7-1-2021

A Study of Sparse Representation of Boolean Functions

Yekun Xu
yxu040@fiu.edu

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

A STUDY OF SPARSE REPRESENTATION OF BOOLEAN FUNCTIONS

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

COMPUTER SCIENCE

by

Yekun Xu

2021

To: Dean John L. Volakis
College of Engineering and Computing

This dissertation, written by Yekun Xu, and entitled A Study of Sparse Representation of Boolean Functions, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

Dong Chen

Mohammad Hadi Amini

S. S. Iyengar

Xiaosheng Li

Ning Xie, Major Professor

Date of Defense: July 1, 2021

The dissertation of Yekun Xu is approved.

Dean John L. Volakis
College of Engineering and Computing

Andrés G. Gil
Vice President for Research and Economic Development
and Dean of the University Graduate School

Florida International University, 2021

DEDICATION

This thesis is dedicated to my parents, Jing Sui and Hanyang Xu, for their unlimited love and support.

And also to my advisor Ning Xie and my labmates Shuai Xu, for their help and companion during my study.

ABSTRACT OF THE DISSERTATION

A STUDY OF SPARSE REPRESENTATION OF BOOLEAN FUNCTIONS

by

Yekun Xu

Florida International University, 2021

Miami, Florida

Professor Ning Xie, Major Professor

The Boolean function is one of the most fundamental computation models in theoretical computer science. The two most common representations of Boolean functions are Fourier transform and real polynomial form. Applying analytic tools under these representations to the study Boolean functions has led to fruitful research in many areas such as complexity theory, learning theory, inapproximability, pseudorandomness, metric embedding, property testing, threshold phenomena, social choice and etc.

In this thesis, we focus on *sparse representations* of Boolean function in both Fourier transform and polynomial form, and obtain the following new results.

A classical result of Rothschild and van Lint asserts that if every non-zero Fourier coefficient of a Boolean function f over $\mathbb{F}_2^n$ has the same absolute value, namely $|\hat{f}(\alpha)| = 1/2^k$ for every α in the Fourier support of f, then f must be the indicator function of some affine subspace of dimension $n - k$. Here we slightly generalize their result, and show that Boolean functions whose Fourier coefficients take values in the set $\{-2/2^k, -1/2^k, 0, 1/2^k, 2/2^k\}$ are indicator functions of two disjoint affine subspaces of dimension $n - k$ or four disjoint affine subspaces of dimension $n - k - 1$. Our main technical tools are results from additive combinatorics which offer tight bounds on the affine span size of a subset of $\mathbb{F}_2^n$ when the doubling constant of the subset is small.

For polynomial representation of a Boolean functions, we study the distribution of the number of non-zero coefficients of *random* Boolean functions. For a random Boolean

function $f : \{0,1\}^n \to \{0,1\}$, i.e., a function whose value at each point on the Boolean cube is chosen independently and uniformly at random from $\{0,1\}$, in real polynomial representation, we give several bounds and concentration results about the distribution of the sparsity of f.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

The fundamental problem in complexity theory [AB09] is to determine which problems are easy to solve computationally and which ones require more time or space for computers to find solutions. During the last half century, many scientists have contributed to this area and have developed many fruitful theories, such as NP-completeness [Coo71, Lev73, GJ79], PCP theorems for hardness of approximations [FGL$^+$96, AS98, ALM$^+$98], and parameterized complexity [FL87, ADF95, PY96, BFR98, DF12]. For a given problem, we may be able to design efficient algorithms, which lead to upper-bounds for the complexity; or the least time needed to solve the problem, which leads to lower-bounds. If the gap between upper and lower bound are asymptotically negligible, we established the tight complexity result of the problem. For example, a classical result is that sorting n numbers using any comparison-based algorithm takes $\Theta(n \log n)$ time.

Although different hardware may have different computation power, causing different running time for the same problem on different computers, we still have theoretical methods to measure the complexity. For example, we could use the number of bit operations, the size of circuit that computes the problem, or the simplicity of the function computes the problem as measurements. Moreover, sometimes the model of question may not rely on a single computing device, where the input data spread among multiple persons or computers. We could measure the difficulty by the number of communications between the multiple parties, which leads to the area of *communication complexity* [Yao79, KN97].

Despite much success, unfortunately, after decades of intensive research, there are numerous computational problems of which tight complexity bounds are still elusive. Of central importance to both computer science and mathematics is the well-know P versus NP problem: for all NP problems, can they be solved in polynomial times in terms of input

size? Any state of the art algorithms for NP-complete problems still take exponential running time.

A systematic approach for studying computation complexity may start with formally modelling computational problems in a most generalized way. Note that almost all problems that can be computed by a computer can be modelled well by one or a collection of *Boolean functions* $f : \{0,1\}^n \to \{0,1\}$, since we can always encode the input and output data into finite binary bits $\{0,1\}$. Any distributed problem involving multiple parties can be transformed to multiple subproblems with only two parties involved. These simple observations explain why the study of *Analysis of Boolean functions* has grown into one of the central research areas in the past thirty years in theoretical computer science. We refer interested readers to [Juk12, O'D14] for comprehensive treatments.

1.1 Boolean functions

'Boolean', which represents the work derived by George Boole, is the subfield of mathematics handling variables only in two values, true or false, or equivalently 1 or 0, respectively. Boolean cube of dimension n, $\{0,1\}^n$ can be considered as the set of all n-length boolean vectors or binary strings. Every n-length binary string $S \in \{0,1\}^n$ can also be considered as a subset of $[n] = \{1, 2, \cdots, n\}$. Any function $f : \{0,1\}^n \to \mathbb{R}$ defined on the Boolean cube will be assigned a value $f(S)$ for any input S.

For every function f defined on a Boolean cube, we may consider the input as n Boolean variables representing true or false, with arbitrary outputs. In the area of computer science, we usually focus on Boolean functions which restrict the output to Boolean variables only. Naturally, we can represent function f as a circuit of logical gates with n bits of Boolean variables as input, returning a single bit as output. The circuit may contain complicated multi-way gates, but using basic 2-way AND/OR/XOR gates would be enough to represent all functions.

The most common alternative way to characterize a function is by providing a polynomial that computes f. In order to compute a polynomial, we have to assign real numbers to represent the input bits.

The standard polynomial representation is assigning 0 to false and 1 to true, while multiplications are equal to logical AND gates. For each monomial, $\prod_{i \in S} x_i$ equals to 1 if and only if all the $x_i \in S$ are 1(true).

Another most widely used representation is called Fourier representation, which assigns 1 to false and -1 to true. In this representation, multiplications are equal to logical XOR gates. Then for each monomial in Fourier representation, $\prod_{i \in S} x_i$ equals to 1 if and only if the number of negative(true) $x_i \in S$ are even.

1.2 Relations between two representations

Both representations have a good property: a unique multi-linear form. For standard representation, we have $x_i^k = x_i$ for any positive k, while for Fourier representation, we have $x_i^2 = 1$ being a constant. And for all S, the 2^n terms $\prod_{i \in S} x_i$ will form a basis of the 2^n-dimension vector space for functions defined on Boolean cube.

Furthermore, there exists a direct relation between the two forms. Supposing that $f_R : \{0,1\}^n \to \mathbb{R}$ and $f_F : \{1,-1\}^n \to \mathbb{R}$ both represents the function $f : \{false, true\}^n \to \mathbb{R}$. They can be rewritten as follows: $f_R = \sum_S c_S \prod_{i \in S} x_i$ and $f_F = \sum_S \hat{f}(S) \prod_{i \in S} \widetilde{x_i}$, while $x_i = \frac{1-\widetilde{x_i}}{2}$ and $\widetilde{x_i} = 1 - 2x_i$. c_S and $\hat{f}(S)$ are called polynomial coefficients and Fourier coefficients.

By multiplying out the equations, we get the following two transformation formulas.

$$f_R(x) = \sum_S c_S \prod_{i \in S} x_i = \sum_S c_S \prod_{i \in S} \frac{1 - \widetilde{x_i}}{2} = \sum_S \frac{c_S}{2^{|S|}} \left(\sum_{T \subset S} (-1)^{|T|} \prod_{i \in T} \widetilde{x_i} \right)$$

$$\Rightarrow \hat{f}(S) = (-1)^{|S|} \sum_{T \supset S} \frac{c_T}{2^{|T|}}$$

$$f_F(x) = \sum_S \hat{f}(S) \prod_{i \in S} \widetilde{x_i} = \sum_S \hat{f}(S) \prod_{i \in S} (1 - 2x_i) = \sum_S \hat{f}(S) \left(\sum_{T \subset S} (-2)^{|T|} \prod_{i \in T} x_i \right)$$

$$\Rightarrow c_S = (-2)^{|S|} \sum_{T \supset S} \hat{f}(T).$$

Although the coefficients differ substantially, the basis generated by Fourier representations is orthogonal, while the basis by standard polynomials is not. These two representations are essentially equivalent, represent the same functions, and can be easily converted via the above formulas. The study of properties of one representation will provide better understanding of the other.

In this thesis, we contributed to the understanding of sparsity of both representations of Boolean functions. For Fourier representation, we provide brand-new results, which could characterize the structure of the function support based only on the magnitude of the Fourier coefficients of the functions under specific circumstances. For real polynomial representation, we give statistical results to improve the understanding of the distribution and concentration of sparsity for all Boolean functions.

CHAPTER 2

A GENERALIZATION OF A THEOREM OF ROTHSCHILD AND VAN LINT

2.1 Introduction

One of the most fruitful approaches in functional analysis is to represent functions as sums of simple and well-structured objects, such as sine wave functions and polynomials. Such representations often provide additional insights on the combinatorial structures of or complexity measures associated with the subjects under consideration. This paradigm in theoretical computer science has witnessed harmonic analysis on the cube, or the discrete Fourier transform of Boolean functions, emerged in the past three decades as a powerful and versatile tool that finds numerous applications in complexity theory (such as PCP and circuit complexity), property testing, learning, cryptography, coding theory, social choice theory and others; see [O'D14] for a comprehensive survey.

Fourier coefficients and function values are two equivalent ways to represent a function. That is, the Fourier spectrum of a function completely determines the function-value at any point on the cube. However, knowing only the *values* of the Fourier spectrum but without the information of the locations of these values in the Fourier space in general leaves the function undetermined to a large extent, even restricted to Boolean functions. To see this, consider the following examples. Generally speaking, we view two Boolean functions as the same function if they are *isomorphic*. More formally, we say that two Boolean functions $f, g : \mathbb{F}_2^n \to \{0,1\}$ are *isomorphic* to each other if there is an invertible linear transformation $L : \mathbb{F}_2^n \to \mathbb{F}_2^n$ such that $g(x) = Lf(x)$ for every $x \in \mathbb{F}_2^n$, where $Lf(x) := f(Lx)$. Now consider the following two families of Boolean functions $\{f_k : \mathbb{F}_2^k \to \{0,1\} \mid k \in \mathbb{N}, k \geq 3\}$ and $\{g_k : \mathbb{F}_2^k \to \{0,1\} \mid k \in \mathbb{N}, k \geq 3\}$, with the Fourier expansions of $f_k(x) = \frac{3}{4} - \frac{1}{4}\chi_{\{1\}}(x) - \frac{1}{4}\chi_{\{2\}}(x) - \frac{1}{4}\chi_{\{1,2\}}(x)$ and $g_k(x) = \frac{3}{4} - \frac{1}{4}\chi_{\{1,2\}}(x) - \frac{1}{4}\chi_{\{1,3\}}(x) - \frac{1}{4}\chi_{\{2,3\}}(x)$. One can check easily that both f_k

and g_k are indeed Boolean functions and the multisets of non-zero Fourier coefficients are both $\{\frac{3}{4}, -\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}\}$. On the other hand, the Fourier dimension — dimension of the subspace spanned by vectors at which the function's Fourier coefficients are non-zero — of f_k is 2 while the Fourier dimension of g_k is 3. Since the Fourier spectrum transforms according to $(L^T)^{-1}$ when the function undergoes the linear transformation L, it follows that there is no invertible linear transformation L that maps f_k to g_k, i.e. they are not isomorphic to each other. Another such example is the class of *address functions* $f_n : \mathbb{F}_2^n \to \{-1, 1\}$, where $n = k + 2^k$ for some positive integer k, together with the class of functions $g_n : \mathbb{F}_2^n \to \{-1, 1\}$ formed by tensoring some *bent function* on $2k$-bits with a δ-function on $n - 2k$ bits. Then both f_n and g_n have 2^{2k} non-zero Fourier coefficients, with $2^{2k-1} + 2^{k-1}$ of them taking value $1/2^k$ and $2^{2k-1} - 2^{k-1}$ of them taking value $-1/2^k$; moreover, since the Fourier dimension of f_n is n and the Fourier dimension of g_n is $2k < n$, these two functions are not isomorphic to each other.

Nevertheless, there are a few exceptions to the general phenomenon in the sense that knowing only the values of the Fourier spectrum completely determine the Boolean function, up to an isomorphism. One such example is the indicator function of an affine subspace, which enjoys a very simple Fourier spectrum. Specifically, if f is the indicator function of an affine subspace in $\mathbb{F}_2^n$ of dimension $n - k$, then it is straightforward to check that every non-zero Fourier coefficient of f is either $1/2^k$ or $-1/2^k$. What about the converse? Namely, if we know that the non-zero Fourier coefficients of a Boolean function all have magnitude $1/2^k$, then what can be said about the function?

2.1.1 Rothschild and van Lint Theorem

Rothschild and van Lint [RvL74] (see also Chapter 13, Lemma 6 in [MS77]) proved the following theorem:

Theorem 2.1.1. *Let $n \geq 1$ and $0 \leq k \leq n$. Let $f = \mathbb{1}_S$ be the indicator function of a set $S \subseteq \mathbb{F}_2^n$ of size $|S| = 2^{n-k}$. If for every $\alpha \in \mathbb{F}_2^n$, $|\hat{f}(\alpha)|$ is equal to either zero or $1/2^k$, then S is an affine subspace of dimension $n - k$.*

In other words, Rothschild and van Lint Theorem shows that, up to an invertible linear transform, we have a complete characterization when the Fourier coefficients of a *Boolean* function are all from the set $\{-1/2^k, 0, 1/2^k\}$: the Boolean function must be the indicator of some affine subspace of co-dimension k.

A natural question is: how far can we extend such a nice characterization in terms of the values of Fourier coefficients only? Following [GOS+11], for a rational number x, the *granularity* $\mathrm{gran}(x)$ of x is defined to be the least nonnegative integer k such that $x = m/2^k$, where m is an (odd) integer. A function $\mathbb{F}_2^n \to \mathbb{R}$ is said to be k-*granular* if the maximum granularity of its Fourier coefficients is k — that is, $k = \max_\alpha\{\mathrm{gran}(\hat{f}(\alpha))\}$. For a Boolean function, its granularity is known to be intimately correlated with its *Fourier sparsity* [GOS+11] — the number of non-zero Fourier coefficients; see discussion in Section 2.1.4 for more details. Therefore, one can view Rothschild and van Lint Theorem as a characterization of k-granular Boolean functions with minimum support size (that is, $\hat{f}(\mathbf{0}) = |\{x : f(x) = 1\}|/2^n = 1/2^k$).

2.1.2 Our results

In this chapter, we slightly generalize Rothschild and van Lint Theorem to give a complete characterization of k-granular Boolean functions of support size $2^n \cdot 2/2^k = 2^{n-k+1}$. Roughly speaking, our main theorem is the following:

Theorem 2.1.2 (Informal statement)**.** *For large enough integers $n \geq k$, if a Boolean function $f : \mathbb{F}_2^n \to \{0, 1\}$ has all its Fourier coefficients in the set $\{0, \frac{\pm 1}{2^k}, \frac{\pm 2}{2^k}\}$, then f is the indicator function of disjoint union of two affine subspaces of dimension $n - k$.*

Our Main Theorem is based on the following Main Lemma, which deals with the general case of $k \geq 5$, together with case analysis[1] for small values of k.

Lemma 2.1.3 (Main)**.** *Let $k \geq 5$ and $n \geq k$ be integers. Let $f : \mathbb{F}_2^n \to \{0,1\}$ be a Boolean function such that $\hat{f}(\mathbf{0}) = 1/2^{k-1}$ and any other Fourier coefficients are either zero or equal to $\pm\frac{1}{2^k}$, then f is the indicator function of a disjoint union of two dimension $n-k$ affine subspaces.*

2.1.3 Proof overview and our techniques

The original form of Rothschild and van Lint Theorem was stated to characterize subspaces in affine geometry and projective geometry. For completeness and more importantly, because the first step in our proof of the main theorem follows a similar strategy, we present a slightly different proof using the notation of Fourier analysis.

A proof of Rothschild and van Lint Theorem. We prove the theorem by induction on n. It is trivial to see that the theorem holds for $n = 1$ (for both $k = 0$ and $k = 1$). Let $n \geq 2$. Clearly there is nothing to prove for $k = 0$ and $k = n$, so we assume $0 < k < n$. Note that $\hat{f}(\mathbf{0}) = |S|/2^n = 1/2^k$, then by Parseval's identity, there exists a non-zero α such that $\hat{f}(\alpha) = 1/2^k$ or $-1/2^k$. Assume that $\hat{f}(\alpha) = 1/2^k$ and the case of $\hat{f}(\alpha) = -1/2^k$ is similar. Applying an invertible linear transform L that maps α to e_1, where e_1 stands for the standard basis vector $(1, 0, \ldots, 0)$. Note that both the Fourier spectrum of f and any affine subspace are invariant under invertible linear transformations, hence it suffices to argue about $g := Lf$. Now we have $\hat{g}(\mathbf{0}) = \hat{g}(e_1) = 1/2^k$. Applying a linear restriction over the first bit of the input to get sub-functions g_0 and g_1 (see Proposition 2.2.3 in 2.2.3

[1]The need for a nasty case analysis stems from a key lemma in the proof, namely Lemma 2.3.9, which holds only when $k \geq 5$.

for details). By (2.2), $\hat{g}_1(\mathbf{0}) = \hat{g}(\mathbf{0}) - \hat{g}(e_1) = 0$, which implies that g_1 is the zero-function. This implies that S is completely contained in the support of g_0 and moreover, by (2.3), $\hat{g}_0(\beta) = 2\hat{f}(0,\beta)$ for every $\beta \in \mathbb{F}_2^{n-1}$. In other words, g_0 is a Boolean function over $\mathbb{F}_2^{n-1}$ and $|\hat{g}(\beta)|$ is equal to either zero or $1/2^{k-1}$, therefore the induction hypothesis applies to g_0. It follows that S is an affine subspace of dimension $n-1-(k-1) = n-k$. This completes the proof of Theorem 2.1.1.

Reducing the dimension of the function domain. The proof of the Main Theorem is much more involved than that of Rothschild and van Lint Theorem. In fact, the proof we described above of Theorem 2.1.1 is the first step toward proving the main theorem. The reduction step in the proof of Theorem 2.1.1 can be regarded as reducing the dimension of function domain while keeping all the support of the function. Equivalently, one may view the reduction step as decomposing the original function f as a *tensor product* between a "core-function" g and a "δ-function" h (see Section 2.2 for definition of tensor product of Boolean functions). Namely, $f(x,y) = g(x) \otimes h(y)$, where $h : \mathbb{F}_2^m \to \{0,1\}$ is the δ-function: $h(y) = 1$ if $y = 0^m$ and $h(y) = 0$ for all other vectors. That is, f is "reduced" to a core-function g with dimension $n - m$. To this end, we say a function $f : \mathbb{F}_2^n \to \{0,1\}$ is *reducible* if there exists an invertible linear transformation L such that Lf can be decomposed as the tensor product of a function $g : \mathbb{F}_2^{n-m} \to \{0,1\}$ and a δ-function h over $\mathbb{F}_2^m$ with $m \geq 1$. f is said to be *irreducible* if f is not reducible.[2] Now we are ready to present our Main theorem more precisely.

Theorem 2.1.4 (Main)**.** *Let $k \geq 1$, $n > k$ be two integers, and let $f : \mathbb{F}_2^n \to \{0,1\}$ be a non-constant Boolean function with all its Fourier coefficients taking values in $\{0, \frac{\pm 1}{2^k}, \frac{\pm 2}{2^k}\}$. Then we have the following complete characterization*

[2]To put it differently, a function f defined on $\mathbb{F}_2^n$ is irreducible if and only if the minimum dimension of the affine subspace containing the support of f is n.

- *If $\hat{f}(\mathbf{0}) = \frac{1}{2^k}$, then f is the indicator function of an affine subspace of dimension $n-k$ (Rothschild and van Lint Theorem);*
- *If $\hat{f}(\mathbf{0}) = \frac{1}{2^{k-1}}$ and f is irreducible, then f is either the indicator function of disjoint union of two affine subspaces of dimension $n-k$, or the indicator function of disjoint union of four affine subspaces of dimension $n-k-1$. Moreover, the latter case is only possible when $k=4$.*

Back to our problem, since $\hat{f}(\mathbf{0}) = 1/2^{k-1}$, it is easy to see that whenever there is a non-zero α such that $|\hat{f}(\alpha)| = 1/2^{k-1}$, we can restrict f either to the subspace $\langle \alpha, x \rangle = 0$ or to the affine subspace $\langle \alpha, x \rangle = 1$ while keeping the entire support of f. We repeat this process until we reach a Boolean function f with $\hat{f}(\mathbf{0}) = 1/2^{k-1}$ and all other non-zero Fourier coefficients have magnitude $1/2^k$.

Additive structures of the Fourier spectrum. The starting point of our main argument is the following well-known *characterization* of Boolean functions in terms of their Fourier spectra: a function $f : \mathbb{F}_2^n \to \mathbb{R}$ on the cube is Boolean if and only if

$$\hat{f}(\alpha) = \sum_{\beta \in \mathbb{F}_2^n} \hat{f}(\beta)\hat{f}(\alpha+\beta)$$

holds for every $\alpha \in \mathbb{F}_2^n$. Our main observation is that, since the non-zero Fourier coefficients f can take only two values when f is irreducible, denoting $A := \{\alpha \mid \hat{f}(\alpha) = 1/2^k\}$ and $B := \{\beta \mid \hat{f}(\beta) = -1/2^k\}$, then these two sets — viewed as subsets of abelian group $\mathbb{F}_2^n$ — must exhibit strong *additive structures*. Indeed, one can show that $B + B \subseteq A \cup \{\mathbf{0}\}$ and consequently $|B+B|/|B| \leq (1+|A|)/|B|$.

What can be said about a set B if its *doubling constant* $K := |B+B|/B$ is small? This is a classical problem extensively studied in additive combinatorics. Additive combinatorics is a burgeoning mathematics sub-area which finds exciting applications in theoretical computer science in recent years [ADL18, BSLRZ14, BSRZ15, BDL13, Sam07].

Green and Tao [GT09] proved that, when the underlying ambient group is $\mathbb{F}_2^n$, then B is contained in a subspace of size $2^{2K+O(K \log K)}|B|$, which is asymptotically optimal. Unfortunately, such *asymptotic* "high end" bounds are not accurate enough to be useful for our problem. In fact, we make crucial use of a "low end" additive combinatorics result of Even-Zohar [EZ12], which provides tight bounds on the size of affine span of B in terms of its doubling constant. It is worth noting that all aforementioned applications of additive combinatorics in theoretical computer science employ theorems regarding *asymptotic* behaviors of certain combinatorial objects. We hope researchers may find further applications of such "low end" additive combinatorics results in other places.

2.1.4 Motivations and related work

To the best of our knowledge, besides the work of Rothschild and van Lint, there is no previous structural result on Boolean functions in terms the *magnitudes* of their Fourier coefficients only. Friedgut [Fri98] showed that if the total influence of a Boolean function is small, then it is close to some junta — a function that depends only on a bounded number of variables. Friedgut *et al.* [FKN02] studied Boolean functions whose Fourier mass are concentrated on the lowest two levels and proved that such functions are close to parity functions or negations of parity functions. For a special class of Boolean functions, the so-called *linear threshold functions*, a celebrated result of Chow [Cho61] states that these functions are completely determined by their lowest two level Fourier coefficients; see [DDFS14, OS11] for recent robust versions as well as algorithmic versions of Chow's theorem. Note that all previous structural theorems mentioned above, except Chow's, are "robust" in the following sense: the structural results are robust against small perturbations in the Boolean function's Fourier spectrum. Our main result is automatically robust: by Parseval's identity, small distance in Fourier spectrum implies small distance in func-

tion space; consequently, any Boolean function whose Fourier coefficients are close to being in the form stated in our Main Theorem must also be close to having the affine subspace structures asserted in the theorem.

Apart from studying to what extent can the values of Fourier coefficients themselves determine a Boolean function, an important motivation of this research is to study the behaviors of *Fourier sparse* Boolean functions [GOS$^+$11]. Gopalan *et al.* [GOS$^+$11] proved that, if a Boolean function f has only s non-zero Fourier coefficients, then every Fourier coefficient of f is of the form $m/2^k$, where m is an integer and $k/2 \leq \log s \leq k$. That is, the granularity and Fourier sparsity of a Boolean function are, up to a constant factor, identical. Our result may be regarded as characterizing Boolean functions of Fourier granularity k when all Fourier coefficients of f are between $-2/2^k$ and $2/2^k$.

Probably the most prominent open problem in communication complexity is the so-called *Log-rank Conjecture* proposed by Lovász and Saks [LS88], which asserts that the deterministic communication complexity of any $F : \mathbb{F}_2^n \times \mathbb{F}_2^n \to \{0,1\}$, $\mathsf{D}^{\mathsf{CC}}(F)$, is upper bounded by a polynomial of the logarithm of the rank of the communication matrix $M_F = [F(x,y)]_{x,y}$, where the rank is taken over the reals. Even after more than 30 years of extensive study, we are still very far from resolving it; the current best bound is Lovett's $\mathsf{D}^{\mathsf{CC}}(F) = O(\sqrt{r}\log r)$ [Lov14], where r is the rank of M_F. Recently, studying the Log-rank conjecture for a special class of two-party functions, the so-called *XOR functions*, has attracted much attention [CP18, HHL18, LZ17, STlV17, TXZ16, TWXZ13, ZS10]. The corresponding conjecture for this special class of functions is sometimes called *Log-rank XOR conjecture*. Specifically, F is an XOR function if there exists an $f : \mathbb{F}_2^n \to \{0,1\}$ such that for all x and y, $F(x,y) = f(x+y)$. The beautiful connection between the Log-rank XOR conjecture and Fourier analysis of Boolean functions is that, if F is an XOR function, then the rank of M_F is just the Fourier sparsity of f [BC99]. Moreover, it is now known that resolving the Log-rank XOR conjecture is equivalent to finding a

parity decision tree of depth $\mathrm{polylog}(s)$, or $\mathrm{poly}(k)$ for any Boolean function f [HHL18, TWXZ13, ZS10], where s is the Fourier sparsity and k is the granularity of f.

The parity kill number of a Boolean function f is defined as

$$C_{\oplus,\min}(f) := \min\{\text{co-dim}(S) \mid S \text{ is an affine subspace on which } f \text{ is constant}\}$$

Tsang *et al.* [TWXZ13] demonstrated that, to resolve the Log-rank XOR conjecture, it is sufficient to prove that the kill number of any Boolean function f is upper bounded by $\mathrm{polylog}(s)$ or $\mathrm{poly}(k)$. See [CMS19, OST+14] for recent developments on constructing Boolean functions with large kill numbers. Our main result can be regarded as showing that any Boolean function with granularity k and $\hat{f}(\mathbf{0}) \leq 2/2^k$ has kill number at most $k+1$. In fact, by induction on m and folding $\hat{f}(\mathbf{0})$ with any other non-zero Fourier coefficient, we immediately have the following corollary.

Corollary 2.1.5. *Let $f : \mathbb{F}_2^n \to \{0,1\}$ be a Boolean function with granularity k and $\hat{f}(\mathbf{0}) = m/2^k$. Then the kill number of f is at most $k+m-1$.*

Of course, Corollary 2.1.5 is still very far from showing the desired kill number bound $\mathrm{poly}(k)$ as m can be as large as 2^{k-1}, but it is hoped that further investigations along this approach may lead to more interesting results.

2.1.5 Organization

The rest of the chapter is organized as follows. Preliminaries and notations that we use throughout the chapter are summarized in Section 2.2. We prove our Main Lemma, which deals with the cases when k is at least 5 in Section 2.3, while the small value cases are discussed in Section 2.4. Then, by combining these two ingredients, we prove our Main Theorem in Section 2.5. Finally we end with a brief section of conclusions and open questions.

2.2 Preliminaries

All logarithms in this chapter are to the base 2. Let $n \geq 1$ be a natural number, then $[n]$ denotes the set $\{1, \ldots, n\}$. We use $\mathbb{F}_2$ for the field with 2 elements $\{0, 1\}$, where addition and multiplication are performed modulo 2. We view elements in $\mathbb{F}_2^n$ as n-bit binary strings, i.e. elements in $\{0, 1\}^n$, interchangeably. If x and y are two n-bit strings, then $x + y$ (or $x - y$) denotes bitwise addition (i.e. XOR) of x and y. For positive integers m and n, if $y \in \mathbb{F}_2^m$ and $z \in \mathbb{F}_2^n$, then we write $x = (y, z)$ to denote the binary string $x \in \mathbb{F}_2^{m+n}$ obtained from concatenating y and z together. We view $\mathbb{F}_2^n$ as a vector space equipped with an inner product $\langle x, y \rangle$, which we take to be the standard dot product: $\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i$, where all operations are performed in $\mathbb{F}_2$.

2.2.1 Boolean functions and Fourier analysis

We often use f to denote a real function defined on $\mathbb{F}_2^n$ and write $\mathrm{supp}(f) = \{x \in \mathbb{F}_2^n \mid f(x) \neq 0\}$ for the *support* of f. Sometimes we view f as a 2^n-dimensional vector, e.g. write $f = \mathbf{0}$ and $f = \mathbf{1}$ to denote the trivial all-zero function and all-one function, respectively. In this paper, a function f is *Boolean* if its range is $\{0, 1\}$.

For every $\alpha \in \mathbb{F}_2^n$, one can define a *linear function* (or *parity function*) mapping $\mathbb{F}_2^n$ to $\{0, 1\}$ as $\ell_\alpha(x) = \langle \alpha, x \rangle$. Let $\chi_\alpha = (-1)^{\ell_\alpha}$, which are commonly known as *characters*. For functions $f, g \colon \mathbb{F}_2^n \to \mathbb{R}$ the inner product is defined as $\langle f, g \rangle := \mathbb{E}_{x \in \mathbb{F}_2^n}(f(x)g(x))$. For $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{F}_2^n$, the corresponding character function χ_α is defined as $\chi_\alpha(x_1, \ldots, x_n) = \prod_{i \colon \alpha_i = 1} (-1)^{x_i} = (-1)^{\langle \alpha, x \rangle}$. For $\alpha, \beta \in \mathbb{F}_2^n$, the inner product between χ_α and χ_β is 1 if $\alpha = \beta$, and 0 otherwise. Therefore the characters form an orthonormal basis for real-valued functions over $\mathbb{F}_2^n$, and we can expand any f defined on $\mathbb{F}_2^n$ using $\{\chi_\alpha\}_{\alpha \in \mathbb{F}_2^n}$ as a basis.

Definition 2.2.1 (Fourier Transform)**.** *Let* $f\colon \mathbb{F}_2^n \to \mathbb{R}$*. The* Fourier transform $\hat{f}\colon \mathbb{F}_2^n \to \mathbb{C}$ *of* f *is defined to be* $\hat{f}(\alpha) = \mathbb{E}_x(f(x)\chi_\alpha(x))$*. The quantity* $\hat{f}(\alpha)$ *is called the* Fourier coefficient *of* f *at* α*.*

The Fourier inversion formula is given by $f(x) = \sum_{\alpha\in\mathbb{F}_2^n} \hat{f}(\alpha)\chi_\alpha(x)$, and the Parseval's identity is $\sum_{\alpha\in\mathbb{F}_2^n} \hat{f}(\alpha)^2 = \mathbb{E}_x(f(x)^2)$. The Fourier sparsity of f, denoted by $\|\hat{f}\|_0$ or $\mathrm{spar}(f)$, is the number of nonzero Fourier coefficients of f.

Fourier characterization of Boolean functions

Our proof crucially relies on the following characterization of Boolean functions in terms of their Fourier spectra. We give a proof for completeness.

Proposition 2.2.2 (Folklore)**.** *A function* $f : \mathbb{F}_2^n \to \mathbb{R}$ *defined on the hypercube is Boolean if and only if for every* $\alpha \in \mathbb{F}_2^n$*,*

$$\hat{f}(\alpha) = \sum_{\beta\in\mathbb{F}_2^n} \hat{f}(\beta)\hat{f}(\alpha+\beta). \tag{2.1}$$

Proof. This follows from the fact that f is Boolean if and only if $f^2(x) - f(x) = 0$ for every x. Now expand the left-hand side in terms of Fourier coefficients and notice that, since the right-hand side is the $\mathbf{0}$-function, all of its Fourier coefficients all zero. Comparing each pair of the corresponding Fourier coefficients on both sides gives the desired equality. □

Linear restrictions

The following is a folklore theorem regarding the effect of linear restrictions on the Fourier spectrum of a function defined over the Boolean hypercube.

Proposition 2.2.3. *Let* $f : \mathbb{F}_2^n \to \mathbb{R}$ *be a function defined on the Boolean hypercube. Let* $f_0, f_1 : \mathbb{F}_2^{n-1} \to \mathbb{R}$ *be the "sub-functions" obtained from restricting the first bit of*

the input to 0 *and* 1*, respectively; that is,* $f_0(y) := f(0,y)$ *and* $f_1(y) := f(1,y)$ *for all* $y \in \mathbb{F}_2^{n-1}$*. Then the Fourier spectra of* f_0 *and* f_1 *satisfy that, for all* $\beta \in \mathbb{F}_2^{n-1}$*,*

$$\hat{f}_0(\beta) = \hat{f}(0,\beta) + \hat{f}(1,\beta), \qquad \hat{f}_1(\beta) = \hat{f}(0,\beta) - \hat{f}(1,\beta). \tag{2.2}$$

Conversely, the Fourier spectrum of f *satisfies*

$$\hat{f}(0,\beta) = \frac{1}{2}(\hat{f}_0(\beta) + \hat{f}_1(\beta)), \qquad \hat{f}(1,\beta) = \frac{1}{2}(\hat{f}_0(\beta) - \hat{f}_1(\beta)). \tag{2.3}$$

Proof. We prove the first part in (2.3), the second part follows analogously. By the definition of Fourier transform,

$$\begin{aligned}
\hat{f}(0,\beta) &= \frac{1}{2^n} \sum_{x \in \mathbb{F}_2^n} f(x)\chi_{(0,\beta)}(x) \\
&= \frac{1}{2^n} \sum_{y \in \mathbb{F}_2^{n-1}} \big(f(0,y)\chi_{(0,\beta)}((0,y)) + f(1,y)\chi_{(0,\beta)}((1,y))\big) \\
&= \frac{1}{2^n} \left(\sum_{y \in \mathbb{F}_2^{n-1}} f(0,y)\chi_\beta(y) + \sum_{y \in \mathbb{F}_2^{n-1}} f(1,y)\chi_\beta(y) \right) \\
&= \frac{1}{2^n} \sum_{y \in \mathbb{F}_2^{n-1}} f_0(y)\chi_\beta(y) + \frac{1}{2^n} \sum_{y \in \mathbb{F}_2^{n-1}} f_1(y)\chi_\beta(y) \\
&= \frac{1}{2}(\hat{f}_0(\beta) + \hat{f}_1(\beta)).
\end{aligned}$$

□

Tensor product

The statement as well as the proof of Main Theorem requires the standard notion of tensor products between functions.

Definition 2.2.4 (Tensor Product of Boolean Functions)**.** *Let* $f : \mathbb{F}_2^{n_1} \to \{0,1\}$ *and* $g : \mathbb{F}_2^{n_2} \to \{0,1\}$ *be two Boolean functions on* n_1 *and* n_2 *variables respectively. Then the tensor product of* f *and* g*, denoted by* $f \otimes g$*, is a Boolean function over* $\mathbb{F}_2^{n_1+n_2}$ *such that* $f \otimes g(x,y) = f(x) \cdot g(y)$ *for all* $x \in \mathbb{F}_2^{n_1}$ *and* $y \in \mathbb{F}_2^{n_2}$*.*

It is easy to verify the following fact.

Fact 2.2.5. *If $h = f \otimes g$ is the tensor product of two Boolean function defined above, then the Fourier spectrum h satisfies that $\hat{h}(\alpha, \beta) = \hat{f}(\alpha) \cdot \hat{g}(\beta)$, for every $\alpha \in \mathbb{F}_2^{n_1}$ and $\beta \in \mathbb{F}_2^{n_2}$.*

Given a Boolean function $f : \mathbb{F}_2^{n_1} \to \{0, 1\}$, two commonly used functions to tensor with f are the all-one function $g_1 = \mathbf{1}$ whose Fourier spectrum is $\hat{g}_1(\mathbf{0}) = 1$ and $\hat{g}_1(\alpha) = 0$ for any $\alpha \neq \mathbf{0}$; and the "δ-function" g_2 defined by $g_2(x) = 1$ if and only if $x = 0^{n_2}$, whose Fourier spectrum is $\hat{g}_2(\alpha) = 1/2^{n_2}$ for every α. Note that tensoring f with g_1 is equivalent to setting each to the 2^{n_2} sub-functions, defined by restricting y to different values in $\mathbb{F}_2^{n_2}$, to f; and tensoring f with g_1 is to set the sub-function with $y = \mathbf{0}$ to f and set all other sub-functions to the all-zero function.

Invertible linear transformations and linear shifts

Let $L : \mathbb{F}_2^n \to \mathbb{F}_2^n$ be an invertible linear transformation. If $f : \mathbb{F}_2^n \to \{0, 1\}$ is a Boolean function, then define $g := Lf$, the function obtained from applying the linear transformation L to f, as $g(x) = f(Lx)$ for all $x \in \mathbb{F}_2^n$. The Fourier spectrum of g is given by $\hat{g}(\alpha) = \hat{f}((L^T)^{-1}\alpha)$, where L^T stands for the transpose of L viewed as an $n \times n$ matrix. One can check that the set of Fourier coefficients as well as the property of being the indicator function of an (affine) linear subspace are invariant under invertible linear transformations. If $a \in \mathbb{F}_2^n$ is a non-zero vector, and let $h(x) := f(x + a)$ be the linear shift of f, then the Fourier spectrum of h is given by $\hat{h}(\alpha) = \chi_a(\alpha)\hat{f}(\alpha)$ for every $\alpha \in \mathbb{F}_2^n$.

2.2.2 Additive combinatorics

Additive combinatorics is the sub-field of mathematics concerned with subsets of integers or more generally abelian groups, and studies the interplay between the structural prop-

erties of a subset and its combinatorial estimates associated with arithmetic operations. Recently additive combinatorics has found many applications in computer science, see the excellent exposition [Lov17] and the textbook [TV06] for comprehensive treatments.

Throughout this chapter, G is the abelian group $\mathbb{F}_2^n$ for some positive integer n and the underlying field is $\mathbb{F}_2$. If $A = \{a_1, \ldots, a_m\} \subset G$, then $\mathrm{span}(A)$ stands for the *linear span* of A: $\mathrm{span}(A) = \{\sum_{i \in S} a_i \mid S \subseteq [m]\}$, where summation over the empty set is understood to be the 0 element by convention. For any $x \in G$ and $A \subset G$, we write $x + A$ to denote the set $\{x + a \mid a \in A\}$. If A and B are two subsets of G, then $A + B$ denotes the *sumset* $\{a + b \mid a \in A \text{ and } b \in B\}$. Similarly, $A - B := \{a - b \mid a \in A \text{ and } b \in B\}$, although $A - B$ is always the same as $A + B$ in this chapter as the underlying ambient group is $\mathbb{F}_2^n$. If $A = B$ then we write $2A := A + A$ and in general write $kA := \underbrace{A + \cdots + A}_{k \text{ times}}$ for integer $k \geq 1$.

The following Lemma of Laba is useful for our proofs.

Lemma 2.2.6 ([Łab01], Theorem 2.5)**.** *Let G be an abelian group and $A \subset G$ be a subset of G such that $|A - A| < \frac{3}{2}|A|$. Then $A - A$ is a subgroup of G.*

2.3 Proof of the Main Lemma

First recall our Main Lemma states the following.

Lemma 2.1.3. *Let $k \geq 5$ and $n \geq k$ be integers. Let $f : \mathbb{F}_2^n \to \{0, 1\}$ be a Boolean function such that $\hat{f}(\mathbf{0}) = 1/2^{k-1}$ and any other Fourier coefficients are either zero or equal to $\pm\frac{1}{2^k}$, then f is the indicator function of a disjoint union of two dimension $n - k$ affine subspaces.*

In Section 2.6, we compute the Fourier spectrum of a Boolean function that is supported on two disjoint affine subspaces such that the two affine subspaces are of the same

dimension and their Fourier spectra have minimum intersection. Our strategy for the proof of the Main Lemma is to show that if the Fourier coefficients of a Boolean function satisfy the condition prescribed in the Main Lemma, then its Fourier spectrum matches the one we show in Section 2.6.

Let us define

$$A = \{\alpha \in \mathbb{F}_2^n \mid \hat{f}(\alpha) = \frac{1}{2^k}\}$$

and

$$B = \{\beta \in \mathbb{F}_2^n \mid \hat{f}(\beta) = -\frac{1}{2^k}\}.$$

Without loss of generality[3], from now on, we may assume $f(\mathbf{0}) = 1$. We begin with calculating the cardinalities of sets A and B.

Claim 2.3.1. *For any $k \geq 1$ and $n \geq k$, we have $|A| = 3t$ and $|B| = t$, where $t = 2^{k-1} - 1$.*

Proof. Since $\hat{f}(\mathbf{0}) = 1/2^{k-1}$, by Parseval's identity $\hat{f}(\mathbf{0}) = 1/2^{k-1} = \sum_{\alpha \in \mathbb{F}_2^n} \hat{f}^2(\alpha)$, we have $|A| + |B| = 2^{k+1} - 4$.

On the other hand,

$$1 = f(\mathbf{0}) = \sum_{\alpha \in \mathbb{F}_2^n} \hat{f}(\alpha)\chi_\alpha(\mathbf{0}) = \frac{1}{2^{k-1}} + \sum_{\alpha \in A} \frac{1}{2^k} + \sum_{\beta \in B} (-\frac{1}{2^k}),$$

which gives $|A| - |B| = 2^k - 2$. Therefore we have $|A| = 3(2^{k-1} - 1)$ and $|B| = 2^{k-1} - 1$.

□

For convenience, we let $A = \{\alpha_1, \ldots, \alpha_{3t}\}$ and $B = \{\beta_1, \ldots, \beta_t\}$ in the following.

[3]This is because if $f(\mathbf{0}) = 0$, then let $a \in \mathbb{F}_2^n$ be any vector such that $f(a) = 1$. We can apply a linear shift a to f to get a new Boolean function, $h(x) = f(x + a)$ for every x, so that $h(\mathbf{0}) = 1$. Note that the conclusions in our Main Theorem are invariant under linear shifts. Moreover, since $\hat{h}(\alpha) = \chi_a(\alpha)\hat{f}(\alpha)$ for every $\alpha \in \mathbb{F}_2^n$, we have $\hat{h}(\mathbf{0}) = 1/2^{k-1}$ and $|\hat{h}(\alpha)| = |\hat{f}(\alpha)|$ for any other nonzero α. Therefore, the assumptions apply to h as well.

2.3.1 Some additive properties of sets A and B

We now study the additive properties of sets A and B. Note that the Fourier coefficients of f are non-zero only at $\mathbf{0}$ and in sets A and B; moreover, the Fourier coefficients are uniform for points in A or B. Therefore, by Proposition 2.2.2, we expect that there are nice additive structures within A and B.

Definition 2.3.2. *We call* $(\alpha, \beta, \alpha+\beta)$ *a* triangle *if* α, β *and* $\alpha+\beta$ *are all in the support of* $\hat{f}$*; that is* $\alpha, \beta, \alpha+\beta \in A \cup B \cup \{\mathbf{0}\}$.

Lemma 2.3.3. *For any* $\beta_i \in B$*, there are exactly* t *triangles passing through* β_i*; namely, the* t *triangles are* $(\beta_i, \beta_i, \mathbf{0})$ *and* $\{(\beta_i, \beta_j, \beta_i+\beta_j)\}_{j=1, j\neq i}^{t}$*. In the language of set addition, we have* $2B \subseteq A \cup \{\mathbf{0}\}$.

Proof. For any $\beta_i \in B$, by Proposition 2.2.2,

$$\begin{aligned}
\hat{f}(\beta_i) = -\frac{1}{2^k} &= \sum_{\gamma \in \mathbb{F}_2^n} \hat{f}(\gamma)\hat{f}(\beta_i+\gamma) \\
&= 2\hat{f}(\mathbf{0})\hat{f}(\beta_i) + \sum_{\substack{j=1 \\ j\neq i}}^{t} \hat{f}(\beta_j)\hat{f}(\beta_i+\beta_j) + \sum_{\ell=1}^{3t} \hat{f}(\alpha_\ell)\hat{f}(\beta_i+\alpha_\ell) \\
&\geq 2 \cdot \frac{1}{2^{k-1}} \cdot (-\frac{1}{2^k}) + 2(t-1)(-\frac{1}{2^k})(\frac{1}{2^k}) \text{ [4]} \\
&= -\frac{1}{2^k},
\end{aligned}$$

where the inequality in the second last line becomes equality if and only if the following two conditions hold: 1) for every $1 \leq j \leq t$, $j \neq i$, $\beta_i + \beta_j \in A$; and 2) there is no triangle of the form $(\beta_i, \alpha_j, \alpha_\ell)$. Hence the lemma follows. □

Corollary 2.3.4. *The set* B *is a sum-free set; namely, for any three elements* $\beta_1, \beta_2, \beta_3 \in B$*,* $\beta_1 + \beta_2 \neq \beta_3$*. Equivalently,* $2B \cap B = \emptyset$.

[4] There is a factor 2 in the second summation because if $\beta_i + \beta_j \in A$, then the triangle $(\beta_i, \beta_j, \beta_i+\beta_j)$ appears twice in the summation $\sum_{\gamma \in \mathbb{F}_2^n} \hat{f}(\gamma)\hat{f}(\beta_i+\gamma)$: once with $\gamma = \beta_j$ and the other with $\gamma = \beta_i + \beta_j$.

Proof. This follows directly from Lemma 2.3.3 and the fact sets A and B are disjoint. □

Corollary 2.3.5. *We have* $2B \cap 3B = \emptyset$.

Proof. Suppose not, then there exist $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ in B such that $\beta_1+\beta_2 = \beta_3+\beta_4+\beta_5$. These five elements must be distinct as otherwise they would give rise to a triangle in B. But then we have a $(\alpha_1, \alpha_2, \beta_5)$ triangle, where $\alpha_1 := \beta_1 + \beta_2$ and $\alpha_2 := \beta_3 + \beta_4$, contradicting to Lemma 2.3.3. □

Let us define

$$R = 2B \cap A = 2B \setminus \{\mathbf{0}\}$$

and

$$L = A \setminus R.$$

Note that L and R are disjoint and $A = L \cup R$. For any $\rho \in R$, let

$$N(\rho) = \{\beta_i \in B \mid \exists \beta_j \in B \text{ s.t. } \rho = \beta_i + \beta_j \}$$

be the set of points in B which has a triangle passing through ρ. Define a set $\Gamma \subset \mathbb{F}_2^n$ as

$$\Gamma = \{\gamma = \rho + \beta \mid \rho \in R, \beta \in B \text{ and } \beta \notin N(\rho)\}.$$

Observe that Γ is nonempty: since for every $\rho \in R$, all its β-neighbors can be paired together, so $|N(\rho)|$ is an even number, but $|B| = 2^{k-1} - 1$ is odd.

Claim 2.3.6. *We have* $\Gamma = 3B \setminus B$.

Proof. On one hand, by the definition of set Γ, $\Gamma \subseteq 3B$; since R and B are disjoint and $\mathbf{0} \notin R$, we have $\Gamma \cap B = \emptyset$, and hence $\Gamma \subseteq 3B \setminus B$. On the other hand, let γ be any element in $3B$; that is $\gamma = \beta_1 + \beta_2 + \beta_3$, where $\beta_1, \beta_2, \beta_3 \in B$. When will γ

actually be in B? This happens only if any two of these three elements are identical, then $\gamma = \beta_i$ for some $i \in \{1, 2, 3\}$, thus $\gamma \in B$. Moreover, assume that these three elements are distinct and suppose $\gamma \in B$, i.e. $\gamma = \beta_j$ for some $j > 3$. Let $\rho := \beta_1 + \beta_2$, then $\rho = \beta_3 + \gamma = \beta_3 + \beta_j$; that is $\gamma = \rho + \beta_3$ and $\beta_3 \in N(\rho)$. Therefore, if $\gamma \in 3B \setminus B$, then we must have $\beta_3 \notin N(\rho)$ and consequently $\gamma \in \Gamma$. It follows that $3B \setminus B \subseteq \Gamma$. This completes the proof of the claim. $\square$

It is easy to see that Γ is disjoint from the Fourier support of f.

Claim 2.3.7. *For every element $\gamma \in \Gamma$, we have $\hat{f}(\gamma) = 0$.*

Proof. Recall that, the support of $\hat{f}$ is $A \cup B \cup \{\mathbf{0}\}$. Suppose $\hat{f}(\gamma) \neq 0$, that is $\gamma \in \mathrm{supp}(\hat{f})$. Since $A = L \cup R$, from Claim 2.3.6, we know that $\Gamma \cap B = \emptyset$; from Claim 2.3.6 and Corollary 2.3.5, we know that $\Gamma \cap 2B = \Gamma \cap (R \cup \{\mathbf{0}\}) = \emptyset$. So there is only one possibility left, which is $\gamma \in L$. However, if this were the case, because $\gamma = \rho + \beta$ with $\rho \in R$, it would give rise to a (γ, ρ, β)-triangle with $\gamma, \rho \in A$, contradicting Lemma 2.3.3, so γ is not in L, hence $\hat{f}(\gamma) = 0$. $\square$

2.3.2 Even-Zohar's tight bound on $F(K)$

Let G be an abelian group and $A \subset G$ be a subset. The fundamental Freiman theorem [Fre73] in additive combinatorics states that if G is $\mathbb{Z}$ and $|A + A| \leq K|A|$ for some constant K, then there exist functions $d(K)$ and $\ell(K)$ such that A is contained in a $d(K)$-dimensional arithmetic progression of length at most $\ell(K)|A|$. The ratio $\sigma[A] := |A + A|/|A|$ is commonly known as the *doubling constant* of set A. Hence Freiman theorem asserts that if a set of integers has small doubling constant, then the set is well-structured. Ruzsa [Ruz99] established an analog of Freiman's theorem for finite abelian groups with torsion r. Specifically, he proved that any subset A with doubling

constant K is contained in a subgroup of G of size at most $K^2 r^{K^4}|A|$. The question for groups $\mathbb{F}_2^n$ was first studied by Green and Ruzsa [GR06] and the bound was later improved by Sanders [San08]. An asymptotically tight bound was first proved in [GT09] and [Kon08].

For a subset $A \subset \mathbb{F}_2^n$, let $\langle A \rangle$ denote the *affine span* of A; namely, the smallest affine subspace that contains A. If $\sigma[A] = K$, then let $F(K) := \max_{A:\sigma[A]=K} |\langle A \rangle|/|A|$ denote the maximum relative size of the affine span of A. Even-Zohar [EZ12] gave the tight bound of $F(K)$ for all values of doubling constant K.

Theorem 2.3.8 ([EZ12], Theorem 2)**.** *Let A be a subset of $\mathbb{F}_2^n$ with doubling constant K, i.e. $|2A|/|A| \leq K$. If s is the unique positive integer satisfying the inequalities*

$$\frac{\binom{s}{2} + s + 1}{s+1} \leq K < \frac{\binom{s+1}{2} + s + 2}{s+2}, \tag{2.4}$$

then $|\langle A \rangle|/|A| \leq F(K)$, where $F(K)$ is given by

$$F(K) = \begin{cases} \frac{2^s}{\binom{s}{2}+s+1} \cdot K & \text{if } \frac{\binom{s}{2}+s+1}{s+1} \leq K < \frac{s^2+s+1}{2s}, \\ \frac{2^{s+1}}{s^2+s+1} \cdot K & \text{if } \frac{s^2+s+1}{2s} \leq K < \frac{\binom{s+1}{2}+s+2}{s+2}. \end{cases} \tag{2.5}$$

2.3.3 Characterizing $2B$ and $\mathrm{span}(B)$

Note that the doubling constant of set B satisfies that

$$\sigma[B] = \frac{|R|+1}{|B|} \leq \frac{|A|+1}{|B|} = 3 + \frac{1}{t}, \tag{2.6}$$

and recall that $t = 2^{k-1} - 1$. Therefore, when $k \geq 5$, $K = \sigma[B] \leq \frac{46}{15}$. Plugging this K into (2.4) gives that $s \leq 5$ and consequently $F(K) \leq 2K < 7$. That is, we have $|\langle B \rangle| < 7|B|$.

The most important step in our proof is establishing the following lemma, which almost completely characterizes the structure of set B.

Lemma 2.3.9. *If $k \geq 5$, then $|\operatorname{span}(B)| = 2^k = 2(|B|+1)$ and $2B$ is a subspace of dimension $k-1$.*

We prove Lemma 2.3.9 in the following two subsections, distinguishing between the case when $\langle B \rangle$ is an affine subspace and the case when $\langle B \rangle$ is a subspace.

If $\langle B \rangle$ is an affine subspace

In the case that $\langle B \rangle$ is an affine subspace, let $\langle B \rangle = a + H$ be the affine subspace, where H is a subspace of $\mathbb{F}_2^n$, $a \in H^\perp$ and $a \neq \mathbf{0}$. Therefore $\operatorname{span}(B) = H \cup (a+H)$. Note that we now have $2\ell B \subseteq H$ and $(2\ell-1)B \subseteq a+H$ for every integer $\ell \geq 1$. Moreover, $|\operatorname{span}(B)| = 2|\langle B \rangle| < 14|B|$. Since $\operatorname{span}(B)$ is a subspace and $|B| = 2^{k-1}-1$, so there are only three possibilities: $|\operatorname{span}(B)| = 8(|B|+1)$, $|\operatorname{span}(B)| = 4(|B|+1)$ and $|\operatorname{span}(B)| = 2(|B|+1)$. In the following, we are going to eliminate the first two possibilities.

Claim 2.3.10. *Set L is nonempty.*

Proof. Suppose not, then $2B = A \cup \{\mathbf{0}\} \subset H$. Recall that by Claim 2.3.6, $\Gamma = 3B \setminus B$, so $\Gamma \subseteq a+H$ and is disjoint from set A. It follows that for any $\gamma \in \Gamma$, $\hat{f}(\gamma) = 0$ (or directly from Claim 2.3.7). However, applying Proposition 2.2.2 to $\hat{f}(\gamma)$, we see that by the definition of set Γ, $\gamma = \rho + \beta$ with $\rho \in A$, $\beta \in B$ and $\beta \notin N(\rho)$. Hence there is at least one negative term contribution on the right-hand side in (2.1) for $\hat{f}(\gamma)$, but since both $2B$ and $2A$ are disjoint from Γ, there is no positive term on the right-hand side in (2.1), a contradiction. □

We discuss the following two possibilities separately.

The case when $|H| = 4(|B| + 1)$. First note that if this were the case, then $F(K) = |\langle B\rangle|/|B| = 4(1 + \frac{1}{|B|})$. By Theorem 2.3.8, the doubling constant of B is at least $K = |2B|/|B| > 2.5$, or $|2B| > 2.5|B|$. Therefore $|L| \leq 0.5|B|$. On the other hand, $4B = 2B + 2B$ and $4B \subseteq H$ so $\sigma[2B] = |4B|/|2B| < \frac{4+\frac{1}{16}}{2.5} < 7/4$. Then by Theorem 2.3.8 again, $|4B| = |\langle 2B\rangle|$, that is $4B = H$.

We next claim that $L \subseteq H$. To see this, let λ be an arbitrary element in L; applying Proposition 2.2.2 to $\hat{f}(\lambda)$ gives

$$\frac{1}{2^k} = \hat{f}(\lambda) = 2\hat{f}(\lambda)\hat{f}(\mathbf{0}) + \sum_{\lambda' \in L} \hat{f}(\lambda')\hat{f}(\lambda + \lambda') + \text{other terms}.$$

The first term and the second summation can contribute at most $\frac{1}{2^{2k}}(2|L| + 2) \leq \frac{|B|+2}{2^{2k}} < \frac{1}{2^k}$. Therefore, the "other terms" on the right-hand side must contain terms of the form $\hat{f}(\alpha_1)\hat{f}(\alpha_2)$, where α_1 and α_2 are two distinct points in A and $\lambda = \alpha_1 + \alpha_2$. That is $\lambda \in 2B + 2B$, hence it follows that $L \subseteq 4B = H$.

Let $D := H \setminus (2B \cup L)$. We have $|D| = 4(|B| + 1) - 3|B| - 1 = |B| + 3 > 0$. Let δ be any point in D. First, since $\delta \notin 2B \cup B$, $\hat{f}(\delta) = 0$. Second, since $\delta \in H$, there is no negative term in the right-hand side of $0 = \hat{f}(\delta) = \sum_{\gamma \in \mathbb{F}_2^n} \hat{f}(\gamma)\hat{f}(\delta + \gamma)$, because if $\gamma \in B$, then $\delta + \gamma \in a + H$ but there is no positive Fourier coefficient in $a + H$ (since $L \subset H$). On the other hand, consider the set $\{\delta + \alpha \mid \alpha \in 2B \cup L\}$. Since $|D| < |H|/2$, this set has non-empty intersection with $2B \cup L$. Therefore, there are positive terms in $\sum_{\gamma \in \mathbb{F}_2^n} \hat{f}(\gamma)\hat{f}(\delta + \gamma)$, this contradicts the fact that $\hat{f}(\delta) = 0$.

The case when $|H| = 2(|B| + 1)$. This case is similar to the previous one. First, if this were the case, then $F(K) = |\langle B\rangle|/|B| = 2(1 + \frac{1}{|B|})$. It follows that, by Theorem 2.3.8, the doubling constant of B is at least $K = |2B|/|B| > 7/4$, and hence $|4B|/|2B| \leq |H|/|2B| < 3/2$, and by Theorem 2.3.8 again $4B = H$. The rest is identical to the case when $|H| = 4(|B| + 1)$.

Proof. [Proof of Lemma 2.3.9 when $\langle B\rangle$ is an affine subspace] Now that the only possibility left is $|\operatorname{span}(B)| = 2 \cdot (|B| + 1)$, and because $\langle B\rangle$ is an affine subspace, it follows that $2B \subseteq H$ and hence $|2B| \leq |B| + 1$. Applying Laba's lemma, Lemma 2.2.6, to set B gives that $2B$ is a subspace. Since $|2B| \geq |B|$, it follows that $2B = H$, a dimension $k - 1$ subspace. □

If $\langle B\rangle$ is a subspace

If the affine span $\langle B\rangle$ is a subspace, and since $|\langle B\rangle| < 7|B|$, then we either have $|\langle B\rangle| = 4(|B| + 1)$ or $|\langle B\rangle| = 2(|B| + 1)$ (because $B \cap 2B = \emptyset$ and $|2B| \geq |B|$, $|\langle B\rangle| \geq 2|B|$). In the following we exclude the first case.

Recall that $R = 2B \setminus \{\mathbf{0}\}$ is the set of non-zero points in the Fourier support of f that can be written as a sum of two β-points in B. Let $R = \{\lambda_1, \ldots, \lambda_m\}$, where m is the cardinality of R.

Claim 2.3.11. *If $\langle B\rangle$ is a subspace, then $m \leq 2.5t$.*

Proof. For the sake of contradiction, suppose that $m > 2.5t$. For every $\lambda_i \in R$, let d_i be the number of β_j's that form a triangle with λ_i. Then we have $\sum_{i=1}^{m} d_i = t(t-1)$ and $d_i \geq 2$ for every $1 \leq i \leq m$. By a standard averaging argument, there is some λ_i with $d_i \leq 0.4t$. By the definition of set Γ, it follows that $|\Gamma| \geq t - d_i = 0.6t$. Recall that $\Gamma = 3B \setminus B$ so $\Gamma \subset \langle B\rangle = \operatorname{span}(B)$, and Γ is disjoint from either $2B$ or B, thus $|\langle B\rangle| \geq |2B| + |B| + |\Gamma| > 4.1t$, contradicting our assumption that $|\langle B\rangle| = 4(|B| + 1)$. □

Proof. [Proof of Lemma 2.3.9 when $\langle B\rangle$ is a subspace] Now since $m \leq 2.5t$, the doubling constant of B is at most $|2B|/|B| \leq 2.5 + 1/|B| < 21/8$, then by Theorem 2.3.8, $|\langle B\rangle|/|B| < 42/11 < 4$, therefore we must have $|\langle B\rangle| = 2(|B| + 1) = 2^k$. Once again, applying Laba's lemma to set B shows that $2B$ is a subspace of dimension $k - 1$. □

2.3.4 Completing the proof of the Main Lemma

By Lemma 2.3.9, $2B$ is a dimension $k-1$ subspace; without loss of generality, we may assume that

$$H = 2B = \mathrm{span}(e_1, \ldots, e_{k-1}). \tag{2.7}$$

Since $|\mathrm{span}(B)| = 2^k = 2|2B|$, and $B \cap 2B = \emptyset$, B is an affine shift of H with one point δ missing. Since $\delta \notin H$, so without loss of generality, we may assume e_k is the missing point. That is

$$B = (e_k + \mathrm{span}(e_1, \ldots, e_{k-1})) \setminus \{e_k\} \quad \text{and} \tag{2.8}$$

$$R = 2B \setminus \{\mathbf{0}\} = \mathrm{span}(e_1, \ldots, e_{k-1}) \setminus \{\mathbf{0}\} = e_k + B. \tag{2.9}$$

Now by Claim 2.3.6, we have $\Gamma = \{e_k\}$ and consequently $\hat{f}(e_k) = 0$. Our last task is to determine the structure of set L. Recall that $A = R \cup L$ and $|A| = 3t$, and because we now have $R = 2B \setminus \{\mathbf{0}\}$, therefore $|L| = 2t = 2^k - 2$.

Claim 2.3.12. *For any $\lambda \in L$, $e_k + \lambda \in L$.*

Proof. Applying Proposition 2.2.2 to the Fourier coefficient of f at e_k and noting that $R = e_k + B$, we have

$$\begin{aligned}
\hat{f}(e_k) = 0 &= \sum_{\gamma \in \mathbb{F}_2^n} \hat{f}(\gamma)\hat{f}(e_k + \gamma) \\
&= 2\sum_{\rho \in R} \hat{f}(\rho)\hat{f}(e_k + \rho) + \sum_{\lambda \in L} \hat{f}(\lambda)\hat{f}(e_k + \lambda) \\
&\leq 2t \cdot (-\frac{1}{2^{2k}}) + 2t \cdot \frac{1}{2^{2k}} \\
&= 0,
\end{aligned}$$

where equality holds in the second last line only if for every $\lambda \in L$, $\hat{f}(e_k + \lambda) = \frac{1}{2^k}$. That is, $e_k + \lambda \in A (= L \cup R)$. As each element in R has already been taken into account in the first summation in the second line, therefore we necessarily have $e_k + \lambda \in L$. $\square$

Claim 2.3.13. *For any $\lambda \in L$ and $\rho \in R$, $\hat{f}(\lambda + \rho) = 0$.*

Proof. Applying Proposition 2.2.2 to $\hat{f}(\rho)$, where ρ is an arbitrary element in R, we have

$$\begin{aligned}
&\hat{f}(\rho) = \frac{1}{2^k} \\
&= 2 \cdot \hat{f}(\mathbf{0})\hat{f}(\rho) + \sum_{\beta \in B} \hat{f}(\beta)\hat{f}(\rho + \beta) + \sum_{\rho' \in R, \rho' \neq \rho} \hat{f}(\rho')\hat{f}(\rho + \rho') + \sum_{\lambda \in L} \hat{f}(\lambda)\hat{f}(\lambda + \rho) \\
&= 2 \cdot \frac{2}{2^k} \cdot \frac{1}{2^k} + (t-1) \cdot (-\frac{1}{2^k}) \cdot (-\frac{1}{2^k}) + (t-1) \cdot (\frac{1}{2^k}) \cdot (\frac{1}{2^k}) + \sum_{\lambda \in L} \frac{1}{2^k} \cdot \hat{f}(\lambda + \rho) \\
&\geq \frac{1}{2^k}, \qquad \text{(as } \lambda + \rho \notin B \text{, therefore } \hat{f}(\lambda + \rho) \geq 0)
\end{aligned}$$

where we have a factor of $(t-1)$ in the second line because $\rho + e_k \in B$ and equality holds in the last line only if $\hat{f}(\lambda + \rho) = 0$ for every $\lambda \in L$ and every $\rho \in R$. $\square$

Claim 2.3.14. *For any $\lambda, \lambda' \in L$, $\lambda + \lambda' \in L$ except that $\lambda + \lambda' = \mathbf{0}$ or e_k.*

Proof. Applying Proposition 2.2.2 to $\hat{f}(\lambda)$, where λ is an arbitrary element in L, we have

$$\begin{aligned}
&\hat{f}(\lambda) = \frac{1}{2^k} \\
&= 2 \cdot \hat{f}(\mathbf{0})\hat{f}(\lambda) + \sum_{\beta \in B} \hat{f}(\beta)\hat{f}(\lambda + \beta) + \sum_{\rho \in R} \hat{f}(\rho)\hat{f}(\lambda + \rho) + \sum_{\lambda' \in L, \lambda' + \lambda \notin \{\mathbf{0}, e_k\}} \hat{f}(\lambda')\hat{f}(\lambda + \lambda') \\
&= 2 \cdot \frac{2}{2^k} \cdot \frac{1}{2^k} + 0 + 0 + \sum_{\lambda' \in L, \lambda' + \lambda \notin \{\mathbf{0}, e_k\}} \frac{1}{2^k} \cdot \hat{f}(\lambda + \lambda') \text{ [5]} \\
&\leq \frac{4}{2^{2k}} + (2t-2) \cdot (\frac{1}{2^k}) \cdot (\frac{1}{2^k}) \\
&= \frac{1}{2^k},
\end{aligned}$$

where equality holds in the second last line only if $\lambda + \lambda' \in L$ for every $\lambda' \in L$, except when λ' is equal to λ or $\lambda + e_k$. $\square$

[5] The second term vanishes because the only triangles passing through a point $\beta_i \in B$ are of the type $(\beta_i, \beta_j, \rho_\ell)$ where $\rho_\ell \in R$; the third term vanishes because of Claim 2.3.13.

Put Claim 2.3.12, Claim 2.3.13 and Claim 2.3.14 together, and since $|L| = 2^k - 2$ we conclude that $H' := L \cup \{\mathbf{0}, e_k\}$ is a subspace of dimension k. Moreover, as $\mathrm{span}(B) = \mathrm{span}(e_1, \ldots, e_k)$ is a subspace of dimension k, and $L \cap \mathrm{span}(B) = \emptyset$, we thus have $H' \cap \mathrm{span}(B) = \{\mathbf{0}, e_k\}$. Therefore, without loss of generality, we may take $H' = \mathrm{span}(e_k, \ldots, e_{2k-1})$ and consequently finally have

$$L = \mathrm{span}(e_k, \ldots, e_{2k-1}) \setminus \{\mathbf{0}, e_k\}. \tag{2.10}$$

It is straightforward to check[6] that the Fourier spectrum calculated in Section 2.6 for a disjoint union of two dimension $n-k$ affine subspaces is identical to the Fourier spectrum of f, which is completely specified by sets in (2.8), (2.9) and (2.10). Therefore the proof of the Main Lemma is complete.

2.4 Dealing with small values of k

When $k = 2$ or $k = 3$, note that since Claim 2.3.1 holds for every $k \geq 2$, this will enable us to prove the same results as Main Lemma by slightly different arguments. That is, when $k = 2$ or $k = 3$, support of f is also a disjoint union of two dimension $n - k$ affine subspaces. However, when $k = 4$ one can not prove the same characterization as Main Lemma. In fact, there are two possibilities: one is that f is still the indicator function of two disjoint dimension $n - 4$ affine subspaces; the other is that support of f are *four* disjoint $n-5$ affine subspaces. Furthermore, we show that this is the only counterexample to Main Lemma for all k. Now we give the precise statements for small values of k and their proofs.

Lemma 2.4.1. *Let $2 \leq k \leq 4$ and $n \geq k$ be integers. Let $f : \mathbb{F}_2^n \to \{0, 1\}$ be a Boolean function such that $\hat{f}(\mathbf{0}) = 1/2^{k-1}$ and any other Fourier coefficients are either zero or*

[6]The second line in (2.11) corresponds to set B, third line in (2.11) corresponds to set R, and the fourth and fifth lines of (2.11) correspond to set L.

equal to $\pm\frac{1}{2^k}$*. If* $k = 2$ *or* $k = 3$*, then* f *is the indicator function of a disjoint union of two dimension* $n - k$ *affine subspaces; If* $k = 4$*, then* f *is either the indicator function of a disjoint union of two dimension* $n - k$ *affine subspaces, or the indicator function of a disjoint union of four dimension* $n - k - 1$ *affine subspaces.*

2.4.1 Proof of the case $k = 2$

In this case, $|A| = 3$ and $|B| = 1$. For convenience, suppose that $\hat{f}(\mathbf{0}) = \frac{1}{2}$, $\hat{f}(\beta) = -\frac{1}{4}$ and $\hat{f}(\alpha_1) = \hat{f}(\alpha_2) = \hat{f}(\alpha_3) = \frac{1}{4}$, where $\beta, \alpha_1, \alpha_2, \alpha_3$ are four distinct non-zero vectors.

We claim that there exists an α_i, $1 \leq i \leq 3$, such that $\hat{f}(\beta + \alpha_i) = 0$. To see this, suppose $\hat{f}(\beta + \alpha_i) \neq 0$ for every $1 \leq i \leq 3$. Because the four vectors are distinct, $\beta + \alpha_i \neq \mathbf{0}$; furthermore, since $\alpha_i \neq \mathbf{0}$, so $\beta + \alpha_i \neq \beta$. It follows that $\beta + \{\alpha_1, \alpha_2, \alpha_3\} = \{\alpha_1, \alpha_2, \alpha_3\}$; that is, adding β to A permutes the three elements in the set. But now adding these three elements together gives $3\beta_1 + \sum \alpha_i = \sum \alpha_i$, a contradiction since $\beta_1 \neq \mathbf{0}$.

Without loss of generality, assume $\hat{f}(\beta + \alpha_1) = 0$ and denote $\beta + \alpha_1$ by γ. Now applying Proposition 2.2.2 to γ gives:

$$\begin{aligned}
\hat{f}(\gamma) = 0 &= \sum_{\alpha} \hat{f}(\alpha)\hat{f}(\alpha + \gamma) \\
&= 2 \cdot \hat{f}(\beta_1)\hat{f}(\alpha_1) + \hat{f}(\alpha_2)\hat{f}(\alpha_2 + \gamma) + \hat{f}(\alpha_3)\hat{f}(\alpha_3 + \gamma) \\
&= 2 \cdot (-\frac{1}{4}) \cdot \frac{1}{4} + \hat{f}(\alpha_2)\hat{f}(\alpha_2 + \gamma) + \hat{f}(\alpha_3)\hat{f}(\alpha_3 + \gamma) \\
&\leq 0,
\end{aligned}$$

where equality holds in the last line only if $\gamma = \alpha_2 + \alpha_3$ so that

$$\hat{f}(\alpha_2)\hat{f}(\alpha_2 + \gamma) = \hat{f}(\alpha_3)\hat{f}(\alpha_3 + \gamma) = \hat{f}(\alpha_2)\hat{f}(\alpha_3) = \frac{1}{4} \cdot \frac{1}{4}.$$

After taking an invertible linear transformation if necessary, we may take $\alpha_1 = e_1, \beta = e_1 + e_2, \alpha_2 = e_3$ and $\alpha_3 = e_2 + e_3$, then it is easy to verify that this is identical to the Fourier spectrum in (2.11) for the case of $k = 2$.

2.4.2 Proof of the case $k = 3$

In this case, $|A| = 9$ and $|B| = 3$. Denote set B by $\{\beta_1, \beta_2, \beta_3\}$. Then by Corollary 2.3.4, $\beta_1 + \beta_2 + \beta_3 \neq 0$, therefore $R = \{\beta_1 + \beta_2, \beta_1 + \beta_3, \beta_2 + \beta_3\}$, and $\Gamma = \{\beta_1 + \beta_2 + \beta_3\}$. Hence Lemma 2.3.9 is established and the rest of the proof is identical to that of the Main Lemma in Section 2.3.4 for the general $k \geq 5$ case.

2.4.3 Proof of the case $k = 4$

First of all, it is easy to see that when $k = 4$, the indicator function of a disjoint union of 2 affine subspaces of dimension $n - k = n - 4$ is still a Boolean function with desired Fourier spectrum, for every $n \geq 4$. Next we construct another Boolean function, which demonstrates that Main Lemma is no longer valid for $k = 4$.

Construction 2.4.2. *Let $G = \mathbb{F}_2^6$ with $e_1, \cdots, e_6$ as the standard basis and let $A, B \subset G$ be two disjoint subsets given as follows:*

- $B = \{e_i \mid 1 \leq i \leq 6\} \cup \{\sum_{i=1}^{6} e_i\}$;
- $A = \{e_i + e_j \mid 1 \leq i < j \leq 6\} \cup \{\sum_{i \in S} e_i \mid S \subset [6], |S| = 5\}$.

Clearly $A = 2B \setminus \{\mathbf{0}\}$, $|B| = 2^{4-1} - 1 = 7$ and $|A| = \binom{7}{2} = 3|B|$, which satisfy the size requirements for A and B for $k = 4$. To see that sets A and B in Construction 2.4.2 satisfy all the additive properties imposed by Proposition 2.2.2, one can explicitly compute a "core" function $f_{CE} : \mathbb{F}_2^6 \to \mathbb{R}$ with $A \cup B \cup \{\mathbf{0}\}$ being its Fourier support to verify that f is indeed a *Boolean* function and $\mathrm{supp}(f_{CE}) = \{\mathbf{0}\} \cup \{\sum_{i \in S} e_i \mid S \subset [6], |S| = 5\} \cup \{\sum_{i=1}^{6} e_i\}$. That is, f is equal to 1 on vectors of weights 0, 5 and 6, and is equal to 0 on all other vectors. Note that $\mathrm{supp}(f_{CE})$ consists of 8 distinct vectors and is a disjoint union of four affine subspaces of dimension $n - 4 - 1 = 1$ each. Moreover, it can be

checked that $\mathrm{supp}(f_{CE})$ is not the union of any two disjoint affine subspaces of dimension 2.

Our next claim shows that, up to an invertible linear transformation, Construction 2.4.2 is essentially the only counterexample to the Main Lemma.

Claim 2.4.3. *When $k = 4$, either f is the indicator function of a disjoint union of two affine subspaces of dimension $n - k$, or the Fourier spectrum of f is given by Construction 2.4.2 under some invertible linear transformation, and consequently f is the indicator function of a disjoint union of four affine subspaces of dimension $n - k - 1$.*

Proof. When $k = 4$, we have $|B| = 2^{4-1} - 1 = 7$. By inequality (2.6), $\sigma[B] = |2B|/|B| \leq 22/7$. But if $|2B| \leq 21$, then plugging $K = \sigma[B] \leq 3$ into (2.4) gives that $s \leq 5$ and consequently $F(K) \leq 2K < 7$. That is, we would have $|\langle B\rangle| < 7|B| = 49$. Then following the same argument, we would be able to establish Lemma 2.3.9 for the case $k = 4$ as well, i.e. to have $|\operatorname{span}(B)| = 2^k = 2(|B| + 1)$ and $2B$ is a subspace of dimension $k - 1$, thereby recovering the regular configuration of f being the indicator function of two disjoint affine subspaces of dimension $n - k$.

Therefore, from now on, we assume that $|2B| = 22$. On the other hand, $|A| = 3|B| = 21$; combining this with Lemma 2.3.3 (i.e. $2B \subseteq A \cup \{\mathbf{0}\}$), we must have $A = 2B \setminus \{\mathbf{0}\}$. By the upper bound on $|\langle B\rangle|$ given in Theorem 2.3.8, we have $|\langle B\rangle| \leq 2^6 = 64$. But if $|\langle B\rangle| < 64$ (hence $|\langle B\rangle| = 32$ or $|\langle B\rangle| = 16$), then the proof of Lemma 2.3.9 would follow again.

Hence, the counter-example is possible only when the dimension of $\operatorname{span}(B)$ is at least 6. Without loss of generality, we may assume $B = \{e_i \mid 1 \leq i \leq 6\} \cup \{\beta\}$. We will determine vector β next.

If $\beta \notin \operatorname{span}(e_1, \cdots, e_6)$, then without loss of generality, let $\beta = e_7$. Now $A = 2B \setminus \{\mathbf{0}\} = \{e_i + e_j \mid 1 \leq i < j \leq 7\}$. But applying Proposition 2.2.2 to the vector $e_1 + e_2 + e_3$

gives that $\hat{f}(e_1 + e_2 + e_3) = -6/2^{2k}$, contradiction to the fact that $\hat{f}(e_1 + e_2 + e_3) = 0$ because $e_1 + e_2 + e_3 \notin A \cup B$. It follows that $\beta \in \text{span}(e_1, \cdots, e_6)$.

Note that every weight-2 vector $e_i + e_j$, $1 \leq i < j \leq 6$, is in A. On the other hand, since $|A| = \binom{|B|}{2}$, it follows that for every $\alpha_k \in A$, there exist a unique pair $\beta_i, \beta_j \in B$ such that $\beta_i + \beta_j = \alpha_k$. Combining these two facts, we conclude that none of the weight-3 vector of the form $e_i + e_j + e_k$ is in B, for every $1 \leq i < j < k \leq 6$, as it would gives two ways to obtain vectors such as $e_i + e_j$ by adding two vectors from B, thus making $|A| < \binom{|B|}{2}$. By Claim 2.3.5, none of the weight-4 vectors can be in B either, which leaves only the possibilities of weight-5 or weight-6 vector for β.

If β is a weight-5 vector, without loss of generality, we may assume $\beta = \sum_{i=1}^{5} e_i$. Then B would contain vectors of weight-1 and weight-5 only, consequently A would contain vectors of weight-2, weight-4 and weight-6 only. Now applying Proposition 2.2.2 to the vector $e_1 + e_2 + e_3$ yields $\hat{f}(e_1 + e_2 + e_3) < 0$, contradicting to the fact that $\hat{f}(e_1 + e_2 + e_3) = 0$ as $e_1 + e_2 + e_3 \notin A \cup B$. Therefore, we have $\beta = \sum_{i=1}^{6} e_i$, completing the proof of the claim. $\square$

2.5 Proof of the Main Theorem

Clearly, if $\hat{f}(\mathbf{0}) = \frac{1}{2^k}$, then, because $|\hat{f}(\alpha)| \leq \hat{f}(\mathbf{0})$ for every α, all non-zero Fourier coefficients of f have absolute value $\frac{1}{2^k}$. Therefore, Rothschild and van Lint Theorem applies and f is the indicator function of an affine subspace of dimension $n-k$. Therefore, from now on, we assume $\hat{f}(\mathbf{0}) = \frac{1}{2^{k-1}}$.

The first step in our proof of the Main Theorem is to follow a similar procedure employed in the proof of Theorem 2.1.1. That is, whenever possible, we reduce the values of n and k simultaneously. This proceeds as follows. Suppose there exists a non-zero α with $\hat{f}(\alpha) = \frac{1}{2^{k-1}}$ or $-\frac{1}{2^{k-1}}$. Without loss of generality, assume that $\hat{f}(\alpha) = \frac{1}{2^{k-1}}$.

Apply an invertible linear transform L that maps α to e_1 and let $g := Lf$. Now we have $\hat{g}(\mathbf{0}) = \hat{g}(e_1) = \frac{1}{2^{k-1}}$. Apply the restriction on the first bit of the input to get sub-functions g_0 and g_1. Then by (2.2), $\hat{g}_1(\mathbf{0}) = \hat{g}(\mathbf{0}) - \hat{g}(e_1) = 0$, which implies that $g_1 \equiv 0$. This implies that $\mathrm{supp}(f)$ is completely contained in the support of g_0 and moreover, by (2.3), $\hat{g}_0(\beta) = 2\hat{f}(0, \beta)$ for every $\beta \in \mathbb{F}_2^{n-1}$. In other words, g_0 is a Boolean function over $\mathbb{F}_2^{n-1}$ and $|\hat{g}(\beta)|$ is equal to either zero, or $\frac{1}{2^{k-1}}$, or $\frac{1}{2^{k-2}}$. That is, by performing a linear restriction, we reduce both the dimension n and the parameter k by one, so that the Main Theorem holds for Boolean functions over $\mathbb{F}_2^n$ as long as it holds for Boolean functions over $\mathbb{F}_2^{n-1}$.

When we arrive at a point that such a linear restriction is no longer possible; equivalently, f is irreducible, then $\hat{f}(\mathbf{0})$ is the only Fourier coefficient whose absolute value is $\frac{1}{2^{k-1}}$. Therefore, the Main Lemma for $k \geq 5$ or Lemma 2.4.1 for $2 \leq k \leq 4$ applies.

2.6 The Fourier spectrum of disjoint union of two affine subspaces

In this section we calculate the Fourier spectrum of a Boolean function whose support is the union of two disjoint affine subspaces satisfying certain properties. In particular, the two affine subspaces are of the same dimension and their Fourier spectra have minimum intersection.

Let $n \geq 1$ and $0 \leq k < n$ be integers. If V is a linear subspace in $\mathbb{F}_2^n$ of dimension $n - k$ and $a \in V^\perp$, where $V^\perp$ denotes the linear subspace that is the *orthogonal complement* of V, then it is well known that the Fourier spectrum of the indicator function of affine subspace $a + V$ is (see e.g. [O'D14]):

$$\hat{\mathbb{1}}_{a+V}(\alpha) = \begin{cases} \frac{1}{2^k}\chi_\alpha(a) & \text{if } \alpha \in V^\perp, \\ 0 & \text{otherwise.} \end{cases}$$

Let $f : \mathbb{F}_2^n \to \{0,1\}$ be a Boolean function whose support is the union of two disjoint affine subspaces of dimension $n-k$. By a shift of the origin if necessary, we may assume that one of the two affine subspaces is a linear subspace. Therefore $f = \mathbb{1}_{a+V_1} + \mathbb{1}_{V_2}$, where V_1 and V_2 are two linear subspaces of dimension $n-k$ in $\mathbb{F}_2^n$ and $a \in V_1^\perp$. In order for $a+V_1$ and V_2 to be disjoint, a necessary condition is that their orthogonal complement subspaces have non-trivial intersection, $V_1^\perp \cap V_2^\perp \neq \{\mathbf{0}\}$. The special configuration we are interested in is when this intersection is minimal, that is when $|V_1^\perp \cap V_2^\perp| = 2$.

To this end, without loss of generality, we let $V_1^\perp = \mathrm{span}(e_1, \ldots, e_k)$ and $V_2^\perp = \mathrm{span}(e_k, \ldots, e_{2k-1})$ so that $V_1^\perp \cap V_2^\perp = \{\mathbf{0}, e_k\}$. Then we necessarily have[7] $\langle e_k, a\rangle = 1$. Therefore for simplicity (and also without loss of generality) we may take $a = e_k$. Therefore the Fourier spectrum of f is

$$\hat{f}(\alpha) = \hat{\mathbb{1}}_{a+V_1}(\alpha) + \hat{\mathbb{1}}_{V_2}(\alpha) = \begin{cases} \frac{1}{2^{k-1}} & \text{if } \alpha = \mathbf{0}, \\ -\frac{1}{2^k} & \text{if } \alpha \in e_k + (\mathrm{span}(e_1, \ldots, e_{k-1}) \setminus \{\mathbf{0}\}), \\ \frac{1}{2^k} & \text{if } \alpha \in \mathrm{span}(e_1, \ldots, e_{k-1}) \setminus \{\mathbf{0}\}, \\ \frac{1}{2^k} & \text{if } \alpha \in e_k + (\mathrm{span}(e_{k+1}, \ldots, e_{2k-1}) \setminus \{\mathbf{0}\}), \\ \frac{1}{2^k} & \text{if } \alpha \in \mathrm{span}(e_{k+1}, \ldots, e_{2k-1}) \setminus \{\mathbf{0}\}, \\ 0 & \text{otherwise.} \end{cases} \tag{2.11}$$

[7] This is because, the affine subspace $a + V_1$ can be expressed as the solutions to a system of linear equations $a + V_1 = \{x \in \mathbb{F}_2^n \mid \langle x, e_i\rangle = a_i \text{ for every } 1 \le i \le k\}$, where $\{e_1, \ldots, e_k\}$ is an orthonormal basis for $V_1^\perp$, and $\{a_i := \langle e_i, a\rangle\}_{i=1}^k$ are the components under this basis. Now if $|V_1^\perp \cap V_2^\perp| = 2$, and because the intersection of the two orthogonal complement subspaces is a subspace, we may take $V_1^\perp \cap V_2^\perp = \{\mathbf{0}, e_k\}$ for convenience. On the other hand, $V_2 = \{x \in \mathbb{F}_2^n \mid \langle x, e_i\rangle = 0 \text{ for every } k \le i \le 2k-1\}$. $a+V_1$ and V_2 are disjoint if and only if there is no solution to the two systems of linear equations combined together, which is equivalent to the condition that $\langle e_k, a\rangle = 1$.

2.7 Concluding Remarks and Open Problems

In this chapter, we extend a classical result of Rothschild and van Lint to give a complete characterization of Boolean functions whose Fourier coefficients take values only in the set $\{-2/2^k, -1/2^k, 0, 1/2^k, 2/2^k\}$. Our work may be regarded as a first step toward understanding the structures of Boolean functions of granularity k. A major motivation for such studies is to prove a polynomial upper bound on the kill number for any k-granular Boolean function, thus resolving the Log-rank XOR conjecture. Another interesting question is to find other sets of Fourier coefficients which uniquely or almost uniquely determine the structures of their corresponding Boolean functions.

CHAPTER 3

STATISTICS OF SPARSITY OF REAL POLYNOMIAL REPRESENTATION

3.1 Introduction

In the previous chapter, we focused our work on Fourier representation of Boolean functions and the structure of the Fourier support. In this chapter, we want to learn more about the sparsity of Boolean functions in real polynomial form.

There is a long history of investigating the relation between polynomials and complexity bounds, from communication complexity to circuit complexity. In 1969, Minski and Papert [MP88] started to use real polynomial representations to prove computational complexity properties, together with works by Razborov [Raz87] and Smolensky [Smo87]. In 1992, Nisan and Szegedy [NS94] built connections between degrees, decision tree complexity and sensitivities of Boolean functions. Nisan and Szegedy [NS94] and Paturi's [Pat92] work gave new results related to polynomials that approximate function f. The work from Beigel [Bei93] showed the relations between real polynomial representations and Fourier transform representations. More details can be found in the survey paper from Buhrman and de Wolf [BdW02].

Though, function analysis has drawn lots of attention in the last decades, but understanding of arbitrary functions still needs more work. A recent breakthrough from Knop *et al.* [KLMY20], is also based on real polynomial representation, reduced the originally exponential gap for log-rank conjecture of AND-functions to only $\log n$. Their work built a strong connection between sparsity, monotone block sensitivity and AND-decision tree complexity.

That fascinating result motivates us to study more about the sparsity of polynomial representation of Boolean functions. Given any Boolean function $f : \{0,1\}^n \to \{0,1\}$,

it can be computed by unique real multi-linear polynomial, in the form of summation of monomials $f(x) = \sum_{S\in\{0,1\}^n} c_S \prod_{i\in S} x_i$.

Like in the previous chapter, we can still define the sparsity of function f, which we denote as $\mathrm{spar}(f)$, as the number of non-zero coefficients c_S. What can we say about $\mathrm{spar}(f)$ of a random Boolean function f?

3.2 Preliminaries

In this chapter, we will need some other common notations. $\mathbb{R}$ denotes all the real numbers and $\mathbb{Z}$ denotes all the integers.

As in the previous chapter, we can also consider any $S \in \{0,1\}^n$ as a subset of $[n]$, and from now on, we will denote the size of S as the corresponding lower-case letter, $s := |S|$ (similarly $t := |T|$ and $r := |R|$). Moreover we will use $S \setminus T := \{\alpha : \alpha \in S \bigcap \alpha \notin T\}$ to represent set S minus set T.

For any set S, we define $B^S := \{T : T \subset S\}$, the Boolean subcube generated by S, or equivalently the set of all subsets of S.

For any random variable X, we define $\mathbb{E}[X] := \sum_i i \Pr[X = i]$ as the expectation of X, and $\mathrm{Var}[X] := \mathbb{E}[X^2] - \mathbb{E}^2[X]$ as the variance of X. For two random variables X and Y, we define $\mathrm{cov}[X,Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ as the covariance of X and Y.

For an event A, we use $\mathbb{1}_A$ to denote the indicator function of A, which means

$$\mathbb{1}_A := \begin{cases} 1 & \text{A happens,} \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

3.2.1 Real polynomial representation of Boolean function

For any functions $f : \{0,1\}^n \to \mathbb{R}$, we can always have a multivariate polynomial computes f as follow, $f(x_1, x_2, \cdots, x_n) = \sum_{S\in\{0,1\}^n} f(S) \prod_{i\in S} x_i \prod_{i\notin S}(1-x_i)$. Beigel [Bei93]

named this as table lookup representation. For any input S, we can directly go to the coefficient of the corresponding term $\prod_{i\in S} x_i \prod_{i\notin S}(1-x_i)$. We can also view those terms as indicator polynomials.

The 2^n terms form a basis for the vector space of functions $f : \{0,1\}^n \to \mathbb{R}$ as that is a vector space of dimension 2^n. Therefore every functions $f : \{0,1\}^n \to R$ has a unique table lookup representation.

Table lookup representation can be directly transformed to real polynomial form $f(x) = \sum_{S\in\{0,1\}^n} c_S \prod_{i\in S} x_i$ by multiplying out the formula with applying distributive law. But the value of c_S looks unclear. The famous Moebius inversion formula would be the answer. And we will include a self-contained proof just for completeness.

Lemma 3.2.1 ([Bei93, BdW02], Moebius inversion formula)**.** *we have* $f(S) = \sum_{T\subset S} c_T$ *and controversially* $c_S = \sum_{T\subset S}(-1)^{s-t} f(T)$.

Proof. The first part is trivial since $\prod_{i\in T} x_i = 1$ if and only if $T \subset S$.

And we could rewrite $c_S = \sum_{T\subset S} \mu(S,T) f(T)$, and key-point is to calculate $\mu(S,T)$.

Let's use induction on $s-t$ to prove $\mu(S,T) = (-1)^{s-t}$, and we should start from $s-t=0$. Obviously $\mu(S,S) = 1$, since $c_S = f(S) - \sum_{T\subsetneq S} c_T$.

Now suppose for all $s-t \le k-1$, we have $\mu(S,T) = (-1)^{s-t}$, we will show that when $s-t=k$, $\mu(S,T) = (-1)^{s-t} = (-1)^k$ still holds.

Since we have $c_S = f(S) - \sum_{R\subset S, R\neq S} c_R = f(S) - \sum_{R\subsetneq S}\left(\sum_{T\subset R} \mu(R,T) f(T)\right)$, we obtain that $\mu(S,T) = -(\sum_{T\subset R\subsetneq S} \mu(R,T)) = -\sum_{i=0}^{k-1} \binom{k}{i}(-1)^i = (1-1)^n + (-1)^k = (-1)^k$, which is just what we want, and the third equation is by Binomial theorem [GKP89, p.174].

□

Moebius inversion formula is a general method, about relation between a pair of functions, first derived by Moebius in 1832. But any further details would be out of our scope.

Although these representations and formulas works for all functions defined on the Boolean cube, $f : \{0,1\}^n \to \mathbb{R}$. In aspect of computer, we have only Boolean digits, hence, we will restrict our work on Boolean functions $f : \{0,1\}^n \to \{0,1\}$.

3.3 Statistics results

3.3.1 Expectation

First let's start with the expectation. For a random Boolean function $f : \{0,1\}^n \to \{0,1\}$, we have $\mathbb{E}[\mathrm{spar}(f)] = \mathbb{E}[\sum_{S\in\{0,1\}^n} \mathbb{1}_{c_S\neq 0}]$. By the linearity of expectation, we could calculate $\mathbb{E}[\mathbb{1}_{c_S\neq 0}] = \Pr[c_S \neq 0]$ for all set S and sum them together. We will need the Moebius inversion formula which we have previously shown.

And also another useful lemma from personal communication between Yaoyun Shi, Buhrman and de Wolf [BdW02, Lemma 4, Theorem 6] will be needed.

Lemma 3.3.1 (Shi & Yao, [BdW02])**.** *For a random Boolean function* $f : \{0,1\}^n \to \{0,1\}$*, we have* $\mathbb{E}[\mathbb{1}_{c_{[n]}=0}] = \Pr[c_{[n]} = 0] = \frac{\binom{2^n}{2^{n-1}}}{2^{2^n}}$.

We noticed that this lemma can be easily generalized to any c_S as below, and that would be the starting point of our method measuring the sparsity.

Lemma 3.3.2. *For a random Boolean function* $f : \{0,1\}^n \to \{0,1\}$*, we have* $\mathbb{E}[\mathbb{1}_{c_S=0}] = \Pr[c_S = 0] = \frac{\binom{2^s}{2^{s-1}}}{2^{2^s}}$.

Proof. Following the Lemma 3.2.1 we may see that only the value of $f(T)$ is affecting the value of c_S when $T \subset S$.

Restricting our function to the Boolean cube B^S, with size $|B^S| = 2^s$, we have a total of 2^{2^s} choices to assign the values. Also, we can partition the cube to the odd layers, denote it as $B^S_{odd} = \{T : T \subset S \bigcap s - t \text{ is odd}\}$ and the even layers $B^S_{even} = \{T : T \subset$

$S \bigcap s-t$ is even$\}$, as any subset T in them such that $f(T)=1$ will contribute -1 and 1 to c_S respectively.

Hence, for $c_S=0$ to happen, we will have a balanced state, which means equal number of T s.t. $f(T)=1$ in both B^S_{odd} and B^S_{even}. Supposing that number as k and k could go from 0 to 2^{s-1}, we have total $\sum_{k=0}^{2^{s-1}} \binom{2^{s-1}}{k}\binom{2^{s-1}}{k} = \binom{2^s}{2^{s-1}}$ Boolean functions satisfying this property. The last equation is implied by the Vandermonde convolution [GKP89, p.174]. $\square$

As we may call $c_S=0$ as a *balanced state*, $c_S=k$ as k-biased state, then we have the following corollary for k-biased state using Vandermonde convolution again.

Corollary 3.3.3. *For all k such that,* $-2^{s-1} \le k \le 2^{s-1}$*, we have* $\Pr[c_S=k] = \frac{\binom{2^s}{2^{s-1}+k}}{2^{2^s}}$.

By Lemma 3.3.2, we may notice that when n and $s=|S|$ goes to infinity, $\Pr[c_S \neq 0]$ is close to 1. We may want to measure the supplementary part $\Pr[c_S=0]$ and $\mathrm{zero}(f) := \sum_{S\in\{0,1\}^n} \mathbb{1}_{c_S=0}$, the number of zero coefficients c_S, hence we have $\mathrm{zero}(f) = 2^n - \mathrm{spar}(f)$. $\mathrm{zero}(f)$ have similar concentration properties as $\mathrm{spar}(f)$ since it's just flipped horizontally over range $[0,2^n]$. Also, we could consider $c_S=0$ as an event, and denote $X_S = \mathbb{1}_{c_S=0}$ as a variable equals to the indicator function of the event $c_S=0$.

With the help of Stirling formula $n! \sim \sqrt{2\pi n}(\frac{n}{e})^n$ and $\binom{n}{n/2} \sim \sqrt{\frac{2}{\pi}}\frac{2^n}{\sqrt{n}}$ as an approximation, we obtain the following corollary.

Corollary 3.3.4. $E[\mathrm{zero}(f)] = O((1+\frac{1}{\sqrt{2}})^n)$.

Proof. $E[\mathrm{zero}(f)] = E[\sum_{S\subset\{0,1\}^n} X_S] = \sum_{S\subset\{0,1\}^n} \frac{\binom{2^s}{2^{s-1}}}{2^{2^s}} = O(\sum_{S\subset\{0,1\}^n} \frac{1}{\sqrt{2}^s}) = O((1+\frac{1}{\sqrt{2}})^n)$ $\square$

The last equation is by Binomial theorem, $\sum_{i=0}^n \binom{n}{i} a^i b^{n-i} = (a+b)^n$.

3.3.2 Variance

The next important measurement would be the variance, since we have $\mathrm{Var}[\mathrm{zero}(f)] = \mathrm{Var}[\mathrm{spar}(f)] = \mathbb{E}[\mathrm{zero}^2(x)] - \mathbb{E}^2[\mathrm{zero}(f)]$, it's equivalent to measure $\mathbb{E}[\mathrm{zero}^2(x)] = (\sum_{S\in\{0,1\}^n} \Pr[c_S = 0])^2 = \sum_{S,T\in\{0,1\}^n} \Pr[c_S = 0, c_T = 0]$.

First, we will list the probability $\Pr(c_S = 0, c_T = 0)$ for all circumstances in the following Lemma.

Lemma 3.3.5.

$$\Pr[c_S = 0, c_T = 0] = \begin{cases} \frac{\binom{2^s}{2^{s-1}}}{2^{2^s}} & \textit{Case 1: } S = T, \\ \frac{\binom{2^s}{2^{s-1}}}{2^{2^s}} \frac{\binom{2^t-2^s}{2^{t-1}-2^{s-1}}}{2^{2^t-2^s}} & \textit{Case 2: } S \subset T, \\ \frac{\binom{2^s}{2^{s-1}}}{2^{2^s}} \frac{\binom{2^t}{2^{t-1}}}{2^{2^t}} & \textit{Case 3: } S \bigcap T = \emptyset, \\ \sum_{k=-2^{r-1}}^{2^{r-1}} \frac{\binom{2^s-2^r}{2^{s-1}-2^{r-1}+k}}{2^{2^s-2^r}} \frac{\binom{2^t-2^r}{2^{t-1}-2^{r-1}+k}}{2^{2^t-2^r}} \frac{\binom{2^r}{2^{r-1}+k}}{2^{2^r}} & \textit{Case 4: } S \cap T = R. \end{cases} \tag{3.2}$$

Proof. Case 1 is trivial, as S and T are the exact same set.

Case 2 could follow the same proof in Lemma 3.3.2, we may noticed that the Boolean cube B^S and set $B^T \setminus B^S$, these two parts will be independent. The probability for each part being balanced could be referred from Lemma 3.3.2, and the result is multiplying the probabilities together.

All the first three cases are actually special cases of Case 4, let's rewrite the probability of Case 4 for a better understanding. If $S \cap T = R$ and $|R| = r$, $\Pr(c_S = 0, c_T = 0) = \sum_{k=-2^{s-1}}^{2^{s-1}} \Pr(c_R = k, c_S = 0, c_T = 0)$.

Obviously, we may partition the union of Boolean cubes $B^S \cup B^T$ into three independent parts, B^R, $B^S \setminus B^R$ and $B^T \setminus B^R$, and the formula is directly summing up the probability according to how bias is B^R.

Case 1 indicates $B^S = B^T = B^R$, while Case 2 implies that $B^S = B^R$. Both will force the k could only be 0. The Case 3 is only different in the way that $r = 0$,

$|B^R| = 2^r = 1$ and $2^{r-1} = \frac{1}{2}$, k could only be $\frac{-1}{2}$ or $\frac{1}{2}$. We may have $\Pr(c_S = 0, c_T = 0) = \frac{\binom{2^s-1}{2^{s-1}+1}}{2^{2^s-1}} \frac{\binom{2^t-1}{2^{t-1}+1}}{2^{2^t-1}} \frac{\binom{1}{0}}{2} + \frac{\binom{2^s-1}{2^{s-1}-1}}{2^{2^s-1}} \frac{\binom{2^t-1}{2^{t-1}-1}}{2^{2^t-1}} \frac{\binom{1}{1}}{2} = \frac{\binom{2^s}{2^{s-1}}}{2^{2^s}} \frac{\binom{2^t}{2^{t-1}}}{2^{2^t}}$. $\square$

Before having more specific calculation, we show that for any $S, T \in \{0,1\}^n$, $\Pr[c_S = 0, c_T = 0] \geq \Pr[c_S = 0] \Pr[c_T = 0]$, or equivalently, random variables X_S and X_T are positive correlated.

Here we present the Fortuin–Kasteleyn–Ginibre (FKG) inequality with a self-contained folklore proof for completeness.

Lemma 3.3.6 (FKG inequality)**.** *let $\mu : \mathbb{Z} \to \mathbb{R}$ be a non-negative function, and $f, g : \mathbb{Z} \to \mathbb{R}$ be two monotonically non-decreasing functions on $\mathbb{Z}$. Then we have the following inequality:*

$(\sum_x f(x)g(x)\mu(x))(\sum_x \mu(x)) \geq (\sum_x f(x)\mu(x))(\sum_x g(x)\mu(x))$.

Proof.

$$
\begin{aligned}
&LHS - RHS \\
=&\sum_x \sum_{y>x} \mu(x)\mu(y)(f(y)g(y) + f(x)g(x)) - \sum_x \sum_{y>x} \mu(x)\mu(y)(f(x)g(y) + f(y)g(x)) \\
=&\sum_x \sum_{y>x} \mu(x)\mu(y)(f(y) - f(x))(g(y) - g(x)) \geq 0.
\end{aligned}
$$

$\square$

Lemma 3.3.7. $\Pr[c_S = 0, c_T = 0] \geq \Pr[c_S = 0] \Pr[c_T = 0]$

Proof. For Case 1, $S = T$, it's true as $\frac{\binom{2^s}{2^{s-1}}}{2^{2^s}} < 1$.

For Case 2, $S \subset T$, it's true since $\frac{\binom{2^t-2^s}{2^{t-1}-2^{s-1}}}{2^{2^t-2^s}} > \frac{\binom{2^t}{2^{t-1}}}{2^{2^t}}$.

For Case 3, $S \bigcap T = \emptyset$, we have $\Pr(c_S = 0, c_T = 0) = \Pr(c_S = 0) \Pr(c_T = 0))$, so their correlation are 0 when $S \bigcap T = \emptyset$.

Case 4 is when $S \cap T = R$ and $|R| = r$.

Let's take $\mu(x) = \frac{\binom{2^r}{2^{r-1}+x}}{2^{2^r}}$, $f(x) = \frac{\binom{2^s-2^r}{2^{s-1}-2^{r-1}+x}}{2^{2^s-2^r}}$ and $g(x) = \frac{\binom{2^t-2^r}{2^{t-1}-2^{r-1}+x}}{2^{2^t-2^r}}$ and use the facts that $\sum_{k=-2^{r-1}}^{2^{r-1}} \mu(k) = \sum_{k=-2^{r-1}}^{2^{r-1}} \frac{\binom{2^r}{2^{r-1}+k}}{2^{2^r}} = 1$,
$\sum_{k=-2^{r-1}}^{2^{r-1}} \mu(k)f(k) = \sum_{k=-2^{r-1}}^{2^{r-1}} \frac{\binom{2^s-2^r}{2^{s-1}-2^{r-1}+k}}{2^{2^s-2^r}} \frac{\binom{2^r}{2^{r-1}+k}}{2^{2^r}} = \frac{\binom{2^s}{2^{s-1}}}{2^{2^s}}$
and $\sum_{k=-2^{r-1}}^{2^{r-1}} \mu(k)g(k) = \sum_{k=-2^{r-1}}^{2^{r-1}} \frac{\binom{2^t-2^r}{2^{t-1}-2^{r-1}+k}}{2^{2^t-2^r}} \frac{\binom{2^r}{2^{r-1}+k}}{2^{2^r}} = \frac{\binom{2^t}{2^{t-1}}}{2^{2^t}}$.

Then after reordering the items and apply Lemma 3.3.6, we get the following inequality.

$$\begin{aligned}
&\Pr[c_S = 0, c_T = 0] = \sum_{k=-2^{r-1}}^{2^{r-1}} \frac{\binom{2^s-2^r}{2^{s-1}-2^{r-1}+k}}{2^{2^s-2^r}} \frac{\binom{2^t-2^r}{2^{t-1}-2^{r-1}+k}}{2^{2^t-2^r}} \frac{\binom{2^r}{2^{r-1}+k}}{2^{2^r}} \\
&= \sum_{k=-2^{r-1}}^{2^{r-1}} \mu(k)f(k)g(k) = \left(\sum_{k=-2^{r-1}}^{2^{r-1}} \mu(k)f(k)g(k)\right)\left(\sum_{k=-2^{r-1}}^{2^{r-1}} \mu(k)\right) \\
&\geq \left(\sum_{k=-2^{r-1}}^{2^{r-1}} \mu(k)f(k)\right)\left(\sum_{k=-2^{r-1}}^{2^{r-1}} \mu(k)g(k)\right) = \frac{\binom{2^s}{2^{s-1}}}{2^{2^s}} \frac{\binom{2^t}{2^{t-1}}}{2^{2^t}} \\
&= \Pr[c_S = 0] \Pr[c_T = 0].
\end{aligned}$$

□

Now let's focus on upper-bounding the variance $\mathrm{Var}[\mathrm{zero}(f)]$. Given set $S, T \in \{0,1\}^n$ and their intersection set R, with size s, t and r respectively, we may upper-bound the summation by upper-bounding every single item $\mathrm{cov}(X_S, X_T) = \Pr[c_S = 0, c_T = 0] - \Pr[c_S = 0] \Pr[c_T = 0]$. The idea is, supposing we could upper-bound $\mathrm{cov}(X_S, X_T) = O(a^{s-r}b^{t-r}c^r)$, we will obtain $\sum_{|S|=s,|T|=t,|S\cap T|=r} \mathrm{cov}(X_S, X_T) = O((1+a+b+c)^n)$ by multinomial theorem. We will handle the four cases separately.

For Case 1, if $S = T$, we have the following upper-bound,

$$\begin{aligned}
&\sum_{S=T} \mathrm{cov}(X_S, X_T) = \sum_S (\Pr[c_S = 0] - \Pr[c_S = 0]^2) \\
\leq &\sum_S \Pr[c_S = 0] = \mathbb{E}[\mathrm{zero}(f)].
\end{aligned} \tag{3.3}$$

For Case 2, if $S \subset T$, with the help of the Stirling formula we have the following approximations and upper-bounds.

$$\begin{aligned}
&\mathrm{cov}(X_S, X_T) = \frac{\binom{2^s}{2^{s-1}}}{2^{2^s}}\left(\frac{\binom{2^t-2^s}{2^{t-1}-2^{s-1}}}{2^{2^t-2^s}} - \frac{\binom{2^t}{2^{t-1}}}{2^{2^t}}\right) \\
&= O\left(\frac{1}{\sqrt{2^s}}\left(\frac{1}{\sqrt{2^t-2^s}}\left(1+O\left(\frac{1}{2^t-2^s}\right)\right) - \frac{1}{\sqrt{2^t}}\left(1+O\left(\frac{1}{2^t}\right)\right)\right)\right) \\
&= O\left(\frac{\sqrt{2^t} - \sqrt{2^t-2^s} + O\left(\frac{\sqrt{2^t}}{2^t-2^s}\right) + O\left(\frac{\sqrt{2^t-2^s}}{2^t}\right)}{\sqrt{2^s}\sqrt{2^t}\sqrt{2^t-2^s}}\right) \\
&= O\left(\frac{\sqrt{2^{2s-t}} + O\left(\frac{\sqrt{2^t}}{2^t-2^s}\right) + O\left(\frac{\sqrt{2^t-2^s}}{2^t}\right)}{\sqrt{2^s}2^t}\right) \\
&= O\left(\frac{(\sqrt{2})^s}{(2\sqrt{2})^t}\right) = O\left(\frac{1}{(2\sqrt{2})^{t-s}2^s}\right).
\end{aligned} \tag{3.4}$$

Therefore, we have the following bound.

$$\begin{aligned}
&\sum_{S\subset T} \mathrm{cov}(X_S, X_T) \le \sum_{s=0}^{n} \binom{n}{s} \left(\sum_{t=s}^{n} \binom{n-s}{t-s} O\left(\frac{1}{(2\sqrt{2})^{t-s}2^s}\right)\right) \\
&= O\left(\left(\frac{3}{2} + \frac{1}{2\sqrt{2}}\right)^n\right).
\end{aligned} \tag{3.5}$$

For Case 3, if $S \bigcap T = \emptyset$, we will have $\mathrm{cov}(X_S, X_T) = 0$, so is the summation.

For Case 4, if $S \cap T = R$ and $|R| = r$, we use magnifying on $\Pr[c_S = 0, c_T = 0]$ such that,

$$\begin{aligned}
&\Pr[c_S = 0, c_T = 0] = \sum_{k=-2^{r-1}}^{2^{r-1}} \frac{\binom{2^s-2^r}{2^{s-1}-2^{r-1}+k}}{2^{2^s-2^r}} \frac{\binom{2^t-2^r}{2^{t-1}-2^{r-1}+k}}{2^{2^t-2^r}} \frac{\binom{2^r}{2^{r-1}+k}}{2^{2^r}} \\
\le &\frac{\binom{2^t-2^r}{2^{t-1}-2^{r-1}}}{2^{2^t-2^r}} \sum_{k=-2^{r-1}}^{2^{r-1}} \frac{\binom{2^s-2^r}{2^{s-1}-2^{r-1}+k}}{2^{2^s-2^r}} \frac{\binom{2^r}{2^{r-1}+k}}{2^{2^r}} = \frac{\binom{2^t-2^r}{2^{t-1}-2^{r-1}}}{2^{2^t-2^r}} \frac{\binom{2^s}{2^{s-1}}}{2^{2^s}}.
\end{aligned} \tag{3.6}$$

Then by using the same method in (3.4), we can get the following bound:

$$
\begin{aligned}
&\operatorname{cov}(X_S, X_T) \le \frac{\binom{2^s}{2^{s-1}}}{2^{2^s}}\left(\frac{\binom{2^t-2^r}{2^{t-1}-2^{r-1}}}{2^{2^t-2^r}} - \frac{\binom{2^t}{2^{t-1}}}{2^{2^t}}\right) \\
&= O\left(\frac{1}{\sqrt{2^s}}\left(\frac{1}{\sqrt{2^t-2^r}}\left(1+O\left(\frac{1}{2^t-2^r}\right)\right) - \frac{1}{\sqrt{2^t}}\left(1+O\left(\frac{1}{2^t}\right)\right)\right)\right) \\
&= O\left(\frac{\sqrt{2^t} - \sqrt{2^t-2^r} + O\left(\frac{\sqrt{2^t}}{2^t-2^r}\right) + O\left(\frac{\sqrt{2^t-2^r}}{2^t}\right)}{\sqrt{2^s}\sqrt{2^t}\sqrt{2^t-2^r}}\right) \\
&= O\left(\frac{\sqrt{2^{2r-t}} + O\left(\frac{\sqrt{2^t}}{2^t-2^s}\right) + O\left(\frac{\sqrt{2^t-2^s}}{2^t}\right)}{\sqrt{2^s}2^t}\right) \\
&= O\left(\frac{2^r}{(2\sqrt{2})^t(\sqrt{2})^s}\right) = O\left(\frac{1}{(2\sqrt{2})^{t-r}(\sqrt{2})^{s-r}2^r}\right).
\end{aligned}
\tag{3.7}
$$

Then we can have upper-bound as below,

$$
\begin{aligned}
&\sum_{S\cap T\neq\emptyset} \operatorname{cov}(X_S, X_T) \\
&\le \sum_{r=0}^{n}\binom{n}{r}\left(\sum_{s=r}^{n}\binom{n-r}{s-r}\left(\sum_{t=r}^{n-s+r}\binom{n-s}{t-r}O\left(\frac{1}{(2\sqrt{2})^{t-r}(\sqrt{2})^{s-r}2^r}\right)\right)\right) \\
&= O\left(\left(\frac{3}{2}+\frac{3}{2\sqrt{2}}\right)^n\right).
\end{aligned}
\tag{3.8}
$$

Combining (3.3), (3.5) and (3.8) together, we have the following corollary:

Corollary 3.3.8. $\operatorname{Var}[\operatorname{zero}(f)] = O\left(\left(\frac{3}{2}+\frac{3}{2\sqrt{2}}\right)^n\right)$.

Now let's present the famous Chebyshev's inequality and its proof from [AS08].

Lemma 3.3.9 ([AS08], Theorem 4.1.1, Chebyshev's inequality)**.** *Let X be a random variable with finite expected value μ and finite non-zero variance σ^2. Then for any positive λ,*

$$
\Pr[|X-\mu| \ge \lambda\sigma] \le \frac{1}{\lambda^2}. \tag{3.9}
$$

Proof. $\sigma^2 = \operatorname{Var}[X] = E[(X-\mu)^2] \ge \lambda^2\sigma^2 \Pr[|X-\mu| \ge \lambda\sigma]$. $\square$

Apply Lemma 3.3.9 to $\operatorname{zero}(f)$, we can establish the following theorem.

Theorem 3.3.10. *For a random function f uniformly chosen from all possible Boolean functions defined on $f : \{0,1\}^n \to \{0,1\}$, we have the following properties:*

$$\mu = \mathbb{E}[\mathrm{zero}(f)] = O(1 + \frac{1}{\sqrt{2}})^n)),$$
$$\sigma^2 = \mathrm{Var}[\mathrm{zero}(f)] = O((\frac{3}{2} + \frac{3}{2\sqrt{2}})^n),$$

and by applying Lemma 3.3.9 we obtain that,

$$\Pr[\mathrm{zero}(f) \geq 2\mu] \leq O(\sigma^2/\mu^2) = O((\frac{3}{2+\sqrt{2}})^n),$$

or more generally, $\mathrm{zero}(f) = (1 + o(1))\mathbb{E}[\mathrm{zero}(f)]$ *almost surely, since* $\mathrm{Var}[\mathrm{zero}(f)] = o(\mathbb{E}^2[\mathrm{zero}(f)])$.

3.4 Concluding remarks and Open Problems

In this chapter, we give several bounds and concentration results about the distribution of the sparsity for the real polynomial representation of random Boolean functions. However, though the bound for expectation of sparsity is asymptotically tight, there still exists a gap for variance. We conjecture that the variance $\mathrm{Var}[\mathrm{zero}(f)] = O((\frac{3}{2} + \frac{1}{\sqrt{2}})^n)$, and that will consequently lead to better concentration results.

A major motivation for this study is to find the exact distribution of sparsity. However, having only expectation and variance is not enough to characterize the distribution. Another interesting question is to obtain an approximate distribution of $\mathrm{spar}(f)$.

BIBLIOGRAPHY

[AB09] S. Arora and B. Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, 2009.

[ADF95] Karl A Abrahamson, Rodney G Downey, and Michael R Fellows. Fixed-parameter tractability and completeness iv: On completeness for w [p] and pspace analogues. *Annals of pure and applied logic*, 73(3):235–276, 1995.

[ADL18] D. Aggarwal, Y. Dodis, and S. Lovett. Non-malleable codes from additive combinatorics. *SIAM Journal on Computing*, 47(2):524–546, 2018. Earlier version in STOC'14.

[ALM+98] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998. Earlier version in FOCS'92.

[AS98] S. Arora and S. Safra. Probabilistic checking of proofs: A new characterization of NP. *Journal of the ACM*, 45(1):70–122, 1998. Earlier version in FOCS'92.

[AS08] N. Alon and J. Spencer. *The Probabilistic Method*. John Wiley and Sons, third edition, 2008.

[BC99] A. Bernasconi and B. Codenotti. Spectral analysis of Boolean functions as a graph eigenvalue problem. *IEEE Transactions on Computers*, 48(3):345–351, 1999.

[BDL13] A. Bhowmick, Z. Dvir, and S. Lovett. New bounds for matching vector families. In *Proc. 45th Annual ACM Symposium on the Theory of Computing*, pages 823–832, 2013.

[BdW02] H. Buhrman and R. de Wolf. Complexity measures and decision tree complexity: A survey. *Theoretical Computer Science*, 288:21–43, 2002.

[Bei93] Richard Beigel. The polynomial method in circuit complexity. In *[1993] Proceedings of the Eigth Annual Structure in Complexity Theory Conference*, pages 82–95. IEEE, 1993.

[BFR98] R Balasubramanian, Michael R Fellows, and Venkatesh Raman. An improved fixed-parameter algorithm for vertex cover. *Information Processing Letters*, 65(3):163–168, 1998.

[BSLRZ14] E. Ben-Sasson, S. Lovett, and N. Ron-Zewi. An additive combinatorics approach relating rank to communication complexity. *Journal of the ACM*, 61(4):22, 2014.

[BSRZ15] E. Ben-Sasson and N. Ron-Zewi. From affine to two-source extractors via approximate duality. *SIAM Journal on Computing*, 44(6):1670–1697, 2015. Earlier version in STOC'11.

[Cho61] C. Chow. On the characterization of threshold functions. In *Proc. 2nd Annual IEEE Symposium on Foundations of Computer Science*, pages 34–38. IEEE, 1961.

[CMS19] A. Chattopadhyay, N. Mande, and S. Sherif. The Log-approximate-rank conjecture is false. In *Proc. 51st Annual ACM Symposium on the Theory of Computing*, 2019. To appear.

[Coo71] Stephen A Cook. The complexity of theorem-proving procedures. In *Proceedings of the third annual ACM symposium on Theory of computing*, pages 151–158, 1971.

[CP18] A. Chistopolskaya and V. Podolskii. Parity decision tree complexity is greater than granularity, October 2018. `http://arxiv.org/abs/1810.08668`.

[DDFS14] A. De, I. Diakonikolas, V. Feldman, and R. Servedio. Nearly optimal solutions for the Chow parameters problem and low-weight approximation of halfspaces. *Journal of the ACM*, 61(2):1–36, 2014. Earlier version in STOC'12.

[DF12] Rodney G Downey and Michael Ralph Fellows. *Parameterized complexity*. Springer Science & Business Media, 2012.

[EZ12] C. Even-Zohar. On sums of generating sets in $\mathbb{Z}_2^n$. *Combinatorics, probability and computing*, 21(6):916–941, 2012.

[FGL+96] U. Feige, S. Goldwasser, L. Lovász, S. Safra, and M. Szegedy. Interactive proofs and the hardness of approximating cliques. *Journal of the ACM*, 43(2):268–292, 1996. Earlier version in FOCS'91.

[FKN02] E. Friedgut, G. Kalai, and A. Naor. Boolean functions whose Fourier transform is concentrated on the first two levels. *Advances in Applied Mathematics*, 29(3):427–437, 2002.

[FL87] Michael R Fellows and Michael A Langston. Nonconstructive advances in polynomial-time complexity. *Information Processing Letters*, 26(3):157–162, 1987.

[Fre73] G. Freiman. *Foundations of a structural theory of set addition*. American Mathematical Society, Providence, RI, 1973. Translated from the Russian. *Translations of Mathematical Monographs*, Vol 37.

[Fri98] E. Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):27–35, 1998.

[GJ79] M. Garey and D. Johnson. *Computers and intractability: A guide to the theory of NP-completeness*. W. H. Freeman, 1979.

[GKP89] Ronald L Graham, Donald E Knuth, and Oren Patashnik. Concrete mathematics: a foundation for computer science, 1989.

[GOS+11] P. Gopalan, R. O'Donnell, R. Servedio, A. Shpilka, and K. Wimmer. Testing Fourier dimensionality and sparsity. *SIAM Journal on Computing*, 40(4):1075–1100, 2011. Earlier version in ICALP'09.

[GR06] B. Green and I. Ruzsa. Sets with small sumset and rectification. *Bulletin of the London Mathematical Society*, 38(1):43–52, 2006.

[GT09] B. Green and T. Tao. Freiman's theorem in finite fields via extremal set theory. *Combinatorics, Probability and Computing*, 18(3):335–355, 2009.

[HHL18] H. Hatami, K. Hosseini, and S. Lovett. Structure of protocols for XOR functions. *SIAM Journal on Computing*, 47(1):208–217, 2018.

[Juk12] Stasys Jukna. *Boolean function complexity: advances and frontiers*, volume 27. Springer Science & Business Media, 2012.

[KLMY20] Alexander Knop, Shachar Lovett, Sam McGuire, and Weiqiang Yuan. Log-rank and lifting for AND-functions. *arXiv preprint arXiv:2010.08994*, 2020.

[KN97] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, 1997.

[Kon08] S. Konyagin. On the Freiman theorem in finite fields. *Mathematical Notes*, 84(3-4):435–438, 2008.

[Łab01] I. Łaba. Fuglede's conjecture for a union of two intervals. *Proceedings of the American Mathematical Society*, 129(10):2965–2972, 2001.

[Lev73] L. A. Levin. Universal sequential search problems. *Problemy Peredachi Informatskii*, 9(3):115–116, 1973.

[Lov14] S. Lovett. Communication is bounded by root of rank. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 842–846, 2014.

[Lov17] S. Lovett. Additive combinatorics and its applications in theoretical computer science. *Theory of Computing*, pages 1–55, 2017.

[LS88] L. Lovász and M. Saks. Lattices, Möbius functions and communication complexity. In *Proc. 29th Annual IEEE Symposium on Foundations of Computer Science*, pages 330–337, 1988.

[LZ17] C. Lin and S. Zhang. Sensitivity conjecture and log-rank conjecture for functions with small alternating numbers. In *Proc. 44th Annual International Conference on Automata, Languages, and Programming*, volume 80, pages 51:1–51:13, 2017.

[MP88] Marvin L Minsky and Seymour Papert. Perceptrons: an introduction to computational geometry. 1988.

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

[NS94] N. Nisan and M. Szegedy. On the degree of Boolean functions as real polynomials. *Computational Complexity*, 4:301–313, 1994. Earlier version in STOC'92.

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

[OS11] R. O'Donnell and R. Servedio. The Chow parameters problem. *SIAM Journal on Computing*, 40(1):165–199, 2011. Earlier version in STOC'08.

[OST+14] R. O'Donnell, X. Sun, L. Y. Tan, J. Wright, and Y. Zhao. A composition theorem for parity kill number. In *Proc. 29th Annual IEEE Conference on Computational Complexity*, pages 144–154, 2014.

[Pat92] Ramamohan Paturi. On the degree of polynomials that approximate symmetric Boolean functions (preliminary version). In *Proceedings of the twenty-fourth annual ACM symposium on Theory of computing*, pages 468–474, 1992.

[PY96] Christos H Papadimitriou and Mihalis Yannakakis. On limited nondeterminism and the complexity of the vc dimension. *Journal of Computer and System Sciences*, 53(2):161–170, 1996.

[Raz87] Alexander A Razborov. Lower bounds for the size of circuits of bounded depth with basis f^; g. *Math. notes of the Academy of Sciences of the USSR*, 41(4):333–338, 1987.

[Ruz99] I. Ruzsa. An analog of Freiman's theorem in groups. *Astérisque*, 258(199):323–326, 1999.

[RvL74] B. L. Rothschild and J. van Lint. Characterizing finite subspaces. *Journal of Combinatorial Theory, Series A*, 16(1):97–110, 1974.

[Sam07] A. Samorodnitsky. Low-degree tests at large distances. In *Proc. 39th Annual ACM Symposium on the Theory of Computing*, pages 506–515, 2007.

[San08] T. Sanders. A note on Freĭman's theorem in vector spaces. *Combinatorics, Probability and Computing*, 17(2):297–305, 2008.

[Smo87] Roman Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proceedings of the nineteenth annual ACM symposium on Theory of computing*, pages 77–82, 1987.

[STlV17] A. Shpilka, A. Tal, and B. lee Volk. On the structure of boolean functions with small spectral norm. *computational complexity*, 26(1):229–273, 2017.

[TV06] T. Tao and V. Vu. *Additive Combinatorics*. Cambridge University Press, 2006.

[TWXZ13] H. Tsang, C. Wong, N. Xie, and S. Zhang. Fourier sparsity, spectral norm, and the Log-rank conjecture. In *Proc. 54th Annual IEEE Symposium on Foundations of Computer Science*, pages 658–667, 2013.

[TXZ16] H. Tsang, N. Xie, and S. Zhang. Fourier sparsity of GF(2) polynomials. In *Proceedings of the International Computer Science Symposium in Russia*, pages 409–424, 2016.

[Yao79] A. C. Yao. Some complexity questions related to distributive computing. In *Proc. 11th Annual ACM Symposium on the Theory of Computing*, pages 209–213, 1979.

[ZS10] Z. Zhang and Y. Shi. On the parity complexity measures of Boolean functions. *Theoretical Computer Science*, 411(26-28):2612–2618, 2010.

VITA

YEKUN XU

	Born, Shenyang, Liaoning, China
2012	B.E., Computer Science Tsinghua University Beijing, China
2016–present	Ph.D., Computer Science, School of Computing & Information Sciences Florida International University Miami, Florida

PUBLICATIONS

- Ning Xie, Shuai Xu, Yekun Xu, *A New Algorithm for Finding Closest Pair of Vectors*, In *Proceedings of the 13th International Computer Science Symposium in Russia (CSR 2018)*, 2018
- Ning Xie, Shuai Xu, Yekun Xu, *A new coding-based algorithm for finding closest pair of vectors*, In *Journal of Theoretical Computer Science*, 2019.
- Ning Xie, Shuai Xu, Yekun Xu, *A Generalization of a Theorem of Rothschild and van Lint*, In *Proceedings of the 16th International Computer Science Symposium in Russia (CSR 2021)*, 2021.
- Daniel Chen, Yekun Xu, Betis Baheri, Samuel A Stein, Chuan Bi, Ying Mao, Qiang Quan, Shuai Xu, *Quantum-Inspired Classical Algorithm for Slow Feature Analysis*, In *24th Annual Conference on Quantum Information Processing (QIP 2021 poster session)*, 2021.
- Daniel Chen, Yekun Xu, Betis Baheri, Chuan Bi, Ying Mao, Qiang Quan, Shuai Xu, *Quantum-Inspired Classical Algorithm for Principal Component Regression (submitted)*, 2021.