Algebraic Methods in Pseudorandomness and Circuit Complexity

by

Zachary Remscrim

B.S. MIT (2009)
M.ENG. MIT (2010)

Submitted to the Department of Electrical Engineering and Computer Science
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

at the

Massachusetts Institute of Technology

June 2016

Signature of Author: Signature redacted

Department of Electrical Engineering and Computer Science
May 20, 2016

Certified by: Signature redacted

Michael Sipser
Dean of the School of Science, Professor of Mathematics
Thesis Supervisor

Accepted by: Signature redacted

Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chairman, Committee for Graduate Students

Algebraic Methods in Pseudorandomness and Circuit Complexity

by

Zachary Remscrim

Submitted to the Department of Electrical Engineering and Computer Science
on May 20, 2016 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy in
Electical Engineering and Computer Science

at the

Massachusetts Institute of Technology

June 2016

ABSTRACT

In this thesis, we apply tools from algebra and algebraic geometry to prove new results concerning extractors for algebraic sets, AC^0-pseudorandomness, the recursive Fourier sampling problem, and VC dimension. We present a new construction of an extractor which works for algebraic sets defined by polynomials over $\mathbb{F}_2$ of substantially higher degree than the previous state-of-the-art construction. We exhibit a collection of natural functions that behave pseudorandomly with regards to AC^0 tests. We also exactly determine the $\mathbb{F}_2$-polynomial degree of the recursive Fourier sampling problem and use this to provide new partial results towards a circuit lower bound for this problem. Finally, we answer a question posed in [MR15] concerning VC dimension, interpolation degree and the Hilbert function.

Thesis Supervisor: Michael Sipser

Title: Dean of the School of Science, Professor of Mathematics

Contents

Chapter 1

Introduction

1.1 Pseudorandomness

1.1.1 Algebraic Extractors

For a finite domain Ω and a collection of distributions $\mathcal{C}$ over Ω, we say that a function $E : \Omega \to \{0,1\}^m$ is an *extractor* (sometimes called a *deterministic extractor*) for $\mathcal{C}$ if, for every random variable X distributed according to any distribution in $\mathcal{C}$, $E(X)$ is close to the uniform distribution. We call each distribution $C \in \mathcal{C}$ a *source*. Of course, in order to have any hope of the collection of distributions $\mathcal{C}$ to have an extractor, some sort of condition must be satisfied by the sources. While it is trivial to exhibit simple conditions on $\mathcal{C}$ such that a random function will, with high probability, be an extractor, the problem becomes far more interesting when one requires an *explicit* construction of E (that is to say, a construction realizable by some deterministic polynomial time Turing machine). The natural question is then: for which $\mathcal{C}$ do there exist explicit constructions of extractors?

Numerous versions of this question have been considered. In this thesis, we consider the case, originally introduced in [Dvi12], where each source is the uniform distribution over the set of common zeros of a collection of polynomials defined over some field. Such a set is called an *algebraic set* and such a source is called an *algebraic source*. Algebraic sources are a

natural generalization of *affine sources* (see, for instance [GR05] and [Bou07]) and *bit-fixing sources* (see, for instance, [GRS04] and [KZ03]) and build naturally on the earlier work of *efficiently samplable sources* (see, for instance, [TV00], [KRVZ06], and [DGW07]).

To be precise, for a finite field $\mathbb{F}$, and a positive integer d, we consider algebraic sets $V \subseteq \mathbb{F}^n$ where V is the set of common zeros of a collection of polynomials $f_1, \ldots, f_t \in \mathbb{F}[x_1, \ldots, x_n]$ such that $\deg(f_i) \leq d$. We say that V has *density* ρ if $|V| \geq \rho|\mathbb{F}^n|$. We say that a function $f : \mathbb{F}^n \to \mathbb{F}$ is an *extractor for algebraic sets* defined by polynomials of degree at most d and density ρ if f is close to uniform on every such algebraic set. A closely related weaker notion is that of a *disperser for algebraic sets*, where we say that a function $f : \mathbb{F}^n \to \mathbb{F}$ is a *disperser for algebraic sets* defined by polynomials of degree at most d and density ρ if, for every such algebraic set V, the image of $f : V \to \mathbb{F}$ (the restriction of f to V) is $\mathbb{F}$. Clearly any extractor is also a disperser.

As shown in [Dvi12], there exist explicit extractors for polynomials of degree d defined over moderately sized fields, where $|\mathbb{F}| = poly(d)$, and density $\rho = 2^{-\frac{n}{2}}$ as well as over large fields, where $|\mathbb{F}| = d^{\Omega(n^2)}$ and very small density. However, very little is known about the extreme case in which $\mathbb{F} = \mathbb{F}_2$, the two element finite field. To the best of our knowledge, the current state of the art construction for extractors and dispersers is that of [CT13], in which an explicit construction was exhibited for an extractor for algebraic sets defined by at most $(\log \log n)^{\frac{1}{2c}}$ polynomials each of degree at most 2, as well as for a disperser for algebraic sets defined by at most t polynomials each of degree at most $d = (1 - o(1))\frac{\log(\frac{n}{t})}{\log^{0.9} n}$ (in particular, when $t \leq n^{\alpha}$ for some $\alpha < 1$, then the requirement on degree is $d < (1 - \alpha - o(1)) \log^{0.1} n$).

In this thesis, we focus on the case in which $\mathbb{F} = \mathbb{F}_2$, and exhibit explicit extractors (and hence explicit dispersers) for algebraic sets defined by polynomials of substantially higher degree than any previous construction. We now formally state our results. For any set $V \subseteq \mathbb{F}_2^n$, we say that a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ has *bias* ϵ on V if

$$\text{bias}(f|_V) := |\mathbb{E}_{x \sim V}[(-1)^{f(x)}]| \leq \epsilon.$$

A function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is called an *extractor for algebraic sets* defined by polynomials of degree at most d of density ρ with bias ϵ if $\text{bias}(f|_V) \leq \epsilon$ for every such algebraic set V. We show that any δ-versatile function, which is a certain natural generalization of the concept of a versatile function [Kop11], is an extractor.

Theorem 1. *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be δ-versatile (on $\mathbb{F}_2^n$), where $\delta \geq \frac{n}{2} - n^\gamma$ for some $0 \leq \gamma < \frac{1}{2}$. Then, there is a constant $c > 0$ such that, for any constants α, β such that $0 < \alpha, \beta < \frac{1}{2}$, and for any $d \leq n^\alpha$ and $\rho \geq 2^{-n^\beta}$, f is an extractor with bias $\frac{c\left(n^\gamma + d\log(\frac{\sqrt{n}}{\rho})\right)}{\sqrt{n}}$ for algebraic sets of density at least ρ that are the common zeros of a collection of polynomials each of degree at most d.*

Much as was the case in [Dvi12] and [CT13], our construction relies on statements involving the set of zeros of a single low degree polynomial defined over $\mathbb{F}$. The key distinction between our construction and earlier constructions, which allows our construction to work even for rather high degree polynomials over $\mathbb{F}_2$, is that our construction exploits the structure of this set of zeros, rather than simply bounds on the size of the set of zeros that follow directly from the degree of the polynomial (that is to say, bounds that follow directly from the fundamental theorem of algebra, or, in other words, Schwartz-Zippel type bounds).

1.1.2 AC^0-pseudorandomness

Random-like behavior occurs naturally in many places in mathematics. For example, the binary representations of numbers π, e and $\sqrt{2}$ look random. Various conjectures about the distribution of prime numbers and the number of prime factors of an integer say that these behave randomly. However, very little progress has been made in proving that such behaviors are indeed pseudorandom in any formal sense. For example, it is not known that the binary representations of π, e or $\sqrt{2}$, contain all substrings with the expected frequencies or even that the substring 11 appears infinitely often.

In this thesis, we propose to study the pseudorandom characteristics of naturally occurring mathematical functions by using the tools of complexity theory. The theory of

pseudorandom generators provides a good starting point, but there the motivation is somewhat different than ours. Pseudorandom generators are used to good effect in cryptographic protocols and in derandomizing probabilistic algorithms, and they are designed with those goals in mind. Our objective is to study the basic operations themselves, such as Boolean convolution and integer multiplication, for their pseudorandom properties. These functions occur naturally—they have not been specifically designed to have pseudorandom behaviour—yet we can show that they do exhibit such behavior.

We use the integer multiplication function as a motivating example. Let X and Y be n-bit binary strings representing non-negative integers and let Z be the $2n$-bit string representing $Z = X \times Y$. Take X and Y to be selected uniformly at random from 0 to $2^n - 1$, and consider the characteristics of Z. Does Z look random? The low-order bit of Z certainly does not; it is 0 with probability 3/4. The other very low order bits look non-random for a similar reason. The very high order bits of Z likewise appear non-random. However, if we discard these problematic very low and very high order bits from Z, the result could conceivably be pseudorandom in some appropriate sense.

We show that, for uniformly randomly selected X, Y, the string consisting of X, Y and all $2n$ bits of $X \times Y$, except the lowest and highest n^α bits, for any $\alpha > 0$, is indistinguishable from truly random strings by AC^0 circuits. In fact, we show something even stronger: for almost all Y, the string consisting of X and all $2n$ bits of $X \times Y$, except the lowest and highest n^α bits, is indistinguishable from random by AC^0 circuits that have Y built-in (the circuit is allowed to depend on Y).

AC^0 circuits are circuits consisting of *AND*, *OR*, and *NOT* gates of unbounded fan-in, such that the size of the circuit (the total number of gates) is polynomial in the size of the input and the depth of the circuit (the number of gates on the longest path from the input to the output) is a constant. Techniques for proving strong lower bounds on low-depth circuits [Ajt83],[FSS84],[Yao85],[Has86] enable us to prove the AC^0-pseudorandomness of explicit functions without using any unproven complexity-theoretic assumptions. Moreover

AC^0 is powerful enough to describe basic tests for pseudorandomness.

We now formally define what it means for a function to look random to AC^0 circuits. For ease of exposition, we consider functions that operate on strings of a specific length, whereas we really have in mind a family of functions and their asymptotic properties. For a function $f : \{0,1\}^{m_1} \times \cdots \times \{0,1\}^{m_h} \to \{0,1\}^k$, define the function $g : \{0,1\}^{m_1} \times \cdots \times \{0,1\}^{m_h} \to \{0,1\}^n$, where $n = m_1 + \cdots + m_h + k$, such that $g(x_1, \ldots, x_h) = x_1 \circ \cdots \circ x_h \circ f(x_1, \ldots, x_h)$ is the concatenation of $x_1, \ldots, x_h$ and $f(x_1, \ldots, x_h)$. Let μ_n denote the distribution of $g(x_1, \ldots, x_h)$, when each x_i is drawn uniformly at random from $\{0,1\}^{m_i}$. For any binary predicate $P_n : \{0,1\}^n \to \{0,1\}$, let $E_{\mu_n}[P_n]$ denote the expected value of P_n when inputs are drawn according to the distribution μ_n and $E[P_n]$ denote the expected value of P_n when inputs are drawn uniformly at random from $\{0,1\}^n$. We say that the distribution μ_n ***ϵ-fools*** the function P_n if $|E_{\mu_n}[P_n] - E[P_n]| < \epsilon$ and that the original function f is ***AC^0-pseudorandom*** if the corresponding distribution μ_n ϵ-fools every P_n that is computable in AC^0, where $\epsilon = O(2^{-n^\kappa})$, for constant $\kappa > 0$. This is, of course, quite similar to the standard pseudorandom generator model for AC^0 circuits (see, for instance, [Nis91], [NW94]), with the exception of the fact that we impose the stronger requirement that both the input and output of the function together are indistinguishable from random bits, instead of only requiring that the output is indistinguishable. Also, while the focus of this thesis is the pseudorandomness of functions, not the difficulty of actually computing the functions, it is still worth noting that the functions considered can be computed in a low complexity class such as $AC^0[2]$ (AC^0 circuits that are allowed unbounded fan-in parity gates) or TC^0 (constant depth circuits with unbounded fan-in majority gates), but still produce strings that are indistinguishable from truly random strings by AC^0 circuits.

A somewhat similar question was considered in [Gre12], concerning the Möbius function $\mu : \mathbb{N} \to \{-1, 0, 1\}$, which is defined such that $\mu(1) = 1$, $\mu(x) = 0$ when x has a nontrivial perfect square factor, and $\mu(x) = (-1)^k$, when x has no nontrivial perfect square factors, where k is the number of distinct primes in the prime factorization of x. It was shown

that μ is asymptotically orthogonal to any AC^0 computable function $f : \mathbb{N} \to \{-1, 1\}$ (that is to say $\frac{1}{N}\sum_{x=1}^{N} f(x)\mu(x) = o(1)$). Tools from complexity theory were used to show that a naturally occurring function looks random to AC^0 circuits. It is worth noting that the functions considered in our thesis have much longer output than the Möbius function; we consider functions which, on an n bit input, produce a $\Omega(n)$ bit output, while the Möbius function maps an n bit input to only a constant sized output.

Another example of a natural problem studied for its pseudorandom properties is the algebraic number problem, which, as noted in [KLL84], was initially proposed by Manuel Blum. An algebraic number is a root of a polynomial with integer coefficients. For example, $\sqrt{2}, \sqrt{3}$, and $(1 + \sqrt{5})/2$ are all algebraic numbers. The algebraic number problem involves selecting, uniformly at random, an algebraic number ζ of bounded degree d and height H (where the degree of ζ is the degree of the (unique) primitive irreducible polynomial that has ζ as a root, and the height is the Euclidean length of the coefficient vector of that polynomial). The string to be considered is a portion of the binary expansion of the fractional part of ζ. In [KLL84], it was shown that, given the first $O(d^2 + d \log H)$ bits of an algebraic number ζ, it is possible, in deterministic polynomial time to determine the minimal polynomial of ζ. Since the next bit of the binary expansion of ζ can easily be obtained if given the minimal polynomial of ζ, this immediately implied that such strings do not pass all polynomial time tests. We consider a closely related problem, which is identical to the above problem, except that we select ζ only from the ring of integers of certain algebraic number fields. By the argument used in [KLL84], this variant also does not pass all polynomial time tests. However, we show that it does pass all AC^0 tests. While this is certainly far away from showing anything about the pseudorandomness properties of a single value, such as $\sqrt{2}$, it might be a step in that direction.

This thesis illustrates two techniques for demonstrating that functions are AC^0-pseudorandom. The first technique makes use of the result in [Bra09] that resolved the long standing Linial-Nisan conjecture [LN90]. We use this technique to show that almost

all "reasonably sized" homomorphisms are AC^0-pseudorandom, and, moreover, that convolution, integer multiplication and matrix multiplication are AC^0-pseudorandom.

Our second technique involves reducing the (provably hard) problem of computing parity to the problem of distinguishing certain distributions from random. The second technique is related to the method in [Nis91],[NW94], in that we show that the structure of certain multiplication problems is a naturally occuring example of the combinatorial designs they employ. We use this technique to show that an alternate form of the multiplication problem, where one multiplicand is substantially longer than the other, is AC^0-pseudorandom. One consequence of this result will be the existence of a simple, multiplication-based pseudorandom generator with the same stretch and hardness parameters as the Nisan-Wigderson generator. An additional consequence is the fact that no AC^0 circuit can compute the product of an n-bit number and a superpolylog(n)-bit number (that is to say, a sequence of numbers whose length grows faster than $\log^c n$, for all constants $c > 0$). This shows that the result from [CSV84], which states that an AC^0 circuit can compute the product of an n-bit and a $O(log^c n)$-bit value, is optimal.

Additionally, we show, via a reduction from the multiplication problem, that a certain variant of the algebraic integer problem looks random to AC^0. These same techniques can be used to show that a variety of additional problems, such as finite field multiplication and division, matrix inversion, computing determinants, and an iterated version of convolution are also AC^0-pseudorandom.

We prove the following theorems:

Let $Hom(\{0,1\}^m, \{0,1\}^k)$ denote the set of homomorphisms from $\{0,1\}^m$ to $\{0,1\}^k$ (or, in other words, the linear maps from the vector space $\{0,1\}^m$ to the vector space $\{0,1\}^k$).

Theorem 2. *If $k = m^u$, for any fixed constant $u > 0$, then all but an exponentially small fraction of all $f \in Hom(\{0,1\}^m, \{0,1\}^k)$ are AC^0-pseudorandom.*

Let $\text{CONV}_{r,s,k} : \{0,1\}^r \times \{0,1\}^s \to \{0,1\}^k$ denote the Boolean convolution function, which takes a $X \in \{0,1\}^r$ and $Y \in \{0,1\}^s$ to the middle k-bits of the $r + s - 1$ bit long

convolution of X and Y.

Theorem 3. *If $s = r^u$ and $k = r + s - (MIN(r,s))^\alpha$, for any fixed constants $u > 0$ and $0 < \alpha < 1$, then $CONV_{r,s,k}$ is AC^0-pseudorandom. In particular, if $r = s$ and $k = 2r - r^\alpha$, for any $0 < \alpha < 1$, then $CONV_{r,s,k}$ is AC^0-pseudorandom.*

Let $\text{MULT}_{r,s,k} : \{0,1\}^r \times \{0,1\}^s \to \{0,1\}^k$ denote the integer multiplication function, which takes a $X \in \{0,1\}^r$ and $Y \in \{0,1\}^s$ to the middle k-bits of the $r + s$ bit long product of X and Y.

Theorem 4. *If $s = r^u$ and $k = r + s - (MIN(r,s))^\alpha$, for any fixed constants $u > 0$ and $0 < \alpha < 1$, then $MULT_{r,s,k}$ is AC^0-pseudorandom. In particular, if $r = s$ and $k = 2r - r^\alpha$, for any $0 < \alpha < 1$, then $MULT_{r,s,k}$ is AC^0-pseudorandom.*

Let $\text{MATRIX-MULT}_{r,s} : \{0,1\}^{rs} \times \{0,1\}^{rs} \to \{0,1\}^{s^2}$ denote the matrix multiplication function, which, on input a $s \times r$ matrix A and a $r \times s$ matrix B (both of which are encoded as strings in $\{0,1\}^{rs}$ in the obvious way), produces the $s \times s$ matrix AB.

Theorem 5. *If $s = r^u$, for any fixed constant $u > 0$, then $MATRIX\text{-}MULT_{r,s}$ is AC^0-pseudorandom.*

1.2 Polynomial Degree

1.2.1 Recursive Fourier Sampling

The recursive Fourier sampling problem is one of the most well studied problems in quantum complexity theory. This problem was first defined, along with the complexity class BQP (Bounded-Error Quantum Polynomial Time), in [BV93], the foundational work of quantum complexity theory. In that thesis, this problem, whose formal definition we delay for now, was used to exhibit an oracle A relative to which BQP is not contained in NP or even MA, that is to say an A such that $\text{BQP}^A \not\subset \text{NP}^A$ and $\text{BQP}^A \not\subset \text{MA}^A$. Such oracle separations are

interesting both because they are, perhaps, suggestive of a unrelativized separation, as well as because they concretely exhibit a measure of complexity in which quantum computers provably outperform classical computers: query complexity, where the resource of interest is the number of queries to the (very long) input string.

