit $X^{2^{d+1}-1}$ has correlation at most $4 \cdot 2^{-n/2^{d+2}}$ with all degree d polynomials (see Theorem 6.2.5).

As an added bonus, this approach gives simple constructions of functions from $\mathbb{F}_2^n$ to $\mathbb{F}_2^m$ (with m large) that have exponentially small *agreement* with all m-tuples of low-degree polynomials, viewed as maps from $\mathbb{F}_2^n$ to $\mathbb{F}_2^m$ (here the agreement of two functions agree(f, g), is the fraction of points in $\mathbb{F}_2^n$ on which they have the same evaluation). Again the constructions are very simple and natural. For example if $m = n$, the same function $p(X) = X^\ell$ from above has exponentially small agreement with m-tuples of degree d polynomials (see Theorem 6.2.4).

One instance of the classical Weil bound for character sums shows that if $f : \mathbb{F}_{2^n} \to \mathbb{F}_2$ is a function of the form $\mathrm{Tr} \circ p$, where $p : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ is a low degree

univariate polynomial, then f has exponentially small correlation with any degree 1 polynomial $g : \mathbb{F}_2^n \to \mathbb{F}_2$, except when f is itself a degree 1 polynomial. Our result is a generalization of this phenomenon to $\mathbb{F}_2$ polynomials of higher degree d, namely, if f is the above form, then f has exponentially small correlation with any polynomial $g : \mathbb{F}_2^n \to \mathbb{F}_2$ of degree at most d, except when f is itself of $\mathbb{F}_2$-degree at most d.

In coding theory terminology, this result gives a dichotomy relating dual BCH codes and Reed-Muller codes. Let C_1 be the dual-BCH code contained in $\{0,1\}^{2^n}$ with parameter $t = O(1)$ (so that it has $n^O(1)$ codewords). Let C_2 bethe Reed-Muller code contained in $\{0,1\}^{2^n}$ of degree d polynomials, with $d = O(1)$. The for any codeword $c \in C_1$, one of the following cases must hold:

1. $c \in C_2$,

2. c is $1/2 - 2^{-\Omega(n)}$ far from all the codewords of C_2.

We now outline the method of proof. Let f denote our exponentially-hard function described above and let g be a degree d polynomial. Recall we are interested in bounding $|\mathbb{E}[(-1)^{f(x)-g(x)}]|$. As in [VW07] (who pioneered the use of the Gowers norm in this context), it suffices to show that the bias of the function after taking $d+1$ directional derivatives is small. This reduces our problem to understanding the $d+1^{\text{st}}$ derivative of our function f and to show that it is unbiased. Now we view f as a univariate polynomial in $\mathbb{F}_{2^n}$. Using the structure of its coefficients, we deduce that its $d+1^{\text{st}}$ derivative is a nonzero polynomial over $\mathbb{F}_{2^n}$ of relatively low degree. We finish the proof with an application of the Weil bound, Theorem 2.3.2, and conclude that f and g have exponentially small correlation.

Organization of this Chapter: In the next section we prove our main results on explicit functions uncorrelated with low-degree polynomials. We conclude with some open problems. For the sake of self-containedness, we include an exposition of an elementary proof of the Weil bound due to Bombieri and Stepanov in the appendix.

6.2 Low degree univariate polynomials hard for multivariate polynomials

In this section, we prove our main results on correlation (for the special case of $\mathbb{F}_2$). Preliminary material on the interplay between polynomials over $\mathbb{F}_2^n$ and $\mathbb{F}_{2^n}$ can be found in Section 2.3.

Let $q = 2^n$. We show that functions over $\mathbb{F}_q$ of low $\mathbb{F}_q$-degree but moderate $\mathbb{F}_2$-degree are "very far" from functions with low $\mathbb{F}_2$-degree. Our proofs use the machinery of discrete derivatives (generalized to functions between arbitrary vector spaces). See Section 2.3.2 for the necessary definitions.

At an intuitive level, the reason behind these results can be explained as follows. Functions that have low $\mathbb{F}_q$-degree are themselves very pseudorandom: they satisfy many equidistribution type properties such as the Schwartz-Zippel lemma and the Weil bound. Now when a function with low $\mathbb{F}_q$-degree is also known to have at least moderate $\mathbb{F}_2$-degree, this gives it some robustness: even after a few derivatives of this function have been taken, we know that the function remains non-zero (by the lower bound on $\mathbb{F}_2$-degree) and we know that the function is still a low-$\mathbb{F}_q$ degree function, which gives it some pseudorandomness properties. This robustness against derivatives is what makes it uncorrelate with low $\mathbb{F}_2$-degree polynomials, which simply vanish after we take a few derivatives.

We begin by proving a result that gives simple explicit functions having low agreement with polynomial-tuples. The proof is somewhat simpler in this case and contains many of the main ideas. A special case to keep in mind is $V = W = \mathbb{F}_q$ and $f(x) = x^{2^{d+1}-1}$.

Theorem 6.2.1 *Let V and W be $\mathbb{F}_q$ vector spaces. Suppose $f : V \to W$ is such that $\deg_{\mathbb{F}_2}(f) > d$ and $\deg_{\mathbb{F}_q}(f) \leq \ell$. Then for all $g : V \to W$ with $\deg_{\mathbb{F}_2}(g) \leq d$, we have*

$$\mathrm{agree}(g, f) \leq \left(\frac{\ell}{q}\right)^{1/2^{d+1}}.$$

Before starting with the proof, we state a simple lemma which gives a lower bound

on the number of "cubes" contained in a subset of $\mathbb{F}_2^n$. It is proved via a repeated application of the Cauchy-Schwarz Lemma, and is closely related to the proof of Lemma 2.2.5 dealing with the Gowers norm.

Lemma 6.2.2 *Let $S \subseteq \mathbb{F}_2^n$. Then*

$$\Pr_{x,a_1,\ldots,a_k \in \mathbb{F}_2^n}\left[\forall I \subseteq [k], (x + \sum_{i \in I} a_i) \in S\right] \geq \left(\frac{|S|}{2^n}\right)^{2^k}.$$

We now prove Theorem 6.2.1.

Proof The idea of the proof is to study how often the $d+1^{\text{st}}$ derivative of f vanishes. We then use this information to conclude that f must have low agreement with any function g with $\mathbb{F}_2$-degree at most d.

Define $h : V \times V^{d+1} \to W$ by

$$h(x, \mathbf{a}) = (D_{\mathbf{a}}(f - g))(x).$$

Fact 2.3.3 shows that $D_{\mathbf{a}}(g) = 0$ for all a. Thus, by linearity of $D_{\mathbf{a}}$, we have $h(x, \mathbf{a}) = (D_{\mathbf{a}} f)(x)$. Fact 2.3.3 implies that h is a non-zero function with $\mathbb{F}_q$-degree at most ℓ. The Schwartz-Zippel Lemma (Lemma 2.1.7) now shows that

$$\Pr_{x,\mathbf{a}}[h(x, \mathbf{a}) = 0] \leq \frac{\ell}{q}.$$

Let $S = \{x \in V : f(x) = g(x)\}$. By Equation (2.4), $h(x, \mathbf{a}) \neq 0$ implies that there exists $I \subseteq [d+1]$ such that $(f - g)(x + \sum_{i \in I} a_i) \neq 0$, or equivalently $x + \sum_{i \in I} a_i \notin S$. Therefore

$$\Pr_{x,a_1,\ldots,a_{d+1} \in V}\left[\forall I \subseteq [d], (x + \sum_{i \in I} a_i) \in S\right] \leq \Pr_{x \in V, \mathbf{a} \in V^{d+1}}[h(x, \mathbf{a}) = 0] \leq \frac{\ell}{q}.$$

The above inequality, combined with Lemma 6.2.2, shows that $\frac{|S|}{q} \leq \left(\frac{\ell}{q}\right)^{1/2^{d+1}}$, as desired. ■

Before we state our correlation bound, we make a definition.

Definition 6.2.3 (Odd function) *Let $f : \mathbb{F}_q \to \mathbb{F}_q$ and let*

$$f(x) = \sum_{i=0}^{q-1} \alpha_i x^i$$

be its (unique) polynomial representation. We say f is an odd function *if for all $i \leq q-1$, whenever i is even, $\alpha_i = 0$.*

We now prove a correlation bound for functions mapping to $\mathbb{F}_2$. In the language of coding theory, it states that if a codeword f of the dual of a BCH code of constant distance is not also an element of the Reed-Muller code of degree d polynomials, then it is $1/2 - 2^{-\Omega(n/2^d)}$ far from every codeword of that Reed-Muller code. As mentioned in the introduction, this is a generalization of the classical Weil bound that states the same thing for $d = 1$.

Theorem 6.2.4 *Suppose $f : \mathbb{F}_q \to \mathbb{F}_q$ is an odd function with $\mathbb{F}_2$-deg$(f) > d$ and $\mathbb{F}_q$-deg$(f) \leq \ell$. Then for all $g : \mathbb{F}_q \to \mathbb{F}_2$ with $\mathbb{F}_2$-deg$(g) \leq d$, we have*

$$\mathrm{Corr}(\mathrm{Tr} \circ f, g) \leq \left(\frac{4\ell^2}{q}\right)^{1/2^{d+2}}.$$

Remark Some hypothesis related to oddness is necessary in the previous theorem, to rule out functions f with $\mathrm{Tr}(f) = 0$ identically. The oddness propagates through the proof and finally plays a crucial role when we apply the Weil bound.

Proof Here we give the proof only for the case $f(x) = x^\ell$. This case already contains the main ideas, the case of general f involves just a few more details.

Our proof follows the same strategy as the proof of Theorem 6.2.1. Define $h : \mathbb{F}_q \times \mathbb{F}_q^{d+1} \to \mathbb{F}_q$ by

$$h(x, \mathbf{a}) = (D_{\mathbf{a}}(f))(x).$$

By Fact 2.3.3, h is a non-zero function of $\mathbb{F}_q$-degree at most ℓ.

$$\begin{aligned}
\left|\mathbb{E}_{x\in\mathbb{F}_q}\left[(-1)^{g(x)+\mathrm{Tr}(f(x))}\right]\right|^{2^{d+1}} &\leq \|(-1)^{g+\mathrm{Tr}\circ f}\|_{U^{d+1}}^{2^{d+1}} \quad \text{By Lemma 2.2.5} \\
&= \|(-1)^{\mathrm{Tr}\circ f}\|_{U^{d+1}}^{2^{d+1}} \quad \text{By Lemma 2.2.8, since } \deg_{\mathbb{F}_2}(g) \leq d. \\
&= \mathbb{E}_{x\in\mathbb{F}_q,\mathbf{a}\in\mathbb{F}_q^{d+1}}\left[(-1)^{D_{\mathbf{a}}(\mathrm{Tr}\circ f)(x)}\right] \\
&= \mathbb{E}_{x\in\mathbb{F}_q,\mathbf{a}\in\mathbb{F}_q^{d+1}}\left[(-1)^{\mathrm{Tr}((D_{\mathbf{a}}f)(x))}\right] \quad \text{as Tr is linear} \\
&= \mathbb{E}_{x\in\mathbb{F}_q,\mathbf{a}\in\mathbb{F}_q^{d+1}}\left[(-1)^{\mathrm{Tr}(h(x,\mathbf{a}))}\right]
\end{aligned}$$

We wish to bound this expression from above, i.e., we wish to show that the function h is unbiased. Our strategy is to partition the domain of h into lines, and show that on most lines the restriction of h to that line is unbiased. We do this by finding a univariate polynomial embedded in h, and applying the Weil bound to it. Proceeding with our chain of inequalities, we get

$$\begin{aligned}
\left|\mathbb{E}_{x\in\mathbb{F}_q}\left[(-1)^{g(x)+\mathrm{Tr}(f(x))}\right]\right|^{2^{d+1}} &\leq \mathbb{E}_{x\in\mathbb{F}_q,\mathbf{a}\in\mathbb{F}_q^{d+1},y\in\mathbb{F}_q^*}\left[(-1)^{\mathrm{Tr}(h(x,y\mathbf{a}))}\right] \\
&\quad \text{(as } \mathbf{a} \text{ and } y\mathbf{a} \text{ are identically distributed)} \\
&\leq \mathbb{E}_{x\in\mathbb{F}_q,\mathbf{a}\in\mathbb{F}_q^{d+1}}\left|\mathbb{E}_{y\in\mathbb{F}_q^*}\left[(-1)^{\mathrm{Tr}(h(x,y\mathbf{a}))}\right]\right|
\end{aligned}$$

At this point we pause to understand the expression $h(x, y\mathbf{a})$. By Fact 2.3.3, the polynomial $h(x, Y\mathbf{a}) \in \mathbb{F}_q[Y]$ may be written as $\sum_i x^{\ell-i}h_i(\mathbf{a})Y^i =: h_{x,\mathbf{a}}(Y) \in \mathbb{F}_q[Y]$. We will show that for most $(x, \mathbf{a})$, the function $h_{x,\mathbf{a}}(Y)$ satisfies the hypotheses of the Weil bound, Theorem 2.3.2.

To show that $h_{x,\mathbf{a}}(Y)$ cannot be written in the form $g(Y)^2 + g(Y) + c$, we will find an $i_0 > \ell/2$ with $2 \nmid i_0$, such that the coefficient of Y^{i_0} in $h_{x,\mathbf{a}}(Y)$ is nonzero. Together with the fact that $\deg(h_{x,\mathbf{a}}(Y) \leq \ell$, this implies that $h_{x,\mathbf{a}}(Y)$ cannot be written in the form $g(Y)^2 + g(Y) + c$.

By the remark after Fact 2.3.4, $i_0 > \ell/2$ with $2 \nmid i_0$, such that the coefficient $x^{\ell-i_0}h_{i_0}(\mathbf{a})$ of the monomial Y^{i_0} in $h(x, Y\mathbf{a})$ is a nonzero polynomial of $\mathbb{F}_q$-degree at most ℓ in the variables $(x, \mathbf{a})$. Thus, whenever $(x, \mathbf{a})$ are such that $x^{\ell-i_0}h_i(\mathbf{a})$ evaluates to something nonzero (which happens for most $(x, \mathbf{a})$, because of the Schwartz-Zippel

lemma), the polynomial $h_{x,\mathbf{a}}(Y)$ is not of the form $g(Y)^2 + g(Y) + c$, which is the hypothesis needed for the Weil bound. We now continue our calculation:

$$\begin{aligned}\left|\mathbb{E}_{x\in\mathbb{F}_q}\left[(-1)^{g(x)+\mathrm{Tr}(f(x))}\right]\right|^{2^{d+1}} &\leq \mathbb{E}_{x\in\mathbb{F}_q,\mathbf{a}\in\mathbb{F}_q^{d+1}}[\mathbf{1}_{x^{\ell-i_0}h_{i_0}(\mathbf{a})=0}]+ \\ &\quad \mathbb{E}_{x\in\mathbb{F}_q,\mathbf{a}\in\mathbb{F}_q^{d+1}}\left[\mathbf{1}_{x^{\ell-i_0}h_{i_0}(\mathbf{a})\neq 0}\cdot\left|\mathbb{E}_{y\in\mathbb{F}_q^*}\left[(-1)^{\mathrm{Tr}(h(x,y\mathbf{a}))}\right]\right|\right] \\ &\leq \frac{\ell}{q} + \ell q^{-1/2} \\ &\quad \text{(We bound the first term using the Schwartz-Zippel lemma 2.1.7.} \\ &\quad \text{We applied the Weil bound, as discussed above, to bound the second term.)} \\ &\leq 2\ell q^{-1/2}.\end{aligned}$$

Taking 2^{d+1}-th roots of both sides, the result follows. ■

Theorem 6.2.5 (Main) *There is an explicit function $f_0 : \mathbb{F}_2^n \to \mathbb{F}_2$ such that for any $g : \mathbb{F}_2^n \to \mathbb{F}_2$ with $\mathbb{F}_2$*

$$\mathrm{Corr}(f_0, g) \leq \frac{3}{2^{\frac{n}{2^{d+2}}}}.$$

Proof Let $f : \mathbb{F}_q \to \mathbb{F}_q$ be given by $f(x) = x^{2^{d+1}-1}$. We see that $\deg_{\mathbb{F}_q}(f) = 2^{d+1} - 1$. By Lemma 2.3.1, $\deg_{\mathbb{F}_2}(f) = d + 1$. Thus $f(x)$ satisfies the hypothesis of Theorem 6.2.4. Taking $f_0(x) = \mathrm{Tr} \circ f(x)$, we get the result. ■

These results generalize in a straightforward manner to low-degree polynomial-tuples mapping from $\mathbb{F}_p^n \to \mathbb{F}_p^m$ for arbitrary $\mathbb{F}_p$, m and n.

6.3 Open Problems

We conclude with some open problems.

1. Find explicit functions which are exponentially hard for polynomials of degree polylog(n). This would give strong average-case lower bounds and pseudorandom generators for $\mathsf{AC}^0[\oplus]$, a cherished goal of modern low-level complexity

theory. We believe that the kinds of functions considered here, namely traces of low-degree polynomials over a big field, should have this property.

Concretely, we conjecture that the function $f : \mathbb{F}_{2^n} \to \mathbb{F}_2$ given by $f(x) = \mathrm{Tr}(x^{-1})$ has $2^{-\Omega(n)}$ correlation with polynomials having $\mathbb{F}_2$-degree at most $\mathrm{polylog}(n)$. Apart from solving the open problem mentioned above, this would also imply that the natural operation of inversion over finite fields is average-case hard for $\mathsf{AC}^0[\oplus]$.

2. Traces of low-degree polynomials over a big field also appeared in our explicit constructions of affine dispersers in Chapter 4. Perhaps there is some interesting notion of pseudorandomness that these functions have which explains all these applications?

Appendix A

The elementary proof of the Weil bound

A.1 The Weil bound

Let $q = 2^n$. Let $\mathrm{Tr} : \mathbb{F}_q \to \mathbb{F}_2$ be the trace map. Our goal is to prove the Weil bound, which gives a bound on the number of zeroes of a function of the form $\mathrm{Tr}(f(x))$, where $f(X)$ is a low degree polynomial over $\mathbb{F}_q$.

Theorem A.1.1 (Weil bound) *For any $f(X) \in \mathbb{F}_q[X]$ of with degree exactly d, where d is an odd integer at most $2^{n/2-1}$, we have*

$$\left| \sum_{x \in \mathbb{F}_q} (-1)^{\mathrm{Tr}(f(x))} \right| \leq 2d2^{\lceil n/2 \rceil}.$$

We briefly comment why this form of the theorem implies the version in Theorem 2.3.2. Consider a polynomial $f(X)$ which is not of the form $g(X)^2 + g(X) + c$. Then by repeatedly replacing some monomials in $f(X)$ of the form aX^{2m} by $a^{1/2}X^m$, we can obtain another polynomial $f'(X)$ of odd degree, such that $\deg(f') \leq \deg(f)$, and for all $x \in \mathbb{F}_{2^n}$, $\mathrm{Tr}(f(x)) = \mathrm{Tr}(f'(x))$. This allows us to deduce Theorem 2.3.2 from Theorem A.1.1 (upto an $O(1)$ multiplicative factor).

The proof that we present is due to Bombieri, building on ideas of Stepanov. A related proof, also building on ideas of Stepanov, was given by Schmidt. An excellent reference for both proofs is [Sch04]. Our aim is to give an exposition of the proof that requires minimal prerequisites.

A.2 The plan

We begin with a standard lemma.

Lemma A.2.1 *For any $\alpha \in \mathbb{F}_q$, $\mathrm{Tr}(\alpha) = 0$ if and only if α is of the form $\beta^2 + \beta$, for some $\beta \in \mathbb{F}_q$.*

The (a, b)-*degree* of a monomial $X^i Y^j$ is defined to be $ai + bj$. The (a, b)-degree of a polynomial is the maximum of the (a, b)-degree of its monomials.

We will need the following version of Bezout's theorem:

Theorem A.2.2 *Let K be a field. Let $A(X, Y) \in K[X, Y]$ with X-degree at most d_X and Y-degree at most d_Y. Let $B(X, Y) \in K[X, Y]$ be relatively prime to $A(X, Y)$ and have (d_Y, d_X)-degree at most D. Then,*

$$|\{(x, y) \in K \times K : A(x, y) = B(x, y) = 0\}| \leq D.$$

This has an elementary proof using resultants.

Let $V = \{(x, y) \in \mathbb{F}_q \times \mathbb{F}_q : P(x, y) = 0\}$, where $P(X, Y) \in \mathbb{F}_q[X, Y]$ is the polynomial $Y^2 + Y + f(X)$. Using Lemma A.2.1, we see that $\sum_{x \in \mathbb{F}_q} (-1)^{\mathrm{Tr}(f(x))} = |V| - 2^n$. Therefore it suffices to show that

$$2^n - 2d2^{\lceil n/2 \rceil} \leq |V| \leq 2^n + 2d2^{\lceil n/2 \rceil}.$$

This is the form in which we will do the main argument.

Our strategy is as follows. We will find a polynomial $Q(X, Y)$ relatively prime to $P(X, Y)$ such that for any $(x, y) \in V$, $Q(x, y)$ is 0. Thus the cardinality of V is at at most the number of points of intersection of $P(X, Y) = 0$ and $Q(X, Y) = 0$. We

will then use Theorem A.2.2 to get an upper bound for this quantity. Applying this upper bound to a different function f, we will get the required lower bound too.

A.3 The execution

Let S be the set of all integers that can be written as either $2i$ or $2i + d$, for some nonnegative integer i. Let $S_j = \{s \in S : s \leq j\}$. For any $j \geq d$, we have $|S_j| = (j - (d-1)/2)$. For any $s \in S$, let $M_s(X,Y)$ be the monomial $X^{s/2}$ if s is even, or $X^{(s-d)/2}Y$ if s is odd. Notice that the $(2,d)$-degree of $M_s(X,Y)$ is s.

First observe that any polynomial $R(X,Y) \in \mathbb{F}_q[X,Y]$ of $(2,d)$-degree j, is congruent modulo $P(X,Y)$ to exactly one polynomial of Y-degree at most 1 (repeatedly replacing every occurrence of Y^2 by $Y + f(X)$). We denote this polynomial $\overline{R(X,Y)}$. In fact, the same argument shows that $\overline{R(X,Y)}$ has $(2,d)$-degree at most j, and is in the $\mathbb{F}_q$ linear span of $\{M_s(X,Y) : s \in S_j\}$. Clearly, the map $R(X,Y) \mapsto \overline{R(X,Y)}$ is $\mathbb{F}_q$-linear.

A.3.1 The upper bound

Let $r = \lfloor n/2 \rfloor$. Let k, ℓ be two integers (to be picked later) satisfying the following 3 conditions:

1. $k, \ell \geq d$,
2. $\ell < 2^{n-r}$,
3. $(k - (d-1)/2)(\ell - (d-1)/2) > (2^r \ell + k - (d-1)/2)$.

Also let $(a_{st})_{s \in S_k, t \in S_\ell}$ be formal variables over $\mathbb{F}_q$.

Consider the polynomial

$$A(X,Y) := \sum_{s \in S_k, t \in S_\ell} a_{st} M_s(X,Y) M_t(X,Y)^{2^r}.$$

Its $(2,d)$-degree is at most $2^r\ell + k$. Thus $\overline{A(X,Y)}$ is in the linear span of $\{M_u(X,Y) : u \in S_{2^r\ell+k}\}$. Thus the map sending $(a_{st})_{s \in S_k, t \in S_\ell}$ to $\overline{A(X,Y)}$ is a linear map from

a space of dimension $|S_k||S_\ell|$ to a space of dimension $|S_{2^r\ell+k}|$. Thus, as k, ℓ satisfy $(k-(d-1)/2)(\ell-(d-1)/2) > 2^r\ell + k - (d-1)/2$, we know that there is a nonzero $A(X,Y)$ of the above form such that $\overline{A(X,Y)} = 0$ (i.e., $P(X,Y)$ divides $A(X,Y)$). Take such an $A(X,Y)$.

We will now see how to construct the polynomial $Q(X,Y)$ that we wanted earlier. Let

$$Q(X,Y) := \sum_{s\in S_k, t\in S_\ell} a_{st}^{2^{n-r}} M_s(X,Y)^{2^{n-r}} M_t(X,Y).$$

Note that the $(2,d)$-degree of $Q(X,Y)$ is at most $2^{n-r}k + \ell$.

Let us now check that for any $(x,y) \in V$, $Q(x,y) = 0$. The crucial observation is:

$$Q(X,Y) = \sum_{s\in S_k, t\in S_\ell} a_{st}^{2^{n-r}} M_s(X,Y)^{2^{n-r}} M_t(X,Y)^{2^n} \mod \langle X^{2^n} - X, Y^{2^n} - Y\rangle \tag{A.1}$$

$$= A(X,Y)^{2^{n-r}} \mod \langle X^{2^n} - X, Y^{2^n} - Y\rangle. \tag{A.2}$$

Take $(x,y) \in V$. As $P(x,y) = 0$ and $P(X,Y)$ divides $A(X,Y)$, we conclude that $A(x,y) = 0$. Furthermore, since $x, y \in \mathbb{F}_q$, we have $x^{2^n} - x = 0$ and $y^{2^n} - y = 0$. The crucial observation above now implies that $Q(x,y) = 0$.

Thus V is contained in the set of all common solutions (x,y) of $Q(x,y) = P(x,y) = 0$.

Finally, let us show that $Q(X,Y)$ is relatively prime to $P(X,Y)$. Because f has odd degree, $P(X,Y)$ is irreducible. Thus it suffices to show that $\overline{Q(X,Y)}$ is nonzero. Note that:

$$\overline{Q(X,Y)} = \sum_{s\in S_k, t\in S_\ell} a_{st}^{2^{n-r}} \overline{M_s(X,Y)^{2^{n-r}} M_t(X,Y)}$$

The $(2,d)$-degree of any single term $a_{st}^{2^{n-r}} \overline{M_s(X,Y)^{2^{n-r}} M_t(X,Y)}$ is $2^{n-r}s+t$. Using the fact that $t \le \ell < 2^{n-r}$, we see that all terms have distinct $(2,d)$-degrees, and hence any nonzero linear combination of them must be nonzero. Thus $\overline{Q(X,Y)} \neq 0$, and so $P(X,Y)$ is relatively prime to $Q(X,Y)$.

We may now apply Theorem A.2.2, and conclude the number of common solutions (x, y) of $Q(x, y) = P(x, y) = 0$ is at most $2^{n-r}k + \ell$. Thus $|V| \leq 2^{n-r}k + \ell$.

Summarizing, we showed that for any k, ℓ satisfying

1. $k, \ell \geq d$,
2. $\ell < 2^{n-r}$,
3. $(k - (d-1)/2)(\ell - (d-1)/2) > (2^r \ell + k - (d-1)/2)$.

we have $|V| \leq 2^{n-r}k + \ell$.

Picking

$$k = 2^r + \frac{d-1}{2} + \frac{2^r \frac{d+1}{2}}{2^{n-r} - \frac{d+1}{2}}$$

and $\ell = 2^{n-r} - 1$, we get $|V| \leq 2^n + 2d2^{n-r}$.

By the discussion in the previous section, we conclude that for any $f(X)$ of degree d,

$$\sum_{x \in \mathbb{F}_q} (-1)^{\mathrm{Tr}(f(x))} \leq 2d2^{\lceil n/2 \rceil}.$$

A.3.2 The lower bound

Applying the upper bound to the polynomial $g(X) := f(X) + \alpha$, where $\alpha \in \mathbb{F}_q$ with $\mathrm{Tr}(\alpha) = 1$, we get

$$\sum_{x \in \mathbb{F}_q} (-1)^{\mathrm{Tr}(g(x))} \leq 2d2^{\lceil n/2 \rceil},$$

which, by choice of g, implies that

$$-\sum_{x \in \mathbb{F}_q} (-1)^{\mathrm{Tr}(f(x))} \leq 2d2^{\lceil n/2 \rceil}.$$

This completes the proof of Theorem A.1.1.

Bibliography

[BEH81] Andreas Blass, Geoffrey Exoo, and Frank Harary. Paley graphs satisfy all first-order adjacency axioms. *J. Graph Theory*, 5(4):435–439, 1981.

[Ber68] Elwyn R. Berlekamp. *Algebraic Coding Theory.* Mc Graw-Hill, revised 1984 edition, 1968.

[BGH+04] Eli Ben-Sasson, Oded Goldreich, Prahladh Harsha, Madhu Sudan, and Salil Vadhan. Robust PCPs of proximity, shorter PCPs and applications to coding. In ACM, editor, *Proceedings of the 36th Annual ACM Symposium on the Theory of Computing: Chicago, Illinois, USA, June 13–15, 2004*, pages 1–10, pub-ACM:adr, 2004. pub-ACM.

[BGH+05] Eli Ben-Sasson, Oded Goldreich, Prahladh Harsha, Madhu Sudan, and Salil P. Vadhan. Short PCPs verifiable in polylogarithmic time. In *IEEE Conference on Computational Complexity*, pages 120–134, 2005.

[BGK85] A. Blass, Y. Gurevich, and D. Kozen. A zero–one law for logic with a fixed point operator. *Information and Control*, 67:70–90, 1985.

[BGK06] J. Bourgain, A. A. Glibichuk, and S. V. Konyagin. Estimates for the number of sums and products and for exponential sums in fields of prime order. *J. London Math. Soc. (2)*, 73(2):380–398, 2006.

[BHRV01] Eli Ben-Sasson, S. Hoory, E. Rozenman, and S. Vadhan. Extractors for affine sources, unpublished manuscript. 2001.

[BIW04] Boaz Barak, Russell Impagliazzo, and Avi Wigderson. Extracting randomness using few independent sources. In *FOCS '04: Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 384–393, Washington, DC, USA, 2004. IEEE Computer Society.

[BKR06] Eli Ben-Sasson, Swastik Kopparty, and Jaikumar Radhakrishnan. Subspace polynomials and list decoding of reed-solomon codes. In *FOCS: IEEE Symposium on Foundations of Computer Science (FOCS)*, 2006.

[BKS+05] Boaz Barak, Guy Kindler, Ronen Shaltiel, Benny Sudakov, and Avi Wigderson. Simulating independence: New constructions of condensers, ramsey graphs, dispersers, and extractors. In *STOC: ACM Symposium on Theory of Computing (STOC)*, 2005.

[BKT04] J. Bourgain, N. Katz, and T. Tao. A sum-product estimate in finite fields, and applications. *Geom. Funct. Anal.*, 14(1):27–57, 2004.

[BNS89] L. Babai, N. Nisan, and M. Szegedy. Multiparty protocols and logspace-hard pseudorandom sequences. In *STOC: ACM Symposium on Theory of Computing (STOC)*, 1989.

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

[BR05] A. Blass and B. Rossman. Explicit graphs with extension properties. *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS*, (86):166–175, 2005.

[BRSW06] Boaz Barak, Anup Rao, Ronen Shaltiel, and Avi Wigderson. 2-source dispersers for sub-polynomial entropy and ramsey graphs beating the frankl-wilson construction. In *STOC '06: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 671–680, New York, NY, USA, 2006. ACM.

[BS05] Eli Ben-Sasson and Madhu Sudan. Simple PCPs with poly-log rate and query complexity. In *STOC '05: Proceedings of the 37th annual ACM Symposium on Theory of Computing*, pages 266–275, New York, NY, USA, 2005. ACM Press.

[BV07] A. Bogdanov and E. Viola. Pseudorandom bits for polynomials. In *FOCS*, pages 41–51, 2007.

[CB04] Yuval Cassuto and Jehoshua Bruck. A combinatorial bound on the list size. *Paradise Laboratory Technical report, California Institute of Technology*, 2004.

[CGH+85] Benny Chor, Oded Goldreich, Johan Håstad, Joel Friedman, Steven Rudich, and Roman Smolensky. The bit extraction problem of t-resilient functions (preliminary version). In *FOCS*, pages 396–407. IEEE, 1985.

[DG09] Matt DeVos and Ariel Gabizon. Simple affine extractors using dimension expansion. *Electronic Colloquium on Computational Complexity (ECCC)*, (63), 2009.

[DS07] Z. Dvir and A. Shpilka. An improved analysis of linear mergers. *Comput. Complex.*, 16(1):34–59, 2007. (Extended abstract appeared in RANDOM 2005).

[Dvi08] Z. Dvir. On the size of Kakeya sets in finite fields. *J. AMS (to appear)*, 2008.

[DW08] Zeev Dvir and Avi Wigderson. Kakeya sets, new mergers and old extractors. In *FOCS*, pages 625–633. IEEE Computer Society, 2008.

[Fag74] R. Fagin. Generalized first–order spectra and polynomial–time recognizable sets. In R. M. Karp, editor, *Complexity of Computation, SIAM-AMS Proceedings, Vol. 7*, pages 43–73, 1974.

[Fag76] R. Fagin. Probabilities on finite models. *Journal of Symbolic Logic*, 41:50–58, 1976.

[GKLT69] Y. V. Glebskii, D. I. Kogan, M. I. Liogonki, and V. A. Talanov. Range and degree of realizability of formulas in the restricted predicate calculus. *Cybernetics*, 5:142–154, 1969.

[Gow01] W. T. Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.*, 11(3):465–588, 2001.

[GR05] Ariel Gabizon and Ran Raz. Deterministic extractors for affine sources over large fields. In *FOCS '05: Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 407–418, Washington, DC, USA, 2005. IEEE Computer Society.

[GR06] Venkatesan Guruswami and Atri Rudra. Explicit capacity-achieving list-decodable codes. In Jon M. Kleinberg, editor, *STOC*, pages 1–10. ACM, 2006.

[GRS06] Gabizon, Raz, and Shaltiel. Deterministic extractors for bit-fixing sources by obtaining an independent seed. *SICOMP: SIAM Journal on Computing*, 36, 2006.

[GS71] R. L. Graham and J. H. Spencer. A constructive solution to a tournament problem. *Canad. Math. Bull.*, 14:45–48, 1971.

[GS99] Venkatesan Guruswami and Madhu Sudan. Improved decoding of Reed-Solomon and algebraic-geometric codes. *IEEE Transactions on Information Theory*, 45:1757–1767, 1999.

[GT08] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions. *Ann. of Math. (2)*, 167(2):481–547, 2008.

[GUV07] Venkatesan Guruswami, Christopher Umans, and Salil P. Vadhan. Unbalanced expanders and randomness extractors from parvaresh-vardy codes. In *IEEE Conference on Computational Complexity*, pages 96–108. IEEE Computer Society, 2007.

[HKL96] L. Hella, Ph.G. Kolaitis, and K. Luosto. Almost everywhere equivalence of logics in finite model theory. *Bulletin of Symbolic Logic*, 2(4):422–443, 1996.

[HKT08] J. W. P. Hirschfeld, G. Korchmaros, and F. Torres. *Algebraic Curves over a Finite Field (Princeton Series in Applied Mathematics)*. Princeton University Press, 2008.

[HLX02] Xiang-Dong Hou, Ka Hin Leung, and Qing Xiang. A generalization of an addition theorem of kneser. *Journal of Number Theory*, 97(1):1 – 9, 2002.

[HS09] Elad Haramaty and Amir Shpilka. On the structure of cubic and quartic polynomials. *Electronic Colloquium on Computational Complexity (ECCC)*, (80), 2009.

[KaS] Algebraic property testing: the role of invariance.

[KRVZ06] Jesse Kamp, Anup Rao, Salil P. Vadhan, and David Zuckerman. Deterministic extractors for small-space sources. In *STOC*, pages 691–700, 2006.

[KV87] Ph. G. Kolaitis and M. Y. Vardi. The decision problem for the probabilities of higher-order properties. In *Proc. 19th ACM Symp. on Theory of Computing*, pages 425–435, 1987.

[KV90] Ph. G. Kolaitis and M. Y. Vardi. 0-1 laws and decision problems for fragments of second-order logic. *Information and Computation*, 87:302–338, 1990.

[KZ07] Jesse Kamp and David Zuckerman. Deterministic extractors for bit-fixing sources and exposure-resilient cryptography. *SIAM J. Comput.*, 36(5):1231–1247, 2007.

[LN97] Rudolf Lidl and Harald Niederreiter. *Finite fields*, volume 20 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, second edition, 1997. With a foreword by P. M. Cohn.

[Lov08] S. Lovett. Unconditional pseudorandom generators for low degree polynomials. In *STOC*, pages 557–562, 2008.

[LRVW03] Chi-Jen Lu, Omer Reingold, Salil Vadhan, and Avi Wigderson. Extractors: Optimal up to constant factors. In *Proceedings of the 35th Annual ACM Symposium on Theory of Computing"*, 2003.

[NNT05] M. Naor, A. Nussboim, and E. Tromer. Efficiently constructible huge graphs that preserve first order properties of random graphs. In *TCC*, pages 66–85, 2005.

[Ore33] O. Ore. On a special class of polynomials. *Trans. Amer. Math. Soc.*, 35(3):559–584, 1933.

[Ore34] O. Ore. Contributions to the theory of finite fields. *Trans. Amer. Math. Soc.*, 36(2):243–274, 1934.

[PS89] L. Pacholski and W. Szwast. The 0-1 law fails for the class of existential second-order Gödel sentences with equality. In *Proc. 30th IEEE Symp. on Foundations of Computer Science*, pages 280–285, 1989.

[PV05] Farzad Parvaresh and Alexander Vardy. Correcting errors beyond the guruswami-sudan radius in polynomial time. In *FOCS*, pages 285–294, 2005.

[Raz87a] Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *MATHNASUSSR: Mathematical Notes of the Academy of Sciences of the USSR*, 41, 1987.

[Raz87b] A. Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *MATHNASUSSR: Mathematical Notes of the Academy of Sciences of the USSR*, 41, 1987.

[RRS00] O. Reingold and and A. Wigderson R. Shaltiel. Extracting randomness via repeated condensing. In *Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science*, 2000.

[RRV99] Ran Raz, Omer Reingold, and Salil P. Vadhan. Extracting all the randomness and reducing the error in trevisan's extractors. In *STOC*, pages 149–158, 1999.

[Sch04] Wolfgang Schmidt. *Equations over finite fields: an elementary approach.* Kendrick Press, Heber City, UT, second edition, 2004.

[Sha02] Ronen Shaltiel. Recent developments in explicit constructions of extractors. *Bulletin of the EATCS*, 77:67–95, 2002.

[Smo87] R. Smolensky. Algebraic methods in the theory of lower bounds for boolean circuit complexity. In *STOC*, pages 77–82, 1987.

[SS87] J. Spencer and S. Shelah. Threshold spectra for random graphs. In *Proc. 19th ACM Symp. on Theory of Computing*, pages 421–424, 1987.

[SS88] S. Shelah and J. Spencer. Zero-one laws for sparse random graphs. *J. Amer. Math. Soc.*, 1:97–115, 1988.

[SS08] Shubhangi Saraf and Madhu Sudan. Improved lower bound on the size of kakeya sets over finite fields. *Analysis and PDE (to appear)*, 2008.

[Sud97] Madhu Sudan. Decoding of Reed-Solomon codes beyond the error-correction bound. *Journal of Complexity*, 13(1):180–193, 1997.

[TS96a] A. Ta-Shma. On extracting randomness from weak random sources. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on the Theory of Computing, Philadelphia, Pennsylvania, USA*, pages 276–285, 1996.

[TS96b] A. Ta-Shma. *Refining Randomness.* PhD thesis, The Hebrew University, Jerusalem, Israel, 1996.

[TV00] Luca Trevisan and Salil P. Vadhan. Extracting randomness from samplable distributions. In *FOCS*, pages 32–42, 2000.

[Vio08] E. Viola. The sum of d small-bias generators fools polynomials of degree d. In *IEEE Conference on Computational Complexity*, pages 124–127, 2008.

[VW07] E. Viola and A. Wigderson. Norms, xor lemmas, and lower bounds for gf(2) polynomials and multiparty protocols. In *22th IEEE Conference on Computational Complexity (CCC)*, 2007.

[Wol99] T. Wolff. Recent work connected with the Kakeya problem. In *Prospects in Mathematics*, pages 129–162, 1999.

[WZ99] Avi Wigderson and David Zuckerman. Expanders that beat the eigenvalue bound: explicit construction and applications. In *Combinatorica*, volume 19, pages 125–138, 1999.

[Zuc06] David Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. In *STOC '06: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 681–690, New York, NY, USA, 2006. ACM.

[Zuc07] David Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. *Theory of Computing*, 3(1):103–128, 2007.