For this reason, it is natural to seek oracle separations between BQP and increasingly larger classical complexity classes. However, very little progress in this direction has been made. While some results are known, such as the fact, proven in [Aar10], that there is an oracle A such that $\mathrm{BQP}^A \not\subset \mathrm{BPP}^A_{\mathrm{path}}$ and $\mathrm{BQP}^A \not\subset \mathrm{SZK}^A$, even the question of whether or not there exists an oracle A such that $\mathrm{BQP}^A \not\subset \mathrm{AM}^A$ remains open, as does, of course, the substantially stronger question of whether or not there exists an oracle A such that $\mathrm{BQP}^A \not\subset \mathrm{PH}^A$.

It is this potential oracle separation between BQP and the polynomial hierarchy that we now focus on. The natural approach to this problem, which has been used successfully to show many other similar oracle separations between certain complexity classes and the polynomial hierarchy, is to exploit the connection between relativized separations from the polynomial hierarchy and lower bounds against constant depth circuits [FSS84],[Yao85]. Here, the key idea is to reinterpret the $\exists$ and $\forall$ quantifiers of a PH machine as *OR* and *AND* gates, respectively, to convert a PH machine solving some oracle problem on a 2^n bit long oracle string, into a constant depth, $2^{poly(n)}$ sized circuit, consisting of AND, OR, and NOT gates that solves the same problem. Using this idea, and a $2^{\omega(poly(n))}$ lower bound on the size of a constant depth circuit computing the PARITY function (on an input of size 2^n), one concludes that there is an oracle A relative to which $\oplus\mathrm{P}^A \not\subset \mathrm{PH}^A$. The same idea can, and has, been used to show other such relativized separations.

Therefore, given this connection between relativized separations from the polynomial hierarchy and lower bounds against constant depth circuits, and the powerful techniques that exist to show lower bounds against constant depth circuits, [FSS84],[Ajt83],[Has86],[Raz87],[Smo87], one might very naturally ask why the question of whether or not there exists an A such that

$\mathrm{BQP}^A \not\subset \mathrm{PH}^A$ remains open. Most fundamentally, the problem is that, in order to show that a particular function f cannot be computed by a small circuit, all of these circuit lower bound techniques either explicitly (in the case of [Raz87] or [Smo87]) or implicitly (in the case of [FSS84],[Ajt83],[Has86] as shown by [LMN93]) argue that f cannot be well approximated by a low-degree polynomial. This is a problem because, as shown in [BBC+98], any function that can be computed by an efficient quantum algorithm is well approximated by a low degree polynomial.

More precisely, however, [BBC+98] only guarantees the existence of a low-degree polynomial over $\mathbb{R}$, whereas the non-existence of a low-degree polynomial over any field $\mathbb{F}$ would suffice (via the Razborov-Smolensky method) to prove a circuit lower bound, and so this certainly does not completely doom the application of traditional circuit lower bound techniques. Nevertheless, the result of [BBC+98] does suggest that a deeper understanding of approximation by low-degree polynomials may be necessary to resolve the question of whether or not there exists an oracle A such that $\mathrm{BQP}^A \not\subset \mathrm{PH}^A$. It is this issue that we focus on within this thesis.

As has been observed by many authors (for instance [BV93],[BV97],[Aar03],[Joh08],[Aar10]) the recursive Fourier sampling problem (or a slight variant) is a prime candidate for exhibiting an oracle A such that $\mathrm{BQP}^A \not\subset \mathrm{PH}^A$, as this problem seems to perfectly exploit the advantages of a quantum computer at the expense of a classical one.

We delay the formal definition of the recursive Fourier sampling problem. For the moment, we will simply state that it is a promise problem (that is to say, a partial Boolean function whose value is only defined on a portion of the input space, called the promise) which is known to have an efficient quantum algorithm. By the result of [BBC+98], this immediately implies that there is a low degree real polynomial that well approximates the recursive Fourier sampling problem on the promise. In fact, from the standpoint of proving a circuit lower bound, the situation is even "worse" than this, due to the result of [Joh11], which shows that there is an even lower degree real polynomial than the one guaranteed by

[BBC+98] which exactly represents the recursive Fourier sampling problem on its promise. Moreover, [Joh11] proves exactly matching upper and lower bounds on any real polynomial that represents the recursive Fourier sampling problem on its promise, thereby completely resolving the question of the polynomial degree of the recursive Fourier sampling problem, with respect to polynomials over $\mathbb{R}$.

In this thesis, we consider the question of the polynomial degree of the recursive Fourier sampling problem for polynomials defined over $\mathbb{F}_2$. That is to say, we consider the question of what is the lowest degree polynomial defined over $\mathbb{F}_2$ that represents the recursive Fourier sampling problem on its promise. Before proceeding further, we briefly note that this question is only non-trivial because the recursive Fourier sampling problem is a promise problem. For any total function $g : \mathbb{F}_2^n \to \mathbb{F}_2$, there is a unique multilinear polynomial $f \in \mathbb{F}_2[x_1, \ldots, x_n]$ that agrees everywhere with g; the degree of f is, of course, the minimal degree of any polynomial in $\mathbb{F}[x_1, \ldots, x_n]$ that agrees everywhere with g. For a promise problem, however, there can be many multilinear polynomials, of varying degrees, that all agree on the promise.

Over $\mathbb{F}_2$, there is a simple, though relatively high degree, polynomial that exactly computes the recursive Fourier sampling problem. Our key result, stated in the following theorems, is that, for a certain appropriate settings of the parameters, this simple polynomial is, in fact, the lowest degree polynomial that agrees with recursive Fourier sampling everywhere on its promise. In fact, we show something even stronger: no polynomial of lower degree can even non-trivially one-sided agree with the recursive Fourier sampling problem (that is to say, if a polynomial is zero everywhere (on the promise) that the recursive Fourier sampling problem is zero, then that polynomial must be zero on the entire promise). We then use these results to prove new statements about the ability of constant depth circuits to compute the recursive Fourier sampling problem.

Theorem 7. *For any positive integers k, h, Let $n = 2^k - 1$ and let $RFS_{n,h}^{MAJ}$ denote the recursive Fourier sampling function with majority. Then $\nexists g \in \mathbb{F}_2[x_1, \ldots, x_m]$ such that*

deg$(g) < \left(\frac{n+1}{2}\right)^h$ *and* $g(x) = RFS^{MAJ}_{n,h}(x)\ \forall x \in U^{MAJ}_{p,h}$. *Moreover, if any* $g \in \mathbb{F}_2[x_1, \ldots, x_m]$ *such that deg*$(g) < \left(\frac{n+1}{2}\right)^h$ *vanishes everywhere on* $U^{MAJ}_{0,h}$, *it vanishes everywhere on* $U^{MAJ}_{1,h}$.

Theorem 8. *For any positive integers* d, n, h *such that* $d|n$, *and* $n \geq d(2^{d^2} + d - 1)$, *Let* $RFS^{GIP_{n,d}}_{n,h}$ *denote the recursive Fourier sampling function with generalized inner product. Then* $\nexists g \in \mathbb{F}_2[x_1, \ldots, x_m]$ *such that deg*$(g) < d^h$ *and* $g(x) = RFS^{GIP_{n,d}}_{n,h}(x)\ \forall x \in U^{GIP_{n,d}}_{p,h}$. *Moreover, if any* $g \in \mathbb{F}_2[x_1, \ldots, x_m]$ *such that deg*$(g) < d^h$ *vanishes everywhere on* $U^{GIP_{n,d}}_{0,h}$, *it vanishes everywhere on* $U^{GIP_{n,d}}_{1,h}$.

1.2.2 VC Dimension

We say that a subset $J \subseteq [n]$ is *shattered* by a family of vectors $C \subseteq \{0,1\}^n$ if, $\forall s : J \to \{0,1\}$, $\exists c \in C$ such that $c_j = s(j)\ \forall j \in J$ (in other words, if one considers the set of all substrings of elements of C comprised of the positions indexed by J, this collection of substrings is precisely $\{0,1\}^{|J|}$). We then write

$$\text{str}(C) = \{J \subseteq [n] : J \text{ is shattered by} C\}$$

to denote the sets that are shattered with respect to C. We then define the *VC dimension* of C as

$$\text{VC}(C) = \max\{|J| : J \in \text{str}(C)\}.$$

For a field $\mathbb{F}$ and a set $C \subseteq \{0,1\}^n$, the interpolation degree of C, denoted by reg(C) is the minimum d such that every function $f : C \to \mathbb{F}$ can be expressed as a multilinear polynomial in $\mathbb{F}[x_1, \ldots, x_n]$ of degree at most d.

Recently, in [MR15], a very interesting connection between VC dimension and interpolation degree was demonstrated. A simple characterization of sets with interpolation degree 1 was provided. This naturally raised the question of whether a similar characterization exists for sets with interpolation degree r, for arbitrary r. In this thesis, we provide such a characterization, in terms of the rank of a certain inclusion matrix, which will be

defined precisely in §2.

Theorem 9. *A set $C \subseteq \{0,1\}^n$ has $reg(C) = r$ if and only if r is the smallest positive integer such that $rank_{\mathbb{F}_2} \mathcal{M}(C, \binom{[n]}{\leq r}) = |C|$.*

Chapter 2

Algebraic Geometry Tools

2.1 Preliminaries

We begin by recalling several standard definitions from algebraic geometry. Let $\mathbb{F}$ denote a (not necessarily algebraically closed) field and $\mathbb{F}[x_1, \ldots, x_n]$ denote the ring of polynomials in n indeterminates. An *algebraic set* in $\mathbb{F}^n$ is the set of common zeros of a collection of polynomials in $\mathbb{F}[x_1, \ldots, x_n]$. More precisely, given a set of polynomials $f_1, \ldots, f_k \in \mathbb{F}[x_1, \ldots, x_n]$, we denote their set of common zeros by $V(f_1, \ldots, f_k)$ where

$$V(f_1, \ldots, f_k) = \{(x_1, \ldots, x_n) \in \mathbb{F}^n : f_i(x_1, \ldots, x_n) = 0 \ \forall i\}.$$

Rather than working with an arbitrary set of polynomials, it will often be convenient to consider an algebraically nicer object: an ideal. For I an ideal in $\mathbb{F}[x_1, \ldots, x_n]$, let $V(I)$ denote the common zero set of all polynomials in I, that is to say

$$V(I) = \{(x_1, \ldots, x_n) \in \mathbb{F}^n : f(x) = 0 \ \forall f \in I\}.$$

Given a set of polynomials $f_1, \ldots, f_k \in \mathbb{F}[x_1, \ldots, x_n]$, let $\langle f_1, \ldots, f_k \rangle$ denote the ideal which they generate in $\mathbb{F}[x_1, \ldots, x_n]$. Clearly, $V(\langle f_1, \ldots, f_k \rangle) = V(f_1, \ldots, f_k)$. For an algebraic

set V, let its *vanishing ideal* $I(V)$ be the ideal of $\mathbb{F}[x_1, \ldots, x_n]$ consisting of all polynomials which vanish on V and let $R(V) = \mathbb{F}[x_1, \ldots, x_n]/I(V)$ denote its *coordinate ring.*

For a polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$, let $\deg(f)$ denote its total degree. Let $\mathbb{F}[x_1, \ldots, x_n]_{\leq d}$ denote the vector space of polynomials over $\mathbb{F}$ with degree at most d. For an ideal I, let $I_{\leq d} = I \cap \mathbb{F}[x_1, \ldots, x_n]_{\leq d}$ denote the subspace consisting of all polynomials in I of degree at most d. For an algebraic set V, with vanishing ideal $I = I(V)$ and coordinate ring $R = R(V)$, let $R_{\leq d} = \mathbb{F}[x_1, \ldots, x_n]_{\leq d}/I_{\leq d}$. The *affine Hilbert function* $h^a(R, d)$ of R is then given by

$$h^a(R, d) = \dim_{\mathbb{F}}(R_{\leq d}).$$

By slight abuse of notation, we will use the term *affine Hilbert function of an algebraic set* V, which we will denote $h^a(V, d)$, to simply be the affine Hilbert function of the coordinate ring $R(V)$.

Throughout this thesis, we consider only zero-dimensional algebraic sets V (that is to say, V is finite). For such a V, we define its *regularity* $\mathrm{reg}(V)$ to be the minimal value of d such that $h^a(V, d) = |V|$. Equivalently, $\mathrm{reg}(V)$ is the minimal value of d such that every function $V \to \mathbb{F}$ can be realized as a polynomial of degree at most d. This quantity is frequently referred to as *interpolation degree.* In the case of zero-dimensional algebraic sets, this quantity is equivalent to the Castelnuovo-Mumford regularity of $R(V)$ (see, for instance [Eis02] Thm.4.1).

For $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$, we define x^α to be the monomial $x_1^{\alpha_1} \cdots x_n^{\alpha_n} \in \mathbb{F}[x_1, \ldots, x_n]$. For any $J \subseteq [n]$ we define the (multilinear) monomial x_J by $x_J = \prod_{j \in J} x_j$. A *degree compatible term order* $<$ is a total order on the monomials x^α which respects multiplication ($x^\alpha < x^\beta \Rightarrow x^\alpha x^\gamma < x^\beta x^\gamma \; \forall x^\alpha, x^\beta, x^\gamma \in \mathbb{F}[x_1, \ldots, x_n]$) and is degree compatible ($\deg(x^\alpha) < \deg(x^\beta) \Rightarrow x^\alpha < x^\beta \; \forall x^\alpha, x^\beta \in \mathbb{F}[x_1, \ldots, x_n]$). For a degree compatible term order $<$, and polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$, we define its *leading monomial* $\mathrm{lm}(f)$ to be the largest monomial in f with respect to $<$. Similarly, for an ideal I in $\mathbb{F}[x_1, \ldots, x_n]$, we define its

leading monomials to be

$$\mathrm{LM}(I) = \{\mathrm{lm}(f) : f \in I\}$$

and its *standard monomials* to be

$$\mathrm{SM}(I) = \{x^\alpha : \alpha \in \mathbb{N}^n\} \setminus \mathrm{LM}(I).$$

For an algebraic set V, we define $\mathrm{LM}(V) = \mathrm{LM}(I(V))$ and $\mathrm{SM}(V) = \mathrm{SM}(I(V))$. We also define

$$\mathrm{SM}(V, d) = \{x^\alpha \in \mathrm{SM}(V) : \deg(x^\alpha) = d\}$$

and

$$\mathrm{LM}(V, d) = \{x^\alpha \in \mathrm{LM}(V) : \deg(x^\alpha) = d\}.$$

Standard monomials provide an extremely convenient tool for computing both the Hilbert function of an algebraic set and its regularity, as illustracted in the following lemma (these are well known facts in algebraic geometry; see, for instance [Fel07]).

Lemma 1. *(a)* $h^a(V, d) = \sum_{i=0}^{d} |SM(V, i)|$

(b) $reg(V) = \max_{x^\alpha \in SM(V)} deg(x^\alpha)$

(c) $|SM(V)| = |V|$

(d) $V_1 \subseteq V_2 \Rightarrow SM(V_1) \subseteq SM(V_2)$

(e) $V_1 \subseteq V_2 \Rightarrow LM(V_1) \supseteq LM(V_2)$

Let M_n denote the semigroup of all monomials in n indeterminates. That is to say, as a set $M_n = \{x^\alpha : \alpha \in \mathbb{N}^n\}$ with multiplication between monomials defined in the usual way. An ideal U of M_n is simply an upwardly closed subset of M_n ($x^\alpha \in U \Rightarrow x^\alpha x^\beta \in U \ \forall \alpha, \beta$). For an algebraic set $V \subseteq \mathbb{F}^n$, $\mathrm{LM}(V)$ is an ideal of M_n. Similarly, $\mathrm{SM}(V)$ is a dual ideal. In

other words, if $x^\alpha \in \mathrm{LM}(V)$, then $x^\alpha x^\beta \in \mathrm{LM}(V)$ and if $x^\alpha \in \mathrm{SM}(V)$ then $x^\beta \in \mathrm{SM}(V)$ for any divisor x^β of x^α.

For I an ideal of $\mathbb{F}[x_1, \ldots, x_n]$, let $a(I)$ denote the minimal degree of any $g \in I$ such that g consists of only monomials from $\mathrm{SM}(\mathbb{F}^n)$. For an algebraic set $V = V(I)$, let $a(V) = a(I)$. The following lemma, proven independently in [Fel07] and [PR08], provides an extremely useful relationship between $\mathrm{reg}(V)$ and $a(\overline{V})$, where $\overline{V}$ denotes the complement of V.

Lemma 2. *[Fel07], [PR08]*

If $V \subseteq \mathbb{F}^n$ is a nonempty zero-dimensional algebraic set, then $a(\overline{V}) + reg(V) = n$.

Lastly, we consider another useful tool for computing the Hilbert function: inclusion matrices. Let $\mathbb{F}_2$ denote the finite field of two elements. Let $2^{[n]}$ denote the collection of all subsets of $[n] = \{1, \ldots, n\}$, and let $\mathcal{F}, \mathcal{G} \subseteq 2^{[n]}$ denote two families of subsets. The *inclusion matrix* $\mathcal{M}(\mathcal{F}, \mathcal{G})$ is a $|\mathcal{F}| \times |\mathcal{G}|$ matrix, with entries in $\mathbb{F}_2$, where for any $F \in \mathcal{F}$ and $G \in \mathcal{G}$ the (F, G) entry is 1 precisely when $G \subseteq F$. Let $\binom{[n]}{\leq k}$ denote the family of all subsets of $[n]$ of size at most k.

Given an algebraic set $V \subseteq \mathbb{F}_2^n$, we associate it with a family of subsets in the natural way: for each $x = (x_1, \ldots, x_n) \in V$ the subset $\{i : x_i = 1\}$ is included in the set family. By a slight abuse of notation, we will also denote this set family by V. The following is immediate from definitions (as a nontrivial linear combination of the columns corresponds to a polynomial in $I(V)$ and hence a leading monomial).

Lemma 3. *For any algebraic set $V \subseteq \mathbb{F}_2^n$, we have*

$$h^a(V, d) = rank_{\mathbb{F}_2} \mathcal{M}\left(V, \binom{[n]}{\leq d}\right).$$

Throughout this thesis, our key object of interest will be the affine Hilbert function of an algebraic set. We briefly note that this is a slight departure from the typical situation in algebraic geometry in which one considers the "ordinary" Hilbert function (which is defined

similarly to the affine Hilbert function, but in which one considers the space of homogeneous polynomials of a particular degree, rather than arbitrary polynomials of a particular degree) of a variety (which is an algebraic set in which the ground field $\mathbb{F}$ is algebraically closed). Much as was the case in [Smo93], this is done in order to allow a better intuitive connection between the Hilbert function and the questions from complexity theory that we consider. However, it should be noted that it is very straightforward to convert between statements involving the affine Hilbert function of an algebraic set and the Hilbert function of a variety as, firstly, one can harmlessly extend the ground field (and, in particular, extend it to its algebraic closure), and, secondly, one can straightforwardly express the value of the affine Hilbert function at degree d as the sum of values of the Hilbert function of degree at most d. While it is true that certain basic statements that would hold over an algebraically closed ground field do not necessarily hold over arbitrary fields, these statements are either facts that we explicitly exploit in the proof (such as the number of roots a particular degree d polynomial has in a particular algebraic set) or are statements that can easily be modified to analagous statements when the ground field is a finite field (for example, Hilbert's Nullstellensatz, which establishes a bijection between varieties and radical ideals can be modified to a bijection between algebraic sets and radical ideals that contain the field polynomials).

2.2 Generalization of Versatile Functions

In this section, we consider a certain natural generalization of the concept of versatile functions (as defined in [Kop11], see also [Smo87] for the concept of $U_P^n - \text{complete}$ elements) to promise problems. We begin with a definition.

Definition 1. *A function* $f : \mathbb{F}_2^n \to \mathbb{F}_2$ *is Versatile if,* $\forall g : \mathbb{F}_2^n \to \mathbb{F}_2$, $\exists u, v \in \mathbb{F}_2[x_1, \ldots, x_n]$ *where* $deg(u), deg(v) \leq \frac{n}{2}$ *and* $g(x) = u(x)f(x) + v(x)\ \forall x \in \mathbb{F}_2^n$.

Versatile functions admit a particularly simple characterization in terms of regularity (this is essentially the same notion as "degree-m independent sets" as considered in [Smo93]),

as shown in the following lemma.

Lemma 4. *For a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$, let $U_0 = f^{-1}(0)$ and $U_1 = f^{-1}(1)$. Then f is versatile if and only if $reg(U_0), reg(U_1) \leq \frac{n}{2}$.*

Proof. If f is versatile, then, by definition, $\forall g : \mathbb{F}_2^n \to \mathbb{F}_2$, $\exists u, v \in \mathbb{F}_2[x_1, \ldots, x_n]$ where $\deg(u), \deg(v) \leq \frac{n}{2}$ and $g(x) = u(x)f(x) + v(x)\ \forall x \in \mathbb{F}_2^n$, and so $g(x) = v(x)\ \forall x \in U_0$ and $g(x) = u(x) + v(x)\ \forall x \in U_1$. Since $\deg(u+v) \leq \max(\deg(u), \deg(v))$, it immediately follows that $\mathrm{reg}(U_0), \mathrm{reg}(U_1) \leq \frac{n}{2}$.

If $\mathrm{reg}(U_0), \mathrm{reg}(U_1) \leq \frac{n}{2}$, then, by definition, $\forall g : \mathbb{F}_2^n \to \mathbb{F}_2$, $\exists u', v' \in \mathbb{F}_2[x_1, \ldots, x_n]$ where $\deg(u'), \deg(v') \leq \frac{n}{2}$ such that $g(x) = u'(x)\ \forall x \in U_0$ and $g(x) = v'(x)\ \forall x \in U_1$. Therefore, $g(x) = u(x)f(x) + v(x)\ \forall x \in \mathbb{F}_2^n$, where $u = u' + v'$ and $v = v'$. Since $\deg(u), \deg(v) \leq \frac{n}{2}$, f is versatile.

□

As shown in [Kop11], the Majority function (the function $\mathrm{MAJ} : \mathbb{F}_2^n \to \mathbb{F}_2$ where $\mathrm{MAJ}(x) = 1$ when $wt(x) \geq \frac{n}{2}$ and $\mathrm{MAJ}(x) = 0$ when $wt(x) < \frac{n}{2}$, where $wt(x)$ denotes the number of 1s in x) is versatile. As a first illustration of the utility of standard monomials, we present a new short proof of this fact.

Lemma 5. *The function $MAJ : \mathbb{F}_2^n \to \mathbb{F}_2$ is versatile.*

Proof. Let $U_0 = \{x \in \mathbb{F}_2^n : \mathrm{MAJ}(x) = 0\}$. Let $S = \{x^\alpha : \alpha \in \{0,1\}^n, wt(\alpha) < \frac{n}{2}\}$. We will show that $\mathrm{SM}(U_0) = S$. Since $|S| = |U_0| = |\mathrm{SM}(U_0)|$, it suffices to show $\overline{S} \subseteq \mathrm{LM}(U_0)$. To see this, note that for any $J \subseteq [n]$, where $|J| \geq \frac{n}{2}$, we clearly have $x_J \in I(U_0)$ (because, for any $x \in U_0$, a strict majority of the x_j are 0 and so any sufficiently large product $x_J = \prod_{j \in J}$ must vanish on U_0) and so $x_J \in \mathrm{LM}(U_0)$. Trivially, $x_j^2 \in \mathrm{LM}(U_0)\ \forall j$, as, of course, $x_j^2 + x_j \in I(U_0)\ \forall j$. Due to the fact that $\mathrm{LM}(U_0)$ is upwardly closed, the previous two facts immediately imply $\overline{S} \subseteq \mathrm{LM}(U_0)$, as desired.

Similarly, if $U_1 = \{x \in \mathbb{F}_2^n : \mathrm{MAJ}(x) = 1\}$, then, by the same logic as above, $\mathrm{SM}(U_1) = \{x^\alpha : \alpha \in \{0,1\}^n, wt(\alpha) \leq \frac{n}{2}\}$. Therefore, by definition, $\mathrm{reg}(U_0), \mathrm{reg}(U_1) \leq \frac{n}{2}$.

□

We now generalize the notion of versatility to functions of the form $f : U \to \mathbb{F}_2$, for some $U \subseteq \mathbb{F}_2^n$. As shown above, a versatile function partitions the set $\mathbb{F}_2^n$, which has regularity n, into two pieces, the preimage of 0 and the preimage of 1, which each have regularity at most $\frac{n}{2}$. We will call a function f δ-versatile on U if the function f induces a partitioning of U with a regularity gap of at least δ. This notion is formalized in the following definition.

Definition 2. *For a function $f : U \to \mathbb{F}_2$, let $U_0 = \{x \in U : f(x) = 0\}$ and $U_1 = \{x \in U : f(x) = 1\}$. We say that f is δ-versatile on U if $\delta \leq reg(U) - reg(U_0), reg(U) - reg(U_1)$.*

Clearly, this notion generalizes the concept of versatility as a versatile function is $\frac{n}{2}$-versatile on $\mathbb{F}_2^n$. We now prove several useful properties of δ-versatile functions which will be used throughout the thesis.

Lemma 6. *If $f : U \to \mathbb{F}_2$ is δ-versatile on U then, $\nexists g \in \mathbb{F}_2[x_1, \ldots, x_n]$ where $deg(g) < \delta$ and $g(x) = f(x)\ \forall x \in U$.*

Proof. Assume, for contradiction, that such a g exists. By the definition of regularity, there exists at least one function $h : U \to \mathbb{F}_2$ such that, $\forall q \in \mathbb{F}_2[x_1, \ldots, x_n]$ with $\deg(q) < \mathrm{reg}(U)$, $\exists x \in U$ such that $h(x) \neq q(x)$.

Let $U_0 = \{x \in U : f(x) = 0\}$ and $U_1 = \{x \in U : f(x) = 1\}$. Due to the fact that f is δ-versatile on U we have, by definition, $\mathrm{reg}(U_0), \mathrm{reg}(U_1) \leq \mathrm{reg}(U) - \delta$. Therefore, $\exists u, v \in \mathbb{F}_2[x_1, \ldots, x_n]$ where $\deg(u), \deg(v) \leq \mathrm{reg}(U) - \delta$ and $h(x) = u(x)\ \forall x \in U_0$, $h(x) = v(x)\ \forall x \in U_1$. If we then define $q \in \mathbb{F}_2[x_1, \ldots, x_n]$ by $q = u(g + 1) + vg$, we clearly have $\deg(q) \leq \max(\deg(u) + \deg(g), \deg(v) + \deg(g)) \leq (\mathrm{reg}(U) - \delta) + \deg(g) < (\mathrm{reg}(U) - \delta) + \delta = \mathrm{reg}(U)$ and $h(x) = u(x)(g(x) + 1) + v(x)g(x) = q(x)\ \forall x \in U$, which is a contradiction.

□

Next, we consider the behavior of δ-versatile functions $f : U \to \mathbb{F}_2$ where the set U has a certain special property. Given any $U \subseteq \mathbb{F}_2^n$, there is, of course, a unique multilinear

polynomial (recall that a polynomial is multilinear if every monomial has degree at most 1 in each variable) $r_U \in \mathbb{F}_2[x_1, \ldots, x_n]$ such that $r_U(x) = 1$ if and only if $x \in U$. Clearly, $r_U \in I(\overline{V})$. Moreover, each monomial of r_U is in $\mathrm{SM}(\mathbb{F}_2^n)$ (due to the fact that the standard monomials of $\mathbb{F}_2^n$ are precisely the multilinear monomials), and so we immediately conclude that $a(\overline{V}) \leq \deg(r_U)$. We call an algebraic set U *critical* if $a(\overline{V}) = \deg(r_U)$.

Lemma 7. *Let $U \subseteq \mathbb{F}_2^n$ be a critical algebraic set, let $f : U \to \mathbb{F}_2$ be δ-versatile on U, and let $U_0 = \{x \in U : f(x) = 0\}$ and $U_1 = \{x \in U : f(x) = 1\}$. Then, $\forall q \in \mathbb{F}_2[x_1, \ldots, x_n]$ such that $\deg(q) < \delta$, $q \in I(U_0)$ if and only if $q \in I(U_1)$.*

Proof. We show that,$\forall q \in \mathbb{F}_2[x_1, \ldots, x_n]$, where $\deg(q) < \delta$, $q \in I(U_0) \Rightarrow q \in I(U_1)$; the reverse implication follows by symmetry. Assume, for contradiction that $q \in I(U_0)$ but $q \notin I(U_1)$. Let $Y = \{x \in U : q(x) = 1\}$. Clearly $Y \subseteq U_1$ and Y is nonempty. Let $t \in \mathbb{F}_2[x_1, \ldots, x_n]$ denote the unique multilinear polynomial such that $t(x) = r_U(x)q(x)\ \forall x \in \mathbb{F}_2^n$, then $t \in I(\overline{Y})$ and $\deg(t) \leq \deg(r_U) + \deg(q)$. Using Lemma 2, we have

$$\begin{aligned}
\mathrm{reg}(Y) &= n - a(\overline{Y}) \\
&\geq n - \deg(t) \\
&\geq n - \deg(r_U) - \deg(q) \\
&= \mathrm{reg}(U) - \deg(q) \\
&> \mathrm{reg}(U) - \delta \\
&\geq \mathrm{reg}(U_1).
\end{aligned}$$

However, we cannot possibly have $\mathrm{reg}(Y) > \mathrm{reg}(U_1)$ because, as noted above, $U_1 \subseteq U$, and so, by Lemma 1(b,d) we must have $\mathrm{reg}(Y) \leq \mathrm{reg}(U_1)$.

□

The following lemma provides an extremely useful characterization of the behavior of

a δ-versatile f on the intersection of a critical U with a certain simple algebraic set, namely the union of the vanishing sets of a collection of low degree polynomials.

Lemma 8. *Let $U \subseteq \mathbb{F}_2^n$ be a critical algebraic set, let $f : U \to \mathbb{F}_2$ be δ-versatile on U, and let $U_0 = \{x \in U : f(x) = 0\}$ and $U_1 = \{x \in U : f(x) = 1\}$. For any $d < \delta$ and for any $g_1, \ldots, g_k \in \mathbb{F}_2[x_1, \ldots, x_n]$ where $deg(g_i) < d$ $\forall i$, let $G = \cup_i V(g_i)$. Then,*

$$SM(U \cap G, j) = SM(U_0 \cap G, j) = SM(U_1 \cap G, j) \ \forall j \leq \delta - d$$

Proof. Clearly, $U_0 \cap G \subseteq U \cap G$, $U_1 \cap G \subseteq U \cap G$ and so by Lemma 1(d), $\mathrm{SM}(U_0 \cap G), \mathrm{SM}(U_1 \cap G) \subseteq \mathrm{SM}(U \cap G)$, from which it immediately follows that $\mathrm{SM}(U_0 \cap G, j), \mathrm{SM}(U_1 \cap G, j) \subseteq \mathrm{SM}(U \cap G, j)$.

We will now show $\mathrm{SM}(U_0 \cap G, j), \mathrm{SM}(U_1 \cap G, j) \supseteq \mathrm{SM}(U \cap G, j) \forall j \leq \delta - d$, which will complete the proof. Consider any $j \leq \delta - d$. Due to the fact that, for any particular algebraic set, every monomial is either a leading monomial or a standard monomial, if suffices to show $\mathrm{LM}(U_0 \cap G, j), \mathrm{LM}(U_1 \cap G, j) \subseteq \mathrm{LM}(U \cap G, j)$.

To see that $\mathrm{LM}(U_0 \cap G, j) \subseteq \mathrm{LM}(U \cap G, j)$, assume, for contradiction, that this is not the case. Then $\exists x^\alpha \in \mathrm{LM}(U_0 \cap G, j) \cap \mathrm{SM}(U \cap G, j)$. Due to the fact that $x^\alpha \in \mathrm{LM}(U_0 \cap G, j)$ we have, by definition, that $\exists q \in \mathbb{F}_2[x_1, \ldots, x_n]$ such that $q \in I(U_0 \cap G)$ and $\mathrm{lm}(q) = x^\alpha$. Clearly, $\deg(q) = j \leq \delta - d$. Due to the fact that $x^\alpha \in \mathrm{SM}(U \cap G, j)$, we have, by definition $q \notin I(U \cap G)$. This immediately implies $q \notin I(U_1 \cap G)$ because $U \cap G = (U_0 \cup U_1) \cap G = (U_0 \cap G) \cup (U_1 \cap G)$, and so if q did vanish on $U_1 \cap G$, then it would vanish on $U \cap G$ (because, by construction, it vanishes on $U_0 \cap G$). Moreover, since $U_1 \cap G = U_1 \cap (\cup_i V(g_i)) = \cup_i (U_1 \cap V(g_i)$ we conclude $\exists i$ such that $q \notin I(U_1 \cap V(g_i))$. Fix such an i and consider the set $Y = \{x \in U : q(x) = 1 \text{ and } g_i(x) = 0\}$. Notice that due to the requirements that $x \in U$ and $g_i(x) = 0$, we immediately have $Y \subseteq U \cap V(g_i)$, and since q vanishes on $U_0 \cap V(g_i)$, we then have $Y \subseteq U_1 \cap V(g_i)$. Let $t \in \mathbb{F}_2[x_1, \ldots, x_n]$ be the (unique) multilinear polynomial equal to $(r_U)(q)(g_i + 1)$. By construction, $t(x) = 1$ if and

only if $x \in Y$, and so $t \in I(\overline{Y})$. We then have

$$a(\overline{Y}) \leq \deg(t)$$

$$\leq \deg(r_U) + \deg(q) + \deg(g_i + 1)$$

$$< a(\overline{U}) + (\delta - d) + d$$

$$= a(\overline{U}) + \delta.$$

Applying Lemma 2, we then have

$$\operatorname{reg}(Y) = n - a(\overline{Y})$$

$$> n - (a(\overline{U}) + \delta)$$

$$= (n - a(\overline{U})) - \delta$$

$$= \operatorname{reg}(U) - \delta$$

$$\geq \operatorname{reg}(U_1),$$

where the last inequality holds due to the fact that f is δ-versatile. However, we cannot possibly have $\operatorname{reg}(Y) > \operatorname{reg}(U_1)$ because, as noted above, $U_1 \subseteq U$, and so, by Lemma 1(b,d) we must have $\operatorname{reg}(Y) \leq \operatorname{reg}(U_1)$. This contradiction allows us to conclude $\operatorname{LM}(U_0 \cap G, j) \subseteq \operatorname{LM}(U \cap G, j)$. By a precisely symmetric argument, $\operatorname{LM}(U_1 \cap G, j) \subseteq \operatorname{LM}(U \cap G, j)$, which completes the proof.

□

Chapter 3

Pseudorandomness

3.1 Algebraic Extractors

In this section, we exhibit a new construction for an extractor for algebraic sets with extremely strong parameters. We begin with the following lemma, which provides a useful bound on the Hilbert function.

Lemma 9. *Let $V \subseteq \mathbb{F}_2^n$ satisfy $reg(V) \geq \frac{n}{2} - \sqrt{n}$. Then, there is a constant $c > 0$ such that, for any $\beta > 0$ and any $k \leq n^{\frac{1}{2}-\beta}$, we have*

$$h^a(V, reg(V)) - h^a(V, reg(V) - k) \leq \frac{ck}{\sqrt{n}}|V|.$$

Proof. Let $r = \text{reg}(V)$ and set t to be the unique value $r - k + 1 \leq t \leq n$ such that $\binom{t}{r-k+1} \leq |\text{SM}(V, r-k+1)| \leq \binom{t+1}{r-k+1}$.

First, notice that $|\text{SM}(V,i)| \geq \binom{t}{i}$, $\forall i \leq r-k+1$. This follows by a straightforward induction on $j = r-k+1-i$. The case in which $j = 0$ follows from the above definition of t. If $|\text{SM}(V, r-k+1-j)| \geq \binom{t}{r-k+1-j}$, then we immediately have a set $S \subseteq \text{SM}(V, r-k+1-j)$ such that $|S| = \binom{t}{r-k+1-j}$. Define the set ΔS to consist of all monomials that lie immediately below some monomial in S in the monomial order (this is frequently called the *shadow* of

S),

$$\Delta(S) = \{x^\alpha : \deg(x^\alpha) = r - k + 1 - (j+1) \text{ and } \exists x^\gamma \in S \text{ such that } x^\alpha < x^\gamma\}.$$

Due to the fact that $\mathrm{SM}(V)$ is a dual ideal, we note that $S \subseteq \mathrm{SM}(V) \Rightarrow \Delta(S) \subseteq \mathrm{SM}(V)$, from which we immediately conclude $\Delta(S) \subseteq \mathrm{SM}(V, r-k+1-(j+1))$. We then have

$$|\mathrm{SM}(V, r-k+1-(j+1))| \geq |\Delta(S)| \geq \binom{t}{r-k+1-(j+1)},$$

where the last inequality follows immediately from Lovász's version [Lov79] of the Kruskal-Katona theorem.

By a precisely analogous argument, we also have $|\mathrm{SM}(V,i)| \leq \binom{t+1}{i}$ $\forall i \geq r-k+1$. By Lemma 1(a) and the above,

$$h^a(V,r) \geq h^a(V, r-k) = \sum_{i=0}^{r-k} |\mathrm{SM}(V,i)| \geq \sum_{i=0}^{r-k} \binom{t}{i} \geq c_1 2^t,$$

for some constant $c_1 > 0$ (where the last inequality follows from the fact that $r - k > \frac{n}{2} - 2\sqrt{n} \geq \frac{t}{2} - 2\sqrt{t}$ combined with elementary bounds on the sum of binomial coefficients). Similarly, $\forall i \geq r-k+1$, we have, for some constant $c_2 > 0$,

$$|\mathrm{SM}(V,i)| \leq \binom{h+1}{i} \leq \binom{h+1}{\lceil \frac{h+1}{2} \rceil} \leq \frac{c_2 2^h}{\sqrt{h}}.$$

We then have, for some constant $c > 0$,

$$\begin{aligned}
\frac{h^a(V,r) - h^a(V,r-k)}{|V|} &= \frac{h^a(V,r) - h^a(V,r-k)}{h^a(V,r)} \\
&= \frac{\sum_{i=r-k}^{r} |\mathrm{SM}(V,i)|}{h^a(V,r)} \\
&\leq \frac{(k+1)(c_2)\frac{2^t}{\sqrt{t}}}{c_1 2^t}
\end{aligned}$$

$$= \frac{(k+1)\frac{c_2}{c_1}}{\sqrt{t}}$$
$$\leq \frac{ck}{\sqrt{n}}.$$

□

Remark 1. *The above bound can be seen to be essentially optimal, as shown by considering the standard monomials of the function MAJORITY computed in the previous section.*

We now show that any δ-versatile function, for appropriately chosen δ is an extractor.

Theorem 1. *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be δ-versatile (on $\mathbb{F}_2^n$), where $\delta \geq \frac{n}{2} - n^\gamma$ for some $0 \leq \gamma < \frac{1}{2}$. Then, there is a constant $c > 0$ such that, for any constants α, β such that $0 < \alpha, \beta < \frac{1}{2}$, and for any $d \leq n^\alpha$ and $\rho \geq 2^{-n^\beta}$, f is an extractor with bias $\frac{c\left(n^\gamma + d\log(\frac{\sqrt{n}}{\rho})\right)}{\sqrt{n}}$ for algebraic sets of density at least ρ that are the common zeros of a collection of polynomials each of degree at most d.*

Proof. Let $U_0 = f^{-1}(0)$ and $U_1 = f^{-1}(1)$. Due to the fact that f is $(\frac{n}{2} - n^\gamma)$-versatile, we immediately have $\mathrm{reg}(U_0), \mathrm{reg}(U_1) \leq \frac{n}{2} + n^\gamma$. We also have $\mathrm{reg}(U_0), \mathrm{reg}(U_1) \geq \frac{n}{2} - n^\gamma$ because $2^n = |U_0| + |U_1| = |\mathrm{SM}(U_0)| + |\mathrm{SM}(U_1)|$, and the regularity of an algebraic set is the size of its largest standard monomial(Lemma 1(b)).

Consider any algebraic set $V = V(g_1, \ldots, g_k)$ where $g_i \in \mathbb{F}_2[x_1, \ldots, x_n]$ and $\deg(g_i) \leq d \ \forall i$. Using the Razborov-Smolensky method [Raz87],[Smo87], we have a collection of polynomials $y_1, \ldots, y_l \in \mathbb{F}_2[x_1, \ldots, x_n]$ such that $\deg(y_i) \leq d$, $V(g_1, \ldots, g_k) \subseteq V(y_1, \ldots, y_l)$ and $|V(y_1, \ldots, y_l) \setminus V(g_1, \ldots, g_k)| \leq 2^{n-l}$. Setting $y = 1 + \prod_{i=1}^{l}(1 + y_i)$, we then have $\deg(y) \leq dl$ and $V(y) = V(y_1, \ldots, y_l)$.

Consider $U_0 \cap V(y)$ and $U_1 \cap V(y)$. By Lemma 8, we have

$$\mathrm{SM}(U_0 \cap V(y), i) = \mathrm{SM}(U_1 \cap V(y), i) = \mathrm{SM}(V(y)) \ \forall i \leq \frac{n}{2} - n^\gamma - dl.$$

From this, and Lemma 1(a), we immediately conclude $h^a(U_0 \cap V(y), \frac{n}{2} - n^\gamma - dl) = h^a(U_1 \cap$

$V(y), \frac{n}{2} - n^\gamma - dl)$. Clearly, $U_0 \cap V(y) \subseteq U_0$ and $U_1 \cap V(y) \subseteq U_1$, and so, by Lemma 1(d,b), we have $\text{reg}(U_0 \cap V(y)) \leq \text{reg}(U_0) \leq \frac{n}{2} + n^\gamma$, and $\text{reg}(U_1 \cap V(y)) \leq \text{reg}(U_1) \leq \frac{n}{2} + n^\gamma$. Moreover, $\text{reg}(U_0 \cap V(y)), \text{reg}(U_1 \cap V(y)) \geq \frac{n}{2} - n^\gamma - dl$. To see this, first notice that Lemma 2 allows us to conclude $\text{reg}(V(y)) \geq n - dl$ (because $y + 1$ vanishes on the complement of $V(y)$), which immediately implies that $\text{SM}(V(y))$ consists of an element x^κ of degree at least $n - dl$. As $\text{SM}(V(y))$ is a dual ideal, we then also conclude that it consists of an element of degree precisely $\frac{n}{2} - n^\gamma - dl$ (simply take any divisor of x^κ of the appropriate degree). By the above relationship between $\text{SM}(V(y))$, $\text{SM}(U_0 \cap V(y))$ and $\text{SM}(U_1 \cap V(y))$, we then conclude that both $\text{SM}(U_0 \cap V(y))$ and $\text{SM}(U_1 \cap V(y))$ contain an element of degree $\frac{n}{2} - n^\gamma - dl$, and so, by Lemma 1(b), the claimed lower bound on regularity follows. In the following, for brevity, we write $H_i(j) = h^a(U_i \cap V(y), j)$, $d_1 = \frac{n}{2} + n^\gamma$,$d_2 = \frac{n}{2} - n^\gamma - dl$.

We then have

$$\text{bias}(f|_{V(g_1,\ldots,g_k)}) = |\mathbb{E}_{x \sim V(g_1,\ldots,g_k)}[(-1)^{f(x)}]|$$

$$= \frac{||U_0 \cap V(g_1,\ldots,g_k)| - |U_0 \cap V(g_1,\ldots,g_k)||}{|V(g_1,\ldots,g_k)|}$$

$$\leq \frac{||U_0 \cap V(y)| - |U_1 \cap V(y)|| + |V(y) \setminus V(g_1,\ldots,g_k)|}{|V(g_1,\ldots,g_k)|}$$

$$\leq \frac{|H_0(d_1) - H_1(d_1)| + 2^{n-l}}{|V(g_1,\ldots,g_k)|}$$

$$= \frac{|H_0(d_2) - H_1(d_2) + (H_0(d_1) - H_0(d_2)) - (H_1(d_1) - H_1(d_2)) + 2^{n-l}}{|V(g_1,\ldots,g_k)|}$$

$$= \frac{|(H_0(d_1) - H_0(d_2) - (H_1(d_1) - H_1(d_2)) + 2^{n-l}}{|V(g_1,\ldots,g_k)|}$$

$$\leq \frac{\frac{c'(2n^\gamma + dl)}{\sqrt{n}}|V(g_1,\ldots,g_k)| + 2^{n-l}}{|V(g_1,\ldots,g_k)|}$$

$$= \frac{c'(2n^\gamma + dl)}{\sqrt{n}} + \frac{2^{n-l}}{|V(g_1,\ldots,g_k)|}$$

$$\leq \frac{c'(2n^\gamma + dl)}{\sqrt{n}} + \frac{2^{n-l}}{\rho 2^n}$$

$$= \frac{c'(2n^\gamma + dl)}{\sqrt{n}} + \frac{1}{\rho 2^l}.$$

Setting $l = \log(\frac{\sqrt{n}}{\rho})$ yields the claimed bound.

□

Next, as in [CT13], we consider a variant of the extractor model in which, rather than explicitly considering algebraic sets which satisfy a certain density bound, we consider algebraic sets defined by a limited number of polynomials. The following is immediate.

Corollary 1. *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be δ-versatile (on $\mathbb{F}_2^n$), where $\delta \geq \frac{n}{2} - n^\gamma$ for some $0 \leq \gamma < \frac{1}{2}$. Then, there is a constant $c > 0$ such that, for any constants α, β such that $0 < \alpha \leq \beta < \frac{1}{2}$, and for any $d \leq n^\alpha$ and $k \leq n^{\beta-\alpha}$, f is an extractor with bias $\frac{c(n^\gamma + d(n^\beta + \frac{1}{2}\log(n))}{\sqrt{n}}$ for algebraic sets that are the common zeros of a collection of at most k polynomials each of degree at most d.*

Proof. Consider any algebraic set $V = V(g_1, \ldots, g_k)$ where $g_i \in \mathbb{F}_2[x_1, \ldots, x_n]$ and $\deg(g_i) \leq d \ \forall i$. Let $g = 1 + \prod_{i=1}^k g_i$. Then $\deg(g) \leq kd \leq n^\beta$ and $V = V(g)$. From Lemma 2 it immediately follows that $\mathrm{reg}(V(g)) \geq n - \deg(g) \geq n - n^\beta$, and so, by definition $\exists x^\kappa \in \mathrm{SM}(V(g))$ such that $\deg(x^\kappa) = n - n^\beta$. Due to the fact that $\mathrm{SM}(V(g))$ is a dual ideal, every divisor of x^κ is also a member of $\mathrm{SM}(V(g))$. As there are precisely 2^{n-n^β} such divisors we have

$$|V| = |V(g)| = |\mathrm{SM}(V(g))| \geq 2^{n-n^\beta},$$

and so V has density $\rho \geq 2^{-n^\beta}$. The result then follows immediately from Theorem 1.

□

3.2 AC^0-pseudorandomness

3.2.1 The Linial-Nisan-Braverman Technique

Braverman's Theorem

Braverman [Bra09] resolved the long standing Linial-Nisan conjecture [LN90]. We now state this theorem, which provides a simple sufficient condition for a distribution to appear random to AC^0 circuits. For a distribution μ_n with support $\{0,1\}^n$, we say that μ_n is a (β, r)-approximation if every restriction of μ_n to r coordinates is β-close to the uniform distribution on $\{0,1\}^r$ (two distributions are β-close if the statistical distance between them is at most β). The theorem states that if a distribution μ_n is a $(\beta, r(s,d,\epsilon))$-approximation, for sufficiently large r and sufficiently small β, then it ϵ-fools all depth d AC^0 circuits of size s.

Theorem. *[Bra09] Every $(\beta, r(s,d,\epsilon))$-approximation ϵ-fools all depth d AC^0 circuits of size s, where*

$$r(s,d,\epsilon) = \left(\log \frac{s}{\epsilon}\right)^{O(d^2)}$$

and

$$\frac{\epsilon}{\beta} > 2n^{r(s,d,\epsilon)}.$$

In particular, every $(2^{-n^\gamma}, n^\delta)$-approximation, for constants $\kappa < \delta < \gamma < 1$, will 2^{-n^κ}-fool polynomial sized circuits of any constant depth, for sufficiently small constant α. In this thesis, any function f for which the corresponding distribution μ_n, as defined above, meets this condition, will be said to have the Linial-Nisan-Braverman property, or LNB property for short. In fact, many of the functions considered will have an even stronger property: their corresponding distributions will be $(0, n^\delta)$-approximations (or, in other words, every restriction of μ_n to n^δ coordinates will simply be the uniform distribution, rather than being only close to the uniform distribution).

Application to Homomorphisms

Let us now restrict our attention to homomorphisms from $\{0,1\}^m$ to $\{0,1\}^k$, the set of which we denote by $\mathrm{Hom}(\{0,1\}^m, \{0,1\}^k)$ (or, in other words, viewing $\{0,1\}^m$ and $\{0,1\}^k$ as vector spaces, we consider the set of linear maps). It will be shown that it is particularly simple to determine if a given homomorphism has the Linial-Nisan-Braverman property, and, moreover, that many homomorphisms have this property, and hence appear random to AC^0 circuits.

Every $f \in \mathrm{Hom}(\{0,1\}^m, \{0,1\}^k)$ corresponds to a $k \times m$ matrix F, with entries in $\{0,1\}$, such that $f(X) = FX$, for $X \in \{0,1\}^m$. For any $R \subseteq \{1,\ldots,k\}$ and $C \subseteq \{1,\ldots,m\}$, let $F_{R,C}$ be the submatrix of F consisting of rows R and columns C. The following lemma shows that having the Linial-Nisan-Braverman property is equivalent to certain submatrices of F being full rank. As before, $n = m + k$.

Lemma 10. *$f \in Hom(\{0,1\}^m, \{0,1\}^k)$ has the Linial-Nisan-Braverman property if and only if $\exists \delta > 0$ such that $\forall R \subseteq \{1,\ldots,k\}, C \subseteq \{1,\ldots,m\}$ with $|R| + |C| = n^\delta$, the submatrix $F_{R,\overline{C}}$ is full rank, where $\overline{C} = \{1,\ldots,m\} \setminus C$.*

Proof. First, consider a function $f : \{0,1\}^m \to \{0,1\}^k$ whose corresponding matrix F meets the above condition. We show that f has the Linial-Nisan-Braverman property. To do this, let $X \in \{0,1\}^m$ be an arbitrary element, $Y \in \{0,1\}^n$ be the concatenation of X and $f(X)$, and μ_n be the distribution of Y given a uniformly randomly selected X. By definition, f has the Linial-Nisan-Braverman property if μ_n is n^δ-independent. To see that f has this property, imagine that an adversary selects some n^δ sized subset of coordinates of Y. We must show that the distribution μ_n, when restricted to these coordinates is the uniform distribution. Each coordinate is either a coordinate of the input X or a coordinate of the output $f(X)$. Of course, since X is selected uniformly at random, any such restriction on just the bits of X yields the uniform distribution. All that needs to be shown is that the conditional distribution of selected output coordinates is uniform, given any value of the selected input

coordinates, or, in other words, that if the adversary is allowed to look at only a small number of input bits (fewer than n^δ) than the distribution of any small number of output bits due to the remaining inputs bits is still uniform. To see this, let $R \subseteq \{1, \ldots, k\}$ and $C \subseteq \{1, \ldots, m\}$ denote the selected coordinates of $f(X)$ and X, respectively, where $|R| + |C| = n^\delta$. Letting $f(X)_R$ denote the bits of the output corresponding to R (that is to say, the selected bits of the output), and defining X_C and $X_{\overline{C}}$ analogously (which are then the selected and unselected bits of the input, respectively), then we can write $f(X)_R = F_{R,C}X_C + F_{R,\overline{C}}X_{\overline{C}}$. Since F meets the above condition, we know that $F_{R,\overline{C}}$ is full rank, and so, as all of the (unseen) bits of $X_{\overline{C}}$ vary uniformly, $F_{R,\overline{C}}X_{\overline{C}}$ varies uniformly. One way to see this is to note that, since $F_{R,\overline{C}}$ is full rank, it contains a $|R| \times |R|$ invertible submatrix. Therefore, as the bits of $X_{\overline{C}}$ that correspond to this invertible submatrix vary over all possible values (with the other bits of $X_{\overline{C}}$ fixed), $F_{R,\overline{C}}X_{\overline{C}}$ indeed varies uniformly. Therefore, for any fixed X_C, $f(X)_R$ varies uniformly, and so f has the Linial-Nisan-Braverman property.

To prove the other direction, assume that F doesn't meet the above condition. This means that, $\forall \delta > 0$, $\exists R \subseteq \{1, \ldots, k\}, C \subseteq \{1, \ldots, m\}$ with $|R| + |C| = n^\delta$ the submatrix $F_{R,\overline{C}}$ is not full rank. Again, we write $f(X)_R = F_{R,C}X_C + F_{R,\overline{C}}X_{\overline{C}}$. Since $F_{R,\overline{C}}$ is not full rank, we have, by definition, that as $X_{\overline{C}}$ varies $F_{R,\overline{C}}X_{\overline{C}}$ doesn't even hit all possible values. In fact, it must miss at least half of all values, and so $f(X)_R$ is far from uniformly randomly distributed for any fixed X_C.

□

Using the above result, we are now able to prove Theorem 2, which states that for any "reasonable" choice of m and k, almost every $f \in \mathrm{Hom}(\{0,1\}^m, \{0,1\}^k)$ is AC^0-pseudorandom. For convenience, we restate the theorem here.

Theorem 2. *If $k = m^u$, for any fixed constant $u > 0$, then all but an exponentially small fraction of all $f \in Hom(\{0,1\}^m, \{0,1\}^k)$ are AC^0-pseudorandom.*

Proof. Let $P_{h,w}$ denote the probability that an $h \times w$ matrix, where $w \geq h$, with entries

drawn uniformly at random from $\{0,1\}$, is full rank (that is to say, has rank h). We have the following useful bound, which follows from the fact that, in order for the matrix not to be full rank, either the first row must be identically zero, or the second row is a multiple of the first, or, in general, the i^{th} row lies in the span of the first $i-1$ rows; combining these probabilities with a union bound gives:

$$P_{h,w} \geq 1 - 2^{-w} \sum_{i=1}^{h} 2^{i-1}.$$

For any particular m, k, the probability that a randomly selected $f \in \mathrm{Hom}(\{0,1\}^m, \{0,1\}^k)$ is AC^0-pseudorandom is, by the above theorem, given by the probability that all appropriately sized submatrices of a randomly selected $k \times m$ matrix are full rank. To be precise, we are interested in the probability that all submatrices $F_{R,\overline{C}}$, where $|R| + |C| = n^\delta$ are full rank, when $m, k \gg n^\delta$. For any $h \leq k$ and $w \leq m$, the number of $h \times k$ submatrices of a $k \times m$ matrix is given by $\binom{k}{h}\binom{m}{w}$, and so, by a simple union bound, we have the following:

$$\Pr(f \text{ doesn't have the LNB property}) \leq \sum_{j=1}^{n^\delta - 1} \binom{k}{j} \binom{m}{m - (n^\delta - j)} (1 - P_{j,m-(n^\delta - j)})$$

$$\leq n^\delta \binom{k}{n^\delta} \binom{m}{n^\delta} 2^{-(m-n^\delta)} \sum_{i=1}^{n^\delta} 2^{i-1}$$

$$\leq n^\delta \frac{k^{n^\delta}}{(n^\delta)!} \frac{m^{n^\delta}}{(n^\delta)!} 2^{-(m-n^\delta)} (2^{n^\delta+1} - 1)$$

$$\leq \frac{(km)^{n^\delta}}{2^m}$$

$$= \frac{(m^{u+1})^{n^\delta}}{2^m}$$

□

Convolution

In the previous section, it was shown that many functions in $\mathrm{Hom}(\{0,1\}^m, \{0,1\}^k)$ appear random to AC^0 circuits, but no explicit example of such a function was given. This section shows that a particular function, namely the convolution function, satisfies this property. We begin by recalling the definition of convolution. Given some $X \in \{0,1\}^r$ and $Y \in \{0,1\}^s$, the convolution of X and Y, which will be denoted $X * Y$, is the $Z \in \{0,1\}^{r+s-1}$ where if X_i, Y_i, and Z_i refer to the i^{th} bit (zero indexed, counting from the least significant bit up) of X, Y, Z, respectively, then

$$Z_i = \sum_{j=0}^{i} X_j Y_{i-j},$$

where $X_j Y_{i-j}$ denotes the AND of X_j and Y_{i-j}, any X_j or Y_j outside of the defined range is understood to be zero, and the sum is, of course, computed modulo 2.

The goal is to show that convolution is AC^0-pseudorandom. There are several reasonable ways to define this. Perhaps the most natural, immediate thought is to consider the function $f : \{0,1\}^r \times \{0,1\}^s \to \{0,1\}^{r+s-1}$, which takes the pair (X, Y) to $X * Y$. Unpacking definitions, this means we consider the distribution (when X and Y are selected uniformly at random) of the string in $\{0,1\}^{2r+2s-1}$ where the first r bits are X, the next s bits are Y, and the final $r+s-1$ bits are $X * Y$. Observe that this distribution clearly does not look random to AC^0 circuits because some of the bits of $X * Y$ can be determined exactly by an AC^0 circuit. To be precise, letting n denote, as usual, the total size of the string ($n = 2r+2s-1$), we see that any of the first (or last) $O(\log^c n)$ bits of $X * Y$ is simply the parity of $O(\log^c n)$ bits, each of which is the AND of some bit of X with some bit of Y. Since a parity of $O(\log^c n)$ bits can (for any constant c) be computed easily in AC^0, we immediately conclude that including any of these bits will cause the resulting distribution to not appear random to AC^0 circuits. However, if we exclude these bits, we can show that the remainder does appear random to AC^0 circuits. We consider the function $\mathrm{CONV}_{r,s,k} : \{0,1\}^r \times \{0,1\}^s \to \{0,1\}^k$ where now $k = k(r,s) < r + s - 1$, and only the k "middle bits" of $X * Y$ are included (the

k centermost bits). Such a function is not a homomorphism and so the technique of the previous section does not directly apply. Instead, we will consider a variant of this problem, to which that technique does apply. Doing so yields a stronger result that also immediately implies that $\text{CONV}_{r,s,k}$ function does, in fact, appear random to AC^0 circuits.

Essentially, the idea is to consider a "fixed" Y (here we mean that there is a single fixed Y of each length; as mentioned earlier, the discussion involves the asymptotic properties of f, defined by a sequence of Y values, one for each length), and define the function $f_Y : \{0,1\}^r \rightarrow \{0,1\}^k$, (where again, as above, $k = k(r) < r + s - 1$) such that f_Y takes the r-bit value X to the middle k bits of $X * Y$. The difference between these two variants can be understood as follows. In the first variant, described in the previous paragraph, the distinguisher would be an AC^0 circuit family where the circuit whose input size is $r + s + k$ would be able to distinguish the string consisting of a uniformly randomly selected $X \in \{0,1\}^r$, a uniformly randomly selected $Y \in \{0,1\}^s$ and the middle $k(r,s)$ bits of $X * Y$ from a truly random string. In the second variant, the distinguisher can have Y built-in, and only needs to distinguish the string consisting of a uniformly randomly selected $X \in \{0,1\}^r$ and the middle $k(r)$ bits of $X * Y$ from a truly random string.

Since each f_Y is clearly a homomorphism, Lemma 10 applies. Moreover, if it can be shown that, for all sufficiently large r, all but an exponentially small fraction of choices for Y produce an f_Y that is AC^0-pseudorandom, then it immediately follows that the variant of the problem described in the previous paragraph, in which both X and Y are selected uniformly at random, also is AC^0-pseudorandom. Loosely speaking, claiming that this second variant is AC^0-pseudorandom is a stronger claim because being able to have a separate circuit for each Y could conceivably give a distinguisher more power.

We now prove Theorem 3, which is restated below.

Theorem 3. *If $s = r^u$ and $k = r + s - (MIN(r,s))^\alpha$, for any fixed constants $u > 0$ and $0 < \alpha < 1$, then $CONV_{r,s,k}$ is AC^0-pseudorandom. In particular, if $r = s$ and $k = 2r - r^\alpha$, for any $0 < \alpha < 1$, then $CONV_{r,s,k}$ is AC^0-pseudorandom.*

By the above logic, it suffices to show the following lemma.

Lemma 11. *For all but an exponentially small fraction of Y, the function $f_Y : \{0,1\}^r \to \{0,1\}^k$, where $k = 2(r - r^\alpha + 1)$ for any small constant $\alpha > 0$, has the Linial-Nisan-Braverman property.*

Proof. Let f denote an arbitrary element of the set $\{f_Y | Y \in \{0,1\}^s\}$. Since f is a homomorphism, there is a corresponding $k \times r$ matrix F such that $f(X) = FX$, for any $X \in \{0,1\}^r$. To show that, for almost all choices of Y, the corresponding function f has the Linial-Nisan-Braverman property, it suffices, by Lemma 10, to show that the appropriate submatrices of F are full rank.

The matrix F has a particularly simple structure, namely it has constant skew-diagonals. That is to say, if $F_{i,j}$ denotes the element of F in row i and column j then $F_{i,j} = F_{i-1,j+1}$. The first row of F consists of, from left to right, $r - r^\alpha$ zeros followed by the lowest r^α bits of Y, starting with the least significant bit of Y. Each subsequent row of F is obtained by shifting Y one index further to the left, filling empty entries with zeros. Consider an arbitrary submatrix $F_{R,\overline{C}}$ where $R \subseteq \{1, \ldots, k\}$ and $C \subseteq \{1, \ldots, r\}$ such that $|R| + |C| = n^\delta$ for $\delta < \alpha$, where $\overline{C} = \{1, \ldots, r\} \setminus C$ and $n = r + k$. For randomly selected Y, this submatrix is full rank with overwhelming probability. To see this, note that if $F_{R,\overline{C}}$ is not full rank, then there is some non-trivial linear combination of its rows that adds to 0. Let h and w be the height and width, respectively, of $F_{R,\overline{C}}$. Then there are $2^h - 1$ potential non-trivial linear combinations of the rows, because a linear combination is, by definition, a sum of the rows of $F_{R,\overline{C}}$ where each row has coefficient 0 or 1 (having all coefficients be 0 is the trivial linear combination). In other words, it is a sum of some subset of the rows of $F_{R,\overline{C}}$. Consider any fixed non-trivial linear combination. Let i denote the lowest row of $F_{R,\overline{C}}$ that has coefficient 1. Note that the probability (over Y) that this particular linear combination of the rows of $F_{R,\overline{C}}$ is zero is very small. While this fact would be immediate if $F_{R,\overline{C}}$ were simply a random unstructured matrix, some care must be given due to the structure of F (constant skew-diagonals) which forces all elements of F in the same skew-

diagonal to be identical. To deal with this, consider the rows of $F_{R,\overline{C}}$ one at a time, from left to right. In order for the linear combination of the rows to be the zero vector, it must be the case, by definition, that the sum in each column is zero (where of course this sum is only over the subset of elements selected by the linear combination). Consider the element in position (i,j). This element is either some element of Y, if some part of Y was shifted over position (i,j), or is simply 0, if no part of Y was shifted to that position. In the first case, this value is completely independent of any previously considered entries that influence the linear combination. This is because, even though the value of the entry in position (i,j) forces the values of all other entries in the same skew-diagonal (in F), all other such entries are either to the right of this entry, and so haven't been considered yet, or to the left and below this entry, in which case they have coefficient 0 in the linear combination (because row i is the lowest row with coefficient 1). Since row i has coefficient 1, flipping the value of the element in position (i,j) flips the value of the sum in column j, and so the sum in this column is 0 with probability $\frac{1}{2}$. From this, we immediately conclude that the probability that the sum in all columns is 0 is 2^{-z}, where z is the number of entries in row i that come from Y (as opposed to being fixed 0s). Since each row of F has at least r^α such elements (because the output of f does not include the first or last r^α bits of $X * Y$), we conclude that this particular linear combination is 0 with probability at most $2^{-r^{-\alpha}}$. Applying a union bound over all $2^h - 1$ non-trivial linear combinations, where $h < n^\delta \ll r^\alpha$, and then another union bound over all choices of R and C (as in the calculation in the previous section), we conclude that, for all but an exponentially small fraction of Y, F has the desired property, which completes the proof that convolution appears random to AC^0 circuits.

□

Integer Multiplication

Let $\text{MULT}_{r,s,k} : \{0,1\}^r \times \{0,1\}^s \to \{0,1\}^k$ denote the integer multiplication function, which takes a $X \in \{0,1\}^r$ and $Y \in \{0,1\}^s$ to the middle k-bits of the $r+s$ bit long product of X

and Y. In this section, we will prove the following theorem.

Theorem 4. *If $s = r^u$ and $k = r + s - (MIN(r,s))^\alpha$, for any fixed constants $u > 0$ and $0 < \alpha < 1$, then $MULT_{r,s,k}$ is AC^0-pseudorandom. In particular, if $r = s$ and $k = 2r - r^\alpha$, for any $0 < \alpha < 1$, then $MULT_{r,s,k}$ is AC^0-pseudorandom.*

As was the case for convolution, there are two natural variants of the multiplication problem to consider. In the first variant, we select $X \in \{0,1\}^r$ and $Y \in \{0,1\}^s$ uniformly at random, then produce the product $P = X \times Y$, and finally we produce the string consisting of X, Y, and part of P. The hope is that the distribution of that string appears random to AC^0 circuits. It is necessary to include only part of P because, as was the case in convolution, the lowest and highest bits of P do not look random to AC^0 circuits. For example, the low $O(\log^c r)$ bits of the product can be calculated exactly, using the technique in [CSV84]. In the second variant, we consider "fixed" Y, in the sense that we have a single Y of each length, and the multiplication problem is defined such that a uniformly randomly selected $X \in \{0,1\}^r$ is multiplied by the fixed Y to produce the product $P = X \times Y$; the string of interest then consists of X and the middle part of P. Again, loosely speaking, the second variant is stronger in the sense that a potential distinguisher is allowed to have Y built-in.

In this section, we focus on the second variant and show that, for sufficiently large r, all but an exponentially small fraction of Y (of length s) lead to a multiplication problem that looks random to AC^0 circuits. Therefore, by the same logic as in the convolution problem, it immediately follows that the first variant is also AC^0-pseudorandom. We consider the function $f_Y : \{0,1\}^r \to \{0,1\}^k$, which takes the r-bit value X to the middle k bits of the product $X \times Y$. We will prove the following lemma, from which the above theorem immediately follows.

Lemma 12. *For all but an exponentially small fraction of $Y \in \{0,1\}^s$, where $s = r^u$, the function $f_Y : \{0,1\}^r \to \{0,1\}^k$, where $k = r + s - 2r^\alpha$ for any small constant $\alpha > 0$, has the Linial-Nisan-Braverman property.*

Proof. It suffices to establish the claim for almost all odd Y (because adding w trailing zeros to Y simply shifts the product $X \times Y$ by w bits to the left; all but an exponentially small fraction of Y have fewer than r^α trailing zeros), and so we restrict our attention to the case in which Y is odd. We begin by establishing some notation. Let $n = r + k$. Let $Z = Z_1 \cdots Z_n$ be the distribution of the set of all strings of the form $X \circ f_Y(X)$ (strings that are the concatenation of X with $f_Y(X)$), where X is an r-bit string. Then, by definition, f_Y has the Linial-Nisan-Braverman property if Z is a $(2^{-n^\gamma}, n^\delta)$-approximation for appropriate small constants $0 < \delta < \gamma < 1$, which is to say that, for every set of n^δ coordinates the restriction of μ_n to those coordinates is 2^{-n^γ}-close to the uniform distribution over $\{0,1\}^{n^\delta}$. To show this, we begin by recalling that the bias of a distribution Z on some set $I \subseteq \{1, \ldots, n\}$ is defined to be

$$\text{bias}_I(Z) = \text{E}[(-1)^{\sum_{i \in I} Z_i}].$$

We make use of the following lemma, variants of which appeared in, for example [Vaz86] and [AGM02].

Lemma 13. *[Vaz86], [AGM02] Every distribution Z that has bias as most ϵ on every non-empty subset I of size at most h is a $(2^{h/2}\epsilon, h)$-approximation.*

We will then show that Z has bias at most 2^{-n^ν}, for some constant $\nu > 0$, on all non-empty sets of size at most n^δ. The above lemma implies that Z is a $(2^{-n^\gamma}, n^\delta)$-approximation, as desired (for any $\delta < \gamma < \nu$). To see why, let X_i denote the i^{th} bit of X and let $f_{Y,j} : \{0,1\}^r \to \{-1,1\}$ be defined such that $f_{Y,j}(X) = 1$ when the j^{th} bit of $X \times Y$ is 0 and $f_{Y,j}(X) = -1$ when the j^{th} bit of $X \times Y$ is 1 (note that $f_{Y,j}$ corresponds to the j^{th} bit of $X \times Y$ not the j^{th} bit of $f_Y(X)$, where $f_Y(X)$ consists of all bits of $X \times Y$ except the lowest and highest r^α; this is done because it will be much cleaner to refer to bits by their position in the entire product). Clearly,

$$f_{Y,j}(X) = (-1)^{\lfloor \frac{XY}{2^{j-1}} \rfloor}.$$

For any $S \subseteq \{1, \ldots, r\}$, let $\hat{f}_{Y,j}(S)$ denote the Fourier-Walsh coefficients of $f_{Y,j}$, which are given by

$$\hat{f}_{Y,j}(S) = \mathrm{E}[f_{Y,j}(X)(-1)^{\sum_{i \in S} X_i}].$$

These are the Fourier coefficients of a function on $\mathbb{F}_2^r$ (we use the term Fourier-Walsh to avoid confusion with the "ordinary" Fourier coefficients of a function defined on $\mathbb{R}$, which will be used shortly). We partition the set I as $I = S \cup J$, where $S \subseteq \{1, \ldots, r\}$ are the indices of Z that correspond to bits of X and $J \subseteq \{r+1, \ldots, n\}$ are the indices of Z that correspond to bits of $f_Y(X)$.

There are two cases. First, if J is empty, then the set I consists only of bits of X, and so, trivially, Z has bias exactly 0 on this set, because X is uniformly random. The interesting case is when J is non-empty. For notational convenience, define the set $J' \subseteq \{r^\alpha + 1, \ldots, r + s - r^\alpha\}$ such that $J' = \{j' | j' + r - r^\alpha \in J\}$ (simply the set J shifted appropriately to index bits of $X \times Y$). Let $f_{Y,J'}(X) = \prod_{j \in J'} f_{Y,j}(X)$. Then the bias of Z on I is simply $\hat{f}_{Y,J'}(S)$. This follows from the fact that

$$\begin{aligned}
\mathrm{bias}_I(Z) &= \mathrm{E}[(-1)^{\sum_{i \in I} Z_i}] \\
&= \Pr[\oplus_{i \in I} Z_i = 0] - \Pr[\oplus_{i \in I} Z_i = 1] \\
&= \Pr[\oplus_{s \in S} Z_s = \oplus_{j \in J} Z_j] - \Pr[\oplus_{s \in S} Z_s \neq \oplus_{j \in J} Z_j] \\
&= \Pr[(-1)^{\sum_{s \in S} X_s} = f_{Y,J'}(X)] - \Pr[(-1)^{\sum_{s \in S} X_s} \neq f_{Y,J'}(X)] \\
&= \Pr[(-1)^{\sum_{s \in S} X_s} f_{Y,J'}(X) = 1] - \Pr[(-1)^{\sum_{s \in S} X_s} f_{Y,J'}(X) = -1] \\
&= \mathrm{E}[f_{Y,J'}(X)(-1)^{\sum_{s \in S} X_s}] \\
&= \hat{f}_{Y,J'}(S).
\end{aligned}$$

Rather than compute $\hat{f}_{Y,J'}(S)$ directly, we instead compute the Fourier coefficients of $f_{Y,J'}$ when viewed as a function on $\{0, \ldots, 2^r - 1\}$ (instead of on $\mathbb{F}_2^r$), and exploit a connection

between these two types of Fourier coefficients. For a function $f : \{0, \ldots, 2^r - 1\} \to \{-1, 1\}$, define

$$\hat{f}(k) = \mathrm{E}[f(t)e^{-\frac{2\pi ikt}{2^r}}],$$

where $k \in \mathbb{Z}$. We have the following lemma, from [Gre12] (see also [Kat86]), which has been modified to fit our notation. We say that an integer k is a (b, m)-sparse number if it can be written in the form $k = k_1 2^{h_1} + \cdots + k_b 2^{h_b}$ where each $k_i \in \mathbb{Z}$, $|k_i| \leq m$, $h_i \in \mathbb{N}$.

Lemma 14. *Let $f : \{0, \ldots, 2^r - 1\} \to \{-1, 1\}$ be a function such that $\exists S \subseteq \{1, \ldots, r\}$ with Fourier-Walsh coefficient $\hat{f}(S)$ of magnitude at least ϵ, where $0 < \epsilon < \frac{1}{2}$. Then there is a $\left(|S|, \left(\frac{10|S|}{\epsilon}\right)\right)$-sparse number k such that the Fourier coefficient $\hat{f}(k)$ has magnitude at least $\left(\frac{\epsilon}{10|S|}\right)^{4|S|}$.*

Applying this lemma to the function $f_{Y,J'}$, with sets S of size at most n^δ, we immediately conclude that, in order to establish the necessary bounds on the Fourier-Walsh coefficients (which then implies that multiplication has the Linial-Nisan-Braverman property), it suffices to show that, for all $(n^\delta, 10n^\delta 2^{n^\nu})$-sparse numbers k, $|\hat{f}_{Y,J'}(k)| < 2^{-n^\rho}$ for a fixed constant ρ such that $\rho > \delta + \nu$. We say that a particular Fourier component is negligible if its magnitude has such a bound.

We now show that, for almost all Y, the required bound on $\hat{f}_{Y,J'}(k)$ holds. The main idea is that, for each j, $f_{Y,j}$ is simply a downsampled version of a square wave. This fact allows us to express the Fourier coefficients of $f_{Y,j}$ in terms of the Fourier coefficients of a square wave. This is useful because the Fourier coefficients of a square wave are particularly simple. In the following, we make use of several standard facts about the Discrete Fourier Transform, which can be found in essentially any text that deal with Fourier Analysis, for example [OSB99]. We begin with a few definitions. Let $D_Y = \{0, \ldots, Y2^{r+s} - 1\}$. Let $s_j : D_Y \to \{-1, 1\}$ be the perfect square wave of period 2^j,

$$s_j(t) = (-1)^{\lfloor \frac{t}{2^{j-1}} \rfloor}.$$

Let $p_Y : D_Y \to \{0,1\}$ be a pulse train with interval Y,

$$p_Y(t) = \begin{cases} 1, & t \equiv 0 \mod Y \\ 0, & t \not\equiv 0 \mod Y \end{cases}.$$

Let $h_Y(t) : D_Y \to \{0,1\}$ be the step function

$$h_Y(t) = \begin{cases} 1, & t < Y2^r \\ 0, & t \geq Y2^r \end{cases}.$$

Finally, let $g_{Y,J'}(t) = Y2^s h_Y(t) p_Y(t) \prod_{j \in J'} s_j(t)$.

We then have

$$\begin{aligned}
\hat{f}_{Y,J'}(k) &= \frac{1}{2^r} \sum_{t=0}^{2^r-1} f_{Y,J'}(t) e^{-\frac{2\pi i k t}{2^r}} \\
&= \frac{1}{2^r} \sum_{t=0}^{2^r-1} \left(\prod_{j \in J'} (-1)^{\lfloor \frac{Yt}{2^{j-1}} \rfloor} \right) e^{-\frac{2\pi i k t}{2^r}} \\
&= \frac{1}{2^r} \sum_{t=0}^{Y2^r-1} p_Y(t) \left(\prod_{j \in J'} (-1)^{\lfloor \frac{t}{2^{j-1}} \rfloor} \right) e^{-\frac{2\pi i k t}{Y2^r}} \\
&= \frac{1}{2^r} \sum_{t=0}^{Y2^r-1} p_Y(t) \left(\prod_{j \in J'} s_j(t) \right) e^{-\frac{2\pi i k t}{Y2^r}} \\
&= \frac{1}{Y2^{r+s}} \sum_{t=0}^{Y2^{r+s}-1} Y2^s h_Y(t) p_Y(t) \left(\prod_{j \in J'} s_j(t) \right) e^{-\frac{2\pi i 2^s k t}{Y2^{r+s}}} \\
&= \frac{1}{Y2^{r+s}} \sum_{t=0}^{Y2^{r+s}-1} g_{Y,J'}(t) e^{-\frac{2\pi i 2^s k t}{Y2^{r+s}}} \\
&= \hat{g}_{Y,J'}(2^s k).
\end{aligned}$$

Therefore, it suffices to show that $\hat{g}_{Y,J'}(2^s k)$ is sufficiently small for the k values of

interest. The convolution theorem implies that

$$\hat{g}_{Y,J'}(k) = Y2^s \hat{h}_Y(k) \otimes \hat{p}_Y(k) \otimes \bigotimes_{j \in J'} \hat{s}_j(k),$$

where $\otimes$ denotes cyclic convolution.

Notice that, for each j, $\hat{s}_j(k)$ has a particularly simply structure.

$$\hat{s}_j(k) = \begin{cases} \frac{1}{2^{j-2}\left(1-e^{-\frac{2\pi i(2v+1)}{2^j}}\right)}, & k = (2v+1)Y2^{r+s-j} \\ 0, & \text{otherwise} \end{cases}.$$

Notice that $\hat{s}_j(k)$ is only nonzero at few locations; specifically, the odd multiples of $Y2^{r+s-j}$. Moreover, notice that the magnitude of the nonzero values falls off quickly. To be precise,

$$\sum_{\substack{v \\ |2v+1|>2^{n^\tau}}} |\hat{s}_j((2v+1)Y2^{r+s-j})| = O(2^{-n^\eta})$$

for constants η and τ such that $\delta < \eta < \tau \ll 1$. In other words, the only non-negligible part of $\hat{s}_j(k)$ is at values k given by small odd multiples of a shift of Y (Y shifted to the left by $r+s-j$ bits).

We then consider $\bigotimes_{j \in J'} \hat{s}_j(k)$. We split $\hat{s}_j(k)$ into a large low frequency component and a small high frequency component. That is to say, we write $\hat{s}_j(k) = \hat{u}_j(k) + \hat{v}_j(k)$, where

$$\hat{u}_j(k) = \begin{cases} \frac{1}{2^{j-2}\left(1-e^{-\frac{2\pi i(2v+1)}{2^j}}\right)}, & k = (2v+1)Y2^{r+s-j}, |2v+1| \leq 2^{n^\tau} \\ 0, & \text{otherwise} \end{cases}.$$

and

$$\hat{v}_j(k) = \begin{cases} \frac{1}{2^{j-2}\left(1-e^{-\frac{2\pi i(2v+1)}{2^j}}\right)}, & k = (2v+1)Y2^{r+s-j}, |2v+1| > 2^{n^\tau} \\ 0, & \text{otherwise} \end{cases}.$$

Therefore,

$$\bigotimes_{j\in J'} \hat{s}_j(k) = \bigotimes_{j\in J'} (\hat{u}_j(k) + \hat{v}_j(k))$$

$$= \sum_{\substack{J_1,J_2 \\ J_1\cup J_2 = J'}} \left(\bigotimes_{j\in J_1} \hat{u}_j(k)\right) \otimes \left(\bigotimes_{j\in J_2} \hat{v}_j(k)\right).$$

Notice that there are at most 2^{n^δ} terms in the above expansion (because $|J'| \leq n^\delta$). The term $\bigotimes_{j\in J'} \hat{u}_j(k)$ is only nonzero at k values of the form $(2v_1+1)Y2^{r+s-j_1} + \cdots + (2v_{|J'|}+1)Y2^{r+s-j_{|J'|}}$, where each v_i satisfies $|2v_i+1| \leq 2^{n^\tau}$. All other terms are extremely small everywhere. To be precise, when $J_1 \neq J'$, every such term involves at least one $\hat{v}_j(k)$ factor and so we can write $\left(\bigotimes_{j\in J_1} \hat{u}_j(k)\right) \otimes \left(\bigotimes_{j\in J_2} \hat{v}_j(k)\right) = \hat{v}_j(k) \otimes \hat{q}(k)$ for some function $q : D_Y \to \{-1, 0, 1\}$. By combining the bound $\sum_k |\hat{v}_j(k)| = O(2^{-n^\eta})$ with the trivial bound $|\hat{q}(k)| \leq 1$, we obtain $\left|\left(\bigotimes_{j\in J_1} \hat{u}_j(k)\right) \otimes \left(\bigotimes_{j\in J_2} \hat{v}_j(k)\right)\right| = O(2^{-n^\eta})$. Therefore, the total contribution of all terms except $\bigotimes_{j\in J'} \hat{u}_j(k)$ is negligible ($O(2^{-(n^\eta - n^\delta)})$). From the above, it is immediate that the only non-negligible Fourier components are values k of the form $(2v_1+1)Y2^{r+s-j_1} + \cdots + (2v_{|J'|}+1)Y2^{r+s-j_{|J'|}}$, where each v_i satisfies $|2v_i+1| \leq 2^{n^\tau}$. Recall that each j satisfies $r^\alpha < j < r+s-r^\alpha$. Therefore, these values k are of the form Yk', where k' is a $(|J'|, 2^{n^\tau})$-sparse number with at least r^α trailing zeros and at most $r+s-r^\alpha$ trailing zeros.

Next, we consider $\hat{g}_{Y,J'}(k)$. We have

$$\hat{g}_{Y,J'}(k) = Y2^s \hat{h}_Y(k) \otimes \hat{p}_Y(k) \otimes \bigotimes_{j\in J'} \hat{s}_j(k)$$

$$= Y2^s \hat{h}_Y(k) \otimes \hat{p}_Y(k) \otimes \bigotimes_{j\in J'} (\hat{u}_j(k) + \hat{v}_j(k))$$

$$= Y2^s \hat{h}_Y(k) \otimes \hat{p}_Y(k) \otimes \sum_{\substack{J_1,J_2 \\ J_1 \cup J_2 = J'}} \left(\bigotimes_{j \in J_1} \hat{u}_j(k) \right) \otimes \left(\bigotimes_{j \in J_2} \hat{v}_j(k) \right)$$

$$= \sum_{\substack{J_1,J_2 \\ J_1 \cup J_2 = J'}} Y2^s \hat{h}_Y(k) \otimes \hat{p}_Y(k) \otimes \left(\bigotimes_{j \in J_1} \hat{u}_j(k) \right) \otimes \left(\bigotimes_{j \in J_2} \hat{v}_j(k) \right).$$

By the same logic as above, the total contribution of every term in the sum except the $J_1 = J'$ term is negligible everywhere (has total magnitude $O(2^{-(n^\eta - n^\delta)})$ at all k) and so if we define the function $\hat{g'}_{Y,J'}(k) = Y2^s \hat{h}_Y(k) \otimes \hat{p}_Y(k) \otimes \bigotimes_{j \in J'} \hat{u}_j(k)$, it suffices to show that $\hat{g'}_{Y,J'}$ is small at the k values of interest.

We have

$$\hat{p}_Y(k) = \begin{cases} 1, & k = u2^{r+s} \\ 0, & \text{otherwise} \end{cases}$$

and

$$Y2^s \hat{h}_Y(k) = \frac{1}{2^r} \frac{1 - e^{-\frac{2\pi i k}{2^s}}}{1 - e^{-\frac{2\pi i k}{Y 2^{r+s}}}}.$$

Therefore, the only non-negligible values of $\hat{g'}_{Y,J'}(k)$ are those that are "close" to values of the form $Yk' \mod 2^{r+s}$. More precisely, the only non-negligible values of $\hat{g'}_{Y,J'}$ are of the form $k \equiv Yk' + u \mod 2^{r+s}$, where $|u| \leq 2^{s+n^\nu}$, and so the only non-negligible values of $\hat{f}_{Y,J'}(k) = \hat{g}_{Y,J'}(2^s k)$ are at values k such that $2^s k \equiv Yk' + u \mod 2^{r+s}$. Or equivalently, values k where $\exists k', u'$ where k' is (as above) a $(|J'|, 2^{n^\tau})$-spare number with at least r^α trailing zeros and at most $r + s - r^\alpha$ trailing zeros, $|u'| \leq 2^{n^\nu}$ such that $k + u'$ is equal to the high 2^r bits of $Yk' \mod 2^{r+s}$.

Therefore, for a particular value Y, the required bound on $|\hat{f}_{Y,J'}(k)|$ holds if, for every $(n^\delta, 10n^\delta 2^{n^\nu})$-sparse number k, we do not have $k + u'$ equal to the high 2^r bits of Yk' mod 2^{r+s}, for any $(|J'|, 2^{n^\tau})$-sparse number k' with at least r^α trailing zeros and at most $r + s - r^\alpha$ trailing zeros. To see that this holds for all but an exponentially small fraction of Y, first notice that if k is a $(n^\delta, 10n^\delta 2^{n^\nu})$-sparse number, then $k + u'$ is a $(n^\delta + 1, 10n^\delta 2^{n^\nu})$-sparse number. Set the constants τ and ν small enough such that $n^{\nu+\tau} \ll r^\alpha$ (this can be

done because $n = 2r + s - 2r^\alpha = 2r + r^u - 2r^\alpha$ and so n is polynomial in r). Therefore, it suffices to show that, for almost all Y, if k' is a $(|J'|, 2^{n^\tau})$-sparse number with at least r^α trailing zeros and at most $r + s - r^\alpha$ trailing zeros then the high 2^r bits of $Yk' \mod 2^{r+s}$ is not a $(n^\delta + 1, 10n^\delta 2^{n^\nu})$-sparse number. To see this, notice that, for each pair of sparse numbers k, k'', there is at most a fraction $\frac{1}{2^{r^\alpha}}$ of all Y such that the high 2^r bits of $Yk' \mod 2^{r+s}$ are equal to k'' and so a simple union bound completes the proof.

□

Matrix Multiplication

We now show that matrix multiplication is AC^0-pseudorandom. Let MATRIX-MULT$_{r,s}$: $\{0,1\}^{rs} \times \{0,1\}^{rs} \to \{0,1\}^{s^2}$ denote the matrix multiplication function, which, on input a $s \times r$ matrix A and a $r \times s$ matrix B (both of which are encoded as strings in $\{0,1\}^{rs}$ in the obvious way), produces the $s \times s$ matrix AB.

Theorem 5. *If $s = r^u$, for any fixed constant $u > 0$, then MATRIX-MULT$_{r,s}$ is AC^0-pseudorandom.*

As was the case for the convolution and multiplication problems, we consider a stronger variant where one of the matrices is held fixed. We then prove the following lemma, from which the above theorem immediately follows.

Lemma 15. *For an $s \times r$ matrix A, let $f_A : \{0,1\}^{rs} \to \{0,1\}^{s^2}$ denote the function that, on input a $r \times s$ matrix B produces the $s \times s$ matrix $Z = AB$. Then all but an exponentially small fraction of A yield an f_A that is AC^0-pseudorandom.*

Proof. To see that almost all such f_A are AC^0-pseudorandom, let B_i and Z_i denote the i^{th} column of B and Z, respectively. Then, of course, $Z_i = AB_i$, and so we can interpret this problem as the concatenation of s independent instances of the homomorphism problem. That is to say, if we let $f'_A : \{0,1\}^r \to \{0,1\}^s$ be the homomorphism corresponding to A, then $Z_i = f'(B_i)$. The result then follows from Theorem 2.

□

3.2.2 The Reduction Technique

Next-Bit Test and Parity

In this section, another technique for proving that a function appears random to AC^0 circuits is presented, specifically, reducing a known hard problem to the next-bit test. The next-bit test is defined as follows. Given a distribution μ_n with support $\{0,1\}^n$, we say that μ_n passes the next-bit test if, given the first i bits of a string selected according to μ_n, no AC^0 circuit can predict the $(i+1)^{\text{th}}$ bit with non-negligible advantage, for any i. Formally, for any $Z \in \{0,1\}^n$, let Z_j denote the j^{th} bit of Z (1 indexed, counting from left to right) and $Z_{[j,k]}$ denote the substring of Z from positions j to k, inclusive. Then we say that μ_n passes the next-bit test if, for all $i \in \{1,\ldots,n\}$, and for all functions $Q_i : \{0,1\}^{i-1} \to \{0,1\}$ computable by AC^0 circuits, $|\Pr(Q_i(Z_{[1,i-1]}) = x_i) - \frac{1}{2}| = O(2^{-n^\kappa})$, for some constant $\kappa > 0$, where the probability is taken over values of $Z \in \{0,1\}^n$ drawn according to the distribution μ_n. It is known [Yao85] that a distribution μ_n passes the next-bit test if and only if μ_n $O(2^{-n^\kappa})$-fools all AC^0 circuits (strictly speaking, the result in [Yao85] was proven for probabilistic polynomial time algorithms, but the same technique applies just as well to AC^0 circuit families). Since, as stated in §1, we say that a function f is AC^0-pseudorandom if the distribution μ_n corresponding to it $O(2^{-n^\kappa})$-fools all AC^0 circuits, showing that μ_n passes the next-bit test is sufficient to prove the corresponding f is AC^0-pseudorandom.

The natural next question is how to prove that distributions arising from particular functions pass the next-bit test. One idea is to reduce a problem that is known to be hard for AC^0, such as the parity problem, to the next-bit test. The parity problem is defined as follows: given some $X \in \{0,1\}^*$, compute $\sum_i X_i \mod 2$. In other words, the parity of a string is 1 if there are an odd number of 1s in the string and 0 if there are an even number of 1s in the string. It is known that no AC^0 circuit family can compute parity [FSS84],[Ajt83]. In fact, parity can't even be non-negligibly approximated in AC^0 [Has86]. To be precise, if

we define $h(s,d,n)$ to be the function such that no depth d circuit of size 2^s computes parity correctly for more than a $\frac{1}{2} + h(s,d,n)$ fraction of the inputs, then we have the following (Theorem 8.1.iii in [Has86])

Theorem. *[Has86]* $h(s,d,n) < 2^{-\Omega((\frac{n}{s})^{\frac{1}{d-1}})}$ *for* $d > 2$ *and* $s < n^{\frac{1}{d}}$.

The goal is then to reduce the parity problem to the problem of computing the next bit of a string drawn according to μ_n, or, in other words, show that if some AC^0 circuit could predict the next bit with non-negligible advantage, then it could be used to produce another AC^0 circuit that approximates the parity problem, with non-negligible advantage. Since the parity problem cannot be approximated by such a circuit, we could then conclude that the original distribution must pass the next-bit test.

Integer Multiplication

As was already shown in Theorem 4, the function $\text{MULT}_{r,s,k}$ is AC^0-pseudorandom when $s = r^u$ and $k = r + s - (\text{MIN}(r,s))^\alpha$, for constants $u > 0$ and $0 < \alpha < 1$. This was done by considering a variant of the multiplication function in which one of the multiplicands is held fixed. Specifically, for $Y \in \{0,1\}^s$, we defined the function $f_Y : \{0,1\}^r \to \{0,1\}^k$ which takes a value $X \in \{0,1\}^r$ to the middle k bits of $X \times Y$. As shown in Lemma 12, f_Y is AC^0-pseudorandom for all but an exponentially small fraction of Y, when $s = r^u$ and $k = r + s - (\text{MIN}(r,s))^\alpha$. In this section, we will be interested in results that hold when s is much greater than r. Specifically, we are interested in the case when $s > r^u$ for all constants $u > 0$, but $r > \log^c s$ for all constants $c > 0$. Recall that we say a given function looks random to AC^0 circuits if the distribution corresponding to it can only be distinguished (by AC^0 circuits) from the uniform distribution with advantage $O(2^{-n^\kappa})$. In this section we relax this condition only slightly, and only require a bound on the advantage of the form $o(2^{-\log^c n})$ for all constants $c > 0$ (in other words, we require that no AC^0 circuit can distinguish with advantage one over any quasipolynomial in n). We show that, for certain Y, f_Y is AC^0 pseudorandom with these parameters. This has several interesting consequences. Firstly,

this yields a simple, multiplication based pseudorandom generator with the same stretch and security parameters as the Nisan-Wigderson generator [Nis91]. Secondly, this shows that the result in [CSV84], which states that an AC^0 circuit can multiply an n-bit value Y by a $O(\log^c n)$ bit value X is tight,

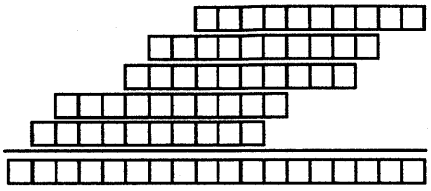
We restrict our attention to $Y \in \{0,1\}^s$ that are "sparse", in the sense that only a small number of the bits of Y are 1s. Specifically, we generate Y as follows: each bit is set to be 1 with probability $r^{-\epsilon}$, for a constant $0 < \epsilon < \frac{1}{2}$. As before, let $f_Y : \{0,1\}^r \to \{0,1\}^k$ be defined such that f_Y takes the value X to the middle k bits of the product $X \times Y$, where here $k = r + s - 2r^{2\epsilon}$. We prove the following theorem.

Theorem 6. *With high probability (where the probability is over the selection of Y according to the above distribution, and the statement high probability means within an exponentially small distance from probability 1), f_Y is AC^0-pseudorandom.*

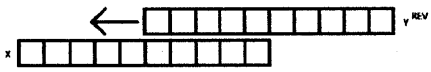
Proof. As usual, we consider strings of the form $X \circ f_Y(X)$. For convenience, we assume that both X and the substring of $Z = X \times Y$ produced by f_Y are written from least significant bit to most significant bit, when read from left to right. We let n denote the total length of the string, and so $n = 2r + s - 2r^{2\epsilon}$. Consider the next-bit test applied to strings generated in this manner. Since the first r bits of the string are bits of the uniformly randomly generated number X, we conclude, for information theoretic reasons, that there is no hope of any AC^0 circuit predicting the i^{th} bit, given the first $i-1$ bits, for $i \in \{1, \ldots, r\}$. All that remains is to prove the same claim for $i \in \{r+1, \ldots, n\}$, which will be done by showing that any AC^0 circuit that predicts such a bit with non-negligible advantage can be used to approximate the parity function, with non-negligible advantage, which we know is impossible. We assume, for contradiction, that we have an AC^0 circuit, call it C, that can predict some next-bit of our pseudorandom string, call it bit i, given the first $i-1$ bits. Using the circuit C, we will produce an AC^0 circuit D that predicts (with non-negligible advantage) the solution to a parity problem T of size r^ν, for some $\nu > 0$, which is impossible.

Begin by noting that, if Y_j denotes the j^{th} bit of Y (0 indexed, counting from least

significant bit up), then we have $X \times Y = X \sum_{j=0}^{r-1} Y_j 2^j = \sum_{j=0}^{r-1} X Y_j 2^j$. Thus, we can understand the multiplication of X by Y as the sum of many shifts of X, where the amount that X is shifted in determined by the locations of the 1s in Y. To be precise, for each j such that $Y_j = 1$, we include a copy of X shifted left by j indices. To produce the product $X \times Y$, we then sum all copies of X. This is illustrated in the figure below.



Each column contains certain bits of X. One way to characterize which bits appear in each particular column is to imagine sliding the strings X and Y^{REV} past one another, where Y^{REV} is the string Y flipped left-to-right. To be precise, start by aligning X and Y^{REV} such that the least significant bits of X and Y line up, and no other bits initially line up. To determine which bits of X lie in column j (where we number the columns from right to left, starting with 0), slide Y^{REV} j bits over; exactly the bits of X that lines up with a 1 in Y appear in column j. This is illustrated in the figure below.



Define sets $U_j \subseteq \{0, \ldots, r-1\}$ such that U_j consists of all indices of X that appear in column j. Let $S_j \subseteq \{0, \ldots, s-1\}$ be a collection of indices of Y. The exact manner in which the S_j are selected will be specified shortly. Let $V_j \subseteq U_j$ be indices of X that appear in column j because they lined up with a 1 in Y at one of the indices S_j. As noted above, we must have $i \in \{r+1, \ldots, n\}$ (the portion of the string containing bits of the product $Z = X \times Y$), and so we are predicting bit $i - r + r^{2\epsilon} - 1 =: k$ of the product. Notice that, if it weren't for the fact that there are carries when computing the sum of the various shifts of X, bit k of the product would simply be the parity of the bits of X selected by U_k. The key idea will be to construct the sets V_k so that they are individually large, $|V_k| > \log^c s$, for

$$FS_n(f,g) = \begin{cases} g(s), & \text{if } \exists s \in \{0,1\}^n \text{ such that } f(x) = x \cdot s \ \forall x \\ *, & \text{otherwise} \end{cases}$$

This function can very naturally be interpreted as encoding a promise problem, called the Fourier sampling problem, in which the promise is that f is a linear function (that is to say a function of the form $f(x) = x \cdot s$), and the value of $FS_n(f,g)$ (when the promise is satisfied) is simply $g(s)$. We will frequently refer to the value s as the *secret* encoded by f.

Next, we define a slight variant of the above problem where the function g is fixed (that is to say that it is not part of the input to the function). Formally, for any positive integer n and any function $g : \{0,1\}^n \to \{0,1\}$, we define the function $FS_n^g : \{0,1\}^{2^n} \to \{0,1\}$ as follows. We now interpret the input to the function as encoding the truth table of a single function $f : \{0,1\}^n \to \{0,1\}$. We then define

$$FS_n^g(f) = \begin{cases} g(s), & \text{if } \exists s \in \{0,1\}^n \text{ such that } f(x) = x \cdot s \ \forall x \\ *, & \text{otherwise} \end{cases}$$

We now define the recursive Fourier sampling function, which is a variant of the Fourier sampling function in which each bit of f is produced, recursively, by a smaller instance of the recursive Fourier sampling problem.

Formally, let $RFS_{n,1} : \{0,1\}^{n+2^n} \to \{0,1\}$ be the (total) Boolean function where the input is interpreted as a pair (s,g) for a secret $s \in \{0,1\}^n$ and a function $g : \{0,1\}^n \to \{0,1\}$ given as a 2^n bit long truth table, and

$$RFS_{n,1}(s,g) = g(s).$$

For each $h > 1$, we define $RFS_{n,h}$ recursively in terms of $RFS_{n,h-1}$ as follows. Let $M_{n,h} = n2^{n(h-1)} + \sum_{j=1}^{h-1} 2^{jn}$. Then $RFS_{n,h} : \{0,1\}^{M_{n,h}} \to \{0,1,*\}$ is the partial Boolean function defined as follows. The input is interpreted as being of the form $(R_0, R_1, \ldots, R_{2^n-1}, g)$,

all constants $c > 0$, but have small intersection with any U_j, $|V_k \cap U_j| \leq 2$, $\forall j < k$, and then fill the bits of X specified by V_k with the bits of an instance of the parity problem. This is very similar to the notion of a combinatorial design, [Nis91], with the exception of the fact that here we consider subsets V_j of U_j.

The circuit D predicts the solution of the parity problem T by producing a multiplication instance to feed to C, that is to say the first $i-1$ bits of a string produced by multiplication. This string consists of a value X and some of the bits of the product XY. We construct this multiplication instance as follows. Begin by setting the bits of X selected by V_k to the bits of the parity instance T. To set the other bits of X, notice that if C can truly predict the next-bit test with non-negligible advantage, then this means, by definition, that the advantage of C, averaged over all choices of X, is non-negligible. In particular, this means that there must exist at least one setting of the other bits of X such that C has non-negligible advantage as just the bits selected by V_k vary (uniformly). We then set the other bits of X to such a fixed value. To be clear, the claim is not that an AC^0 circuit can find a proper setting to the other bits of X, but rather that such a value can simply be built into D (because it is only a single fixed value, which depends only on the input size t of circuit D). In order to calculate the lowest $k-1$ bits of XY that must be fed to C, we write $X = X_{input} + X_{fixed}$ where X_{input} consists of the t bits of the input to D, which are assigned to the positions specified by V_k, as X_{fixed} corresponds to the fixed setting of the other bits of X. Since both Y and X_{fixed} are fixed values, we can also build the value YX_{fixed} into D. Therefore, if it were possible to compute in AC^0 the low $k-1$ bits of YX_{input}, then it would be possible to compute the low $k-1$ bits of XY because $XY = YX_{input} + YX_{fixed}$, and we can, of course, perform addition in AC^0. The key observation is that, with high probability over the choice of Y, it will be easy to compute YX_{input}.

To see this, notice that, with high probability over Y, there will be a choice of S_k such that $|V_k \cap U_j| \leq 2$, for $j \in \{0, \ldots, k-1\}$. This is simply the statement that each column of multiplication problem illustrated in the figure above contains at most two bits of X_{input}.

Therefore, these bits can be packed into two numbers, whose sum (which is calculable in AC^0) will be the low bits of YX_{input}. To see that $|V_k \cap U_j| \leq 2$, with high probability, let Y' be identical to Y except that all bits outside of S_k are set to 0, and note that $|V_k \cap U_j|$ is simply the number of 1s that line up when Y and Y' are slid over one another, or, in other words, the number of h such that Y'_h and $Y_{h-(k-j)}$ are both 1. To bound the probability that $|V_k \cap U_j|$ fails to be at most 2 for every j, we show this failure probability (where, again, the probability is taken over the choice of Y) is extremely small for a single fixed j and union bound over the j. Fix j and define $Q_h = Y'_h Y_{h-(k-j)}$; then $|V_k \cap U_j| = \sum_h Q_h$. Unfortunately, the Q_h are not independent. To deal with this, partition the indices h into two classes, where the first class contains all h such that $h \mod 2(k-j)$ falls in the range $[0, k-j-1]$ and the second class contains all other h. Notice that h and $h-(k-j)$ always are in separate classes, and so the set of all Q_h such that h is in the first class are independent, and, similarly, the set of all Q_h such that h is in the second class are independent. We show that $\sum_h Q_h \leq 1$, where the sum is restricted to a single class. Recall that the bits of Y are generated (independently) such that each bit is 1 with probability $r^{-\epsilon}$ and that, if we select the special bits S_k at random (which is allowed because we need only show $\exists S_K$ that satisfies the above) such that each of the bits of Y that line up with a portion of X (when sliding Y over X, only part of Y lines up with actual indices of X at any given shift) are included in S_k with probability $r^{-(1-\epsilon)}$ then a bit of Y' is 1 with probability $r^{-(1-2\epsilon)}$. The result follows from a simple application of the Chernoff bound.

Thus far, we have shown that D can produce a multiplication instance to feed to C. To use the result produced by C (namely, the predicted next bit of the product) to determine the parity of T, notice that the correct value of the next bit of the product is simply the exclusive-or of the parity of T, the parity of those bits of X_{fixed} that appear in column k of the multiplication problem, and the carry bit that enters column k when the low $k-1$ bits of YX_{input} and YX_{fixed} are added to produce the low $k-1$ bits of the product XY. Since X_{fixed} is a single fixed value, the parity of those bits that appear in column k can be

built in to D. As noted earlier, it is possible, in AC^0, to compute the sum of the low $k-1$ bits of YX_{input} and YX_{fixed}, including the carry into column k. Thus, if the next bit can be predicted with some advantage, then the parity of T can be predicted with the exact same advantage. This contradiction completes the proof that the multiplication problem, as defined above, looks random to AC^0.

□

It is worth noting that, while the above proof was only carried out in the case when $r < s^\alpha$ for all constants $\alpha > 0$, but $r > \log^c s$ for all constants c, the same technique would also work for other parameters, such as if $s = r^u$, for some constant u (the parameters of Lemma 12). Moreover, a similar argument would show that, if $r = O(\log^c s)$, then f_Y passes all AC^0 tests of depth at most d, where d depends on c.

3.2.3 The Algebraic Integer Problem

In this section, it is shown that the algebraic integer problem looks random to AC^0 circuits. We begin with a few definitions. An algebraic integer is a root of some monic polynomial with integer coefficients. An algebraic number field is a finite field extension of $\mathbb{Q}$. Given some algebraic number field K, the ring of integers of K, denoted O_K, is the ring that consists of all algebraic integers in K. For every K, O_K is a free $\mathbb{Z}$-module, and so has an integral basis (that is to say, $\exists b_1, \ldots b_h \in O_K$ such that every element of O_K can be uniquely expressed as $\sum_i a_i b_i$, for $a_i \in \mathbb{Z}$). For a particular basis B, we define the function $f_B : \{0,1\}^{m_1} \times \cdots \times \{0,1\}^{m_h} \to \{0,1\}^k$ such that $f_B(a_1, \ldots, a_h)$ is the first k bits of the binary expansion of the fractional real part of $\sum_i a_i b_i$, where for $i > 1$, $m_i = m_1^{u_i}$ for some constant $u_i > 0$, and $k = m_1^u$, for any constant u. We show, via reduction from the multiplication problem, that certain f_B are AC^0-pseudorandom.

As an example, consider the algebraic number field $K = \mathbb{Q}(\sqrt{d})$, for d a squarefree positive integer. It can be shown that, when $d \equiv 2, 3 \mod 4$, then $\{1, \sqrt{d}\}$ is an integral basis for O_K and that when $d \equiv 1 \mod 4$, $\{1, (1+\sqrt{d})/2\}$ is an integral basis for O_K (of

course, since d is squarefree, we can't have $d \equiv 0 \mod 4$). Let b_1 and b_2 denote the basis elements, in the order they appear above. Then $f_B(a_1, a_2)$ is simply the first k bits of the fractional part of $a_1 b_1 + a_2 b_2$, which is identical to the first k bits of the fractional part of $a_2 b_2$ (because $a_1, b_1 \in \mathbb{Z}$). It is straightforward to show that, for all sufficiently large n, and all strings $Y \in \{0,1\}^{\lfloor n/2 \rfloor - 1}$, there is an n bit value d for which the binary expansion of the fractional part of $\sqrt{d}$ starts with the string Y. In particular, if we consider a string Y such that the multiplication function f_Y is AC^0-pseudorandom, then the corresponding f_B is also AC^0-pseudorandom, because it is just the multiplication problem $a_2\sqrt{d}$ bit-shifted, possibly with $1/2$ added.

In general, consider any basis B of some O_K such that there is some basis element b_j in B such that the binary expansion of the fractional real part of b_j starts with a value Y for which f_Y is AC^0-pseudorandom. Rather than consider f_B directly, it will again be convenient to consider a variant of the function in which some of the inputs are held fixed. In particular, we wish to fix a_i for each $i \neq j$. Define the function $f_{B,j,a_1,\ldots,a_{j-1},a+j+1,a_h} : \{0,1\}^{m_j} \to \{0,1\}^k$ such that it maps the value a_j to the first k bits of $\sum_i a_i b_i$. By a straightforward reduction from the multiplication problem, it follows that $f_{B,j,a_1,\ldots,a_{j-1},a+j+1,a_h}$ is AC^0-pseudorandom, which then immediately implies that f_B is AC^0-pseudorandom.

Chapter 4

Polynomial Degree

4.1 Recursive Fourier Sampling

In this section, we consider the recursive Fourier sampling problem. Numerous variants of this problem have been considered by many authors (see, for instance, [BV93], [BV97], [Aar03], [Aar10], [Joh08]). The version considered in this thesis, and the notation used, follows most closely [Joh08], but essentially the same claims hold for all other standard variants. We begin by precisely defining the problem.

4.1.1 Definition of the Problem

First, we define the Fourier sampling function. For every positive integer n, we define the partial Boolean function $FS_n : \{0,1\}^{2^{n+1}} \to \{0,1,*\}$ as follows. We interpret the 2^{n+1} bit long input to FS_n as a pair of truth tables defining the functions $f, g : \{0,1\}^n \to \{0,1\}$. For $x, s \in \{0,1\}^n$, let x_i and s_i denote the i^{th} bit of x and s, respectively. Let $x \cdot s = \sum_i x_i s_i$ denote the usual Boolean inner product (where of course the sum is evaluated modulo 2). Then

where for each $\sigma \in \{0,1\}^n$, R_σ is an instance of $RFS_{n,h-1}$ and g is a funciton $g : \{0,1\}^n \to \{0,1\}$ given as a 2^n bit long truth table. We then define

$$RFS_{n,h}(R_0, \ldots, R_{2^n-1}, g) = \begin{cases} g(s), & \text{if } \exists s \in \{0,1\}^n \text{ such that } \forall \sigma \in \{0,1\}^n \ RFS_{n,h-1}(R_\sigma) = \sigma \cdot s \\ *, & \text{otherwise} \end{cases}$$

In a precisely analogous fashion, we define $RFS^g_{n,h}$ where now there is a single fixed g used throughout the problem, rather than a collection of functions provided as part of the input.

We very naturally interpret $RFS_{n,h}$ and $RFS^g_{n,h}$ as encoding a particular promise problem, where the promise is that, at every node in the tree, there exists some $s \in \{0,1\}^n$ such that the function $f : \{0,1\}^n \to \{0,1\}$ defined at this node is of the form $f(x) = x \cdot s$.

Fix the entire input to the recursive Fourier sampling function in any way such that every promise is satisfied. For any node t in the tree, we define the *value* of the node, which we denote by $b(t)$ to be the output of the instance of recursive Fourier sampling corresponding to the subtree rooted at t.

Notice that, due to the structure of the promise, in order to determine the value of node t, it is only necessary to know the values of n linearly independent children of t. That is to say, if the children of t are given by $C(t) = \{t_\sigma : \sigma \in \{0,1\}^n\}$, then $b(t)$ is completely determined by the value of a subset of children C' for any $C' \subseteq C$ such that $C' = \{t_{\sigma_1}, \ldots, t_{\sigma_n}\}$ where $\{\sigma_1, \ldots, \sigma_n\}$ are linearly independent (as vectors in $\{0,1\}^n$, in other words the σ_i form a basis of $\{0,1\}^n$).

For $i \in [n]$, let $\chi_i \in \{0,1\}^n$ denote the i^{th} elementary basis element. That is to say χ_i has value 1 in position i and 0 elsewhere. Clearly, the set of χ_i form a basis of $\{0,1\}^n$, and so, for any node t, the value of node t is completely determined by the values of these

children. We call this set of children the *elementary children* of t, which we denote by

$$C_e(t) = \{t_{\chi_i} : i \in [n]\}.$$

Therefore, given an instance (a particular single setting of the input) of $RFS_{n,h}$ or $RFS^g_{n,h}$ that is guaranteed to satisfy the promise, the answer (the value of the root of the tree) can be determined by first determining the value of the n elementary children of t. The value of each of these children can be determined from their n elementary children. This process can be repeated until the leaves of the tree are reached, at which point the value of each node is simply the output of an instance of $RFS_{n,1}$. We refer to this collection of leaves obtained by repeatedly finding elementary children as the *elementary leaves.* For a tree of height h, there are clearly n^{h-1} elementary leaves.

4.1.2 Recursive Fourier Sampling is δ-versatile

In this section, we show that for certain natural choices of the function g, such as the majority function or the generalized inner product function, $RFS^g_{n,h}$ is δ-versatile, for suitably chosen δ.

Fix n, and let m denote the total length of the input to $RFS^g_{n,h}$. Clearly $m = n2^{(h-1)n}$. Let $U^g_{p,h} \subseteq \mathbb{F}_2^m$ denote the set of all points at which all promises are satisfied (that is to say, the set of all values of inputs to the recursive Fourier sampling function such that, at every node of the tree, every linearity constraint is satisfied). We frequently refer to $U^g_{p,h}$ as the "promise". On the promise, the recursive Fourier sampling problem is, of course, a total function. By slight abuse of notation, we also denote this induced total function as $RFS^g_{n,h} : U^g_{p,h} \to \mathbb{F}_2$. Similarly, we define $U^g_{0,h} = (RFS^g_{n,h})^{-1}(0)$ and $U^g_{1,h} = (RFS^g_{n,h})^{-1}(1)$ as the points at which the recursive Fourier sampling problem evaluates to 0 and 1, respectively. The superscript g will often be omitted when the function is clear from context.

The first key result of this section, which holds for any g, is the following lower bound

on regularity of $U^g_{p,h}$,$U^g_{0,h}$, and $U^g_{1,h}$.

Lemma 16. *For any positive integers n, h and for any $g \in \mathbb{F}_2[x_1, \ldots, x_n]$, let $d = deg(g)$ and let $RFS^g_{n,h} : \mathbb{F}_2^m \to \mathbb{F}_2$ denote the recursive Fourier sampling function. Then*

$$reg(U^g_{p,h}) \geq nd^{h-1} + (n-d)\sum_{j=1}^{h-1} 2^{jn} d^{h-j-1}$$

$$reg(U^g_{0,h}), reg(U^g_{1,h}) \geq (n-d)\sum_{j=0}^{h-1} 2^{jn} d^{h-j-1}.$$

Proof. Let $r_{U_{p,h}} \in \mathbb{F}_2[x_1, \ldots, x_m]$ denote the unique squarefree polynomial such that $r_{U_{p,h}}(x) = 1$ if and only if $x \in U_{p,h}$. By a straightforward counting of the number of promises of each degree, we have $\deg(r_{U_{p,h}}) \leq (2^n - n)\sum_{j=1}^{h-1} 2^{(j-1)n} d^{h-j}$. By construction $r_{U_{p,h}}$ vanishes on $\overline{U_{p,h}}$ and so

$$\begin{aligned}
a(\overline{U_{p,h}}) &\leq \deg(r_{U_{p,h}}) \\
&\leq (2^n - n)\sum_{j=1}^{h-1} 2^{(j-1)n} d^{h-j} \\
&= 2^n \left(\sum_{j=1}^{h-1} 2^{(j-1)n} d^{h-j}\right) - n\left(\sum_{j=1}^{h-1} 2^{(j-1)n} d^{h-j}\right) \\
&= \left(\sum_{j=1}^{h-1} 2^{jn} d^{h-j}\right) - n\left(\sum_{j=1}^{h-1} 2^{(j-1)n} d^{h-j}\right) \\
&= d\left(\sum_{j=2}^{h} 2^{(j-1)n} d^{h-j}\right) - n\left(\sum_{j=1}^{h-1} 2^{(j-1)n} d^{h-j}\right) \\
&= d2^{(h-1)n} - nd^{h-1} - (n-d)\sum_{j=2}^{h-1} 2^{(j-1)n} d^{h-j}.
\end{aligned}$$

Applying Lemma 2, we then have

$$\mathrm{reg}(U_{p,h}) = n2^{(h-1)n} - a(\overline{U_{p,h}})$$

$$\geq (n-d)2^{(h-1)n} + nd^{h-1} + (n-d)\sum_{j=2}^{h-1} 2^{(j-1)n} d^{h-j}$$

$$= nd^{h-1} + (n-d)\sum_{j=2}^{h} 2^{(j-1)n} d^{h-j}$$

$$= nd^{h-1} + (n-d)\sum_{j=1}^{h-1} 2^{jn} d^{h-j-1}.$$

Similarly, define $r_{U_{0,h}}, r_{U_{1,h}} \in \mathbb{F}_2[x_1, \ldots, x_m]$ as the unique squarefree polynomials such that $r_{U_{0,h}}(x) = 1$ if and only if $x \in U_{0,h}$ and $r_{U_{1,h}}(x) = 1$ if and only if $x \in U_{1,h}$. We then immediately have $\deg(r_{U_{0,h}}), \deg(r_{U_{1,h}}) \leq \deg(r_{U_{p,h}}) + d^i$, and so, by a precisely analogous argument as above

$$\operatorname{reg}(U_{0,h}), \operatorname{reg}(U_{1,h}) \geq (n-d)d^{h-1} + (n-d)\sum_{j=1}^{h-1} 2^{jn} d^{h-j-1}$$

$$= (n-d)\sum_{j=0}^{h-1} 2^{jn} d^{h-j-1}.$$

□

We now exhibit certain functions for which the above lower bounds on regularity are exact. The first such example is the majority function, for certain appropriately chosen input sizes. For a $x \in \{0,1\}^n$, let $x = (x_1, \ldots, x_n)$ and let $wt(x) = |\{i : x_i = 1\}|$ denote the number of 1s in x. Let $\mathrm{MAJ} : \mathbb{F}_2^n \to \mathbb{F}_2$ be defined such that $\mathrm{MAJ}(x) = 1$ if and only if $wt(x) \geq \frac{n}{2}$. We begin by determining the unique squarefree polynomial in $\mathbb{F}_2[x_1, \ldots, x_n]$ that represents MAJ. Let $e_i(x) = \sum_{J \subseteq [n], |J|=i} \prod_{j \in J} x_j$ denote the i^{th} elementary symmetric polynomial. For $y, z \in \{0,1\}^l$, write $y \geq_b z$ if and only if $y_i \geq z_i \ \forall i$.

Lemma 17. *For any positive integer n, the unique squarefree polynomial in $\mathbb{F}_2[x_1, \ldots, x_n]$ that is identically equal to $MAJ : \mathbb{F}_2^n \to \mathbb{F}_2$ on $\mathbb{F}_2^n$ is given by*

$$\sum_{l \geq \frac{n}{2}} \sum_{j \geq_b l} e_j(x).$$

Proof. Begin by noticing that

$$e_i(x) = \binom{wt(x)}{i} \mod 2.$$

By a straightforward application of Kummer's lemma, we then conclude

$$e_i(x) = \begin{cases} 1, & wt(x) \geq_b i \\ 0, & \text{otherwise} \end{cases}.$$

Next, define functions $E_i : \mathbb{F}_2^n \to \mathbb{F}_2$ and $G_i : \mathbb{F}_2^n \to \mathbb{F}_2$ such that $E_i(x) = 1$ if and only if $wt(x) = i$ and $G_i(x) = 1$ if and only if $wt(x) \geq i$. We then have

$$E_i(x) = \sum_{j \geq_b i} e_j(x).$$

To see this, simply notice that if $E_i(x) = 1$ then $wt(x) = i$ and so $e_i(x) = 1$, but $e_j(x) = 0$ for all other terms in the above sum. If $E_i(x) = 0$, then $wt(x) = t \neq i$. There are then two cases: if $i \leq_b t$, then the only terms in the above sum that evaluate to one are precisely all values j such that $i \leq_b j \leq t$, of which there are an even number; if $i \not\leq_b t$, then $\forall j$ such that $j \geq_b i$, $j \not\leq_b t$, and so every term in the above sum evaluates to zero.

We then have

$$\begin{aligned} G_i(x) &= \sum_{l \geq i} E_l(x) \\ &= \sum_{l \geq i} \sum_{j \geq_b l} e_j(x), \end{aligned}$$

and so

$$\begin{aligned} \text{MAJ}(X) &= G_{\frac{n}{2}}(x) \\ &= \sum_{l \geq \frac{n}{2}} \sum_{j \geq_b l} e_j(x). \end{aligned}$$

□

We now consider $RFS_{n,h}^{\mathrm{MAJ}}$. We begin by demonstrating a useful symmetry in $U_{p,h}^{\mathrm{MAJ}}$. Define the value $\hat{1}_h \in U_{p,h}^{\mathrm{MAJ}}$ as follows. Consider the recursive Fourier sampling tree. We define $\hat{1}_h$ by first defining $b(t)$ for every node t in the tree (that is to say, we define the value $b(t)$ that node t has with input $\hat{1}_h$). First, assign the root of the tree the value 1. Then, for each node that has been assigned a value, assign values to the children of that node as follows. If node t has value $b(t)$, then set $b(t_\sigma) = b(t)$ for each $t_\sigma \in C_e(t)$. Assign all other children the value forced by the promise: for each $t_\sigma \in C(t) \setminus C_e(t)$, set $b(t_\sigma) = \sum_{j \in [n], \sigma_j = 1} b(t_{\chi_j})$. Equivalently, if a node has value 0, all of its children have value 0; if a node has value 1, then each child t_σ has value given by the parity of the string σ. Once the entire tree has been labeled in such a fashion, define $\hat{1}_h$ by setting the portion of the input corresponding to each leaf (that is to say, the n places of the input representing the secret at that leaf) to the value of that leaf.

It is clear that the value $\hat{1}_h \in U_{p,h}^{\mathrm{MAJ}}$ as claimed, due to the fact that $\hat{1}_h$ was constructed in a way such that the promise is satisfied at every node. Moreover, $\hat{1}_h \in U_{1,h}^{\mathrm{MAJ}}$ as, by construction, the value of the root is 1. For any $x \in U_{p,h}^{\mathrm{MAJ}}$, let $\hat{x} = x \oplus \hat{1}_h$ (where $\oplus$ denotes bitwise parity). We then have the following.

Lemma 18. *For any odd positive integer n and any positive integer h, $x \in U_{0,h}^{MAJ}$ if and only if $\hat{x} \in U_{1,h}^{MAJ}$.*

Proof. Given any $x \in U_{0,h}^{\mathrm{MAJ}}$, the root of the corresponding recursive Fourier sampling tree has value 0. The key observation is that adding $\hat{1}_h$ flips the value at every elementary leaf of the tree. That is to say, if on input x, a particular elementary leaf t has value $b \in \{0,1\}$, then on input $\hat{x}$, that leaf has value $\overline{b}$. This occurs because, by construction, $\hat{1}_h$ is 1 at every position in the elementary leaves. It is then straightforward to see that value of the root of the tree flips and that every promise is preserved, which implies $\hat{x} \in U_{1,h}^{\mathrm{MAJ}}$. The reverse implication follows from the fact that $\hat{\hat{x}} = x$ and symmetry.

□

We now show that $U_{0,h}^{\text{MAJ}}$ and $U_{1,h}^{\text{MAJ}}$ have identical standard monomials.

Lemma 19. *For any odd positive integer n and any positive integer h, $SM(U_{0,h}^{MAJ}) = SM(U_{1,h}^{MAJ})$.*

Proof. For any algebraic set, every monomial is either a leading monomial or a standard monomial, and so it suffices to show $\text{LM}(U_{0,h}^{\text{MAJ}}) = \text{LM}(U_{1,h}^{\text{MAJ}})$.

We first show $\text{LM}(U_{0,h}^{\text{MAJ}}) \subseteq \text{LM}(U_{1,h}^{\text{MAJ}})$. Consider any $x^\alpha \in \text{LM}(U_{0,h}^{\text{MAJ}})$. By definition, $\exists q_\alpha \in \mathbb{F}_2[x_1, \ldots, x_m]$ such that $q_\alpha \in I(U_{0,h}^{\text{MAJ}})$ and $\text{lm}(q_\alpha) = x^\alpha$. Define $\hat{q}_\alpha \in \mathbb{F}_2[x_1, \ldots, x_m]$ such that $\hat{q}_\alpha(x) = q_\alpha(\hat{x})$. Notice that

$$\text{lm}(\hat{q}_\alpha) = \text{lm}(q_\alpha) = x^\alpha.$$

Moreover, for any $x \in U_{1,h}^{\text{MAJ}}$, Lemma 18 implies that $\hat{x} \in U_{0,h}^{\text{MAJ}}$ and so

$$\hat{q}_\alpha(x) = q_\alpha(\hat{x}) = 0,$$

where the last follows from the fact that q vanishes on $U_{0,h}^{\text{MAJ}}$. This implies that $\hat{q}_\alpha \in I(U_{1,h}^{\text{MAJ}})$, and so $x^\alpha \in \text{LM}(U_{1,h}^{\text{MAJ}})$. Therefore, $\text{LM}(U_{0,h}^{\text{MAJ}}) \subseteq \text{LM}(U_{1,h}^{\text{MAJ}})$.

A precisely symmetric argument implies $\text{LM}(U_{0,h}^{\text{MAJ}}) \supseteq \text{LM}(U_{1,h}^{\text{MAJ}})$.

□

Next, we provide upper bounds for the regularity of $U_{p,h}^{\text{MAJ}}$, $U_{0,h}^{\text{MAJ}}$, and $U_{1,h}^{\text{MAJ}}$.

Lemma 20. *For any odd positive integer n and any positive integer h, let $RFS_{n,h}^{MAJ} : \mathbb{F}_2^m \to \mathbb{F}_2$ denote the recursive Fourier sampling function with majority. Then*

$$reg(U_{p,h}^{MAJ}) \leq n\left(\frac{n+1}{2}\right)^{h-1} + \left(\frac{n-1}{2}\right)\sum_{j=1}^{h-1} 2^{jn}\left(\frac{n+1}{2}\right)^{h-j-1}$$

$$reg(U_{0,h}^{MAJ}), reg(U_{1,h}^{MAJ}) \leq \left(\frac{n-1}{2}\right) \sum_{j=0}^{h-1} 2^{jn} \left(\frac{n+1}{2}\right)^{h-j-1}.$$

Proof. We show this by induction on h. First, consider the case in which $h = 1$. Clearly, $U_{p,1}^{\text{MAJ}} = \mathbb{F}_2^n$ and so $\text{reg}(U_{p,1}^{\text{MAJ}}) = n$. Moreover, $U_{0,1}^{\text{MAJ}} = (\text{MAJ})^{-1}(0)$ and $U_{1,1}^{\text{MAJ}} = (\text{MAJ})^{-1}(1)$, and so, by Lemma 5, we have $\text{reg}(U_{0,1}) = \text{reg}(U_{1,1}) = \frac{n-1}{2}$.

We now consider the case in which $h > 1$. First, consider $U_{p,h}^{\text{MAJ}}$. By the definition of regularity, $\text{reg}(U_{p,h})$ is the minimal value of d such that $h^a(U_{p,h}, d) = |U_{p,h}|$. Therefore, if, for some d, $h^a(U_{p,h}, d) = |U_{p,h}|$, then $\text{reg}(U_{p,h}) \leq d$. In particular, let

$$d(h) = n\left(\frac{n+1}{2}\right)^{h-1} + \left(\frac{n-1}{2}\right) \sum_{j=1}^{h-1} 2^{jn} \left(\frac{n+1}{2}\right)^{h-j-1}.$$

Then, in order to show $\text{reg}(U_{p,h}) \leq d(h)$, it suffices to show $h^a(U_{p,h}, d(h)) = |U_{p,h}|$.

To show this, as before, let $m = n2^{n(h-1)}$ denote the total size of the input to $RFS_{n,h}^{\text{MAJ}}$, and let

$$M_d = \mathcal{M}\left(U_{p,h}, \binom{[m]}{\leq d}\right)$$

denote the inclusion matrix in which the rows are indexed by elements of $U_{p,h}$ and the columns are indexed by all squarefree monomials of degree at most d. By Lemma 3, $h^a(U_{p,h}, d) = \text{rank}_{\mathbb{F}_2}(M_d)$, and so it suffices to show $\text{rank}_{\mathbb{F}_2}(M_{d(h)}) = |U_{p,h}|$. Observe that $|U_{p,h}|$ is precisely the number of rows of $M_{d(h)}$ (and is, of course, substantially smaller than the number of columns), and so this is equivalent to showing that the matrix $M_{d(h)}$ is full rank.

To see that $M_{d(h)}$ is full rank, assume, for contradiction, that it is not. By definition, this means that there exists some non-empty $T \subseteq U_{p,h}$ such that the sum of the rows of $M_{d(h))}$ indexed by T is 0 in every column. We now show that, for any $T \subseteq U_{p,h}$, $\exists \alpha$ such that the rows indexed by T have the sum 1 in the column indexed by the monomial x^α, which is, of course, a contradiction.

Let x_i denote the i^{th} input variable. Let $E \subseteq [m]$ denote the indices of all variables that are inputs to the elementary leaves of the recursive Fourier sampling tree. Clearly,

$|E| = n^h$ as there are n^{h-1} elementary leaves, each of which have n input variables. Define $\sigma : U_{p,h} \to \{0,1\}^{n^h}$ such that, for any $x \in U_{p,h}$, $\sigma(x)$ is the portion of x at indices E. We refer to this value as the *signature* of x. Consider a partial ordering on the set of signatures given by the usual bitwise ordering. That is to say, for any $y, z \in \{0,1\}^{n^h}$, let y_i and z_i denote the i^{th} bits of y and z, respectively. Define $y \leq z$ if $y_i \leq z_i \; \forall i$. Similarly, define $y < z$ if $y \leq z$ and $y \neq z$. Let $S_T = \{\sigma(x) : x \in T\}$ and M_T denote an (arbitrary) maximal element of S_T with respect to the partial order on signatures. That is to say, M_T is any single value that satisfies $M_T \in S_T$ and $\nexists y \in S_T$ such that $M_T < y$.

Recall that each column of $M_{d(h)}$ is indexed by a squarefree monomial $x^\alpha = x_1^{\alpha_1} \cdots x_m^{\alpha_m}$. Consider any column of $M_{d(h)}$ that is indexed by some x^α such that α agrees with M_T (that is to say, for each $i \in E$, α_i is equal to the corresponding value of M_T). The key observation is that the only rows $x \in T$ that could possibly have value 1 in column x^α are those such that $\sigma(x) = M_T$. To see this, notice that in order for a particular row $x \in T$ to have entry 1 in column x^α, it must be the case that $x_i = 1$ at every $i \in E$ such that $\alpha_i = 1$, and so, by definition, $\sigma(x) \geq M_T$. If $\sigma(x) \neq M_T$, then $\sigma(X) > M_T$, which contradicts the definition of M_T, and so we must have $\sigma(X) = M_T$, as claimed.

Let $Z \subseteq T$ be defined such that $Z = \{x \in T : \sigma(x) = M_T\}$. Then, for any column indexed by an x^α such that α agrees with M_T, the sum over all $x \in T$ and the sum over only those $x \in Z$ must be equal. Therefore, it suffices to exhibit a column indexed by x^α such that α agrees with M_T and the sum over all rows $x \in Z$ in column x^α is 1.

To do this, notice that the set Z is an algebraic set (as it is simply a set of elements in $\mathbb{F}_2^m$) where every $x \in Z$ lies within a particular subspace, namely the subspace consisting of the set of x that satisfy $\sigma(x) = M_T$. We now consider $\widetilde{Z}$, which is the induced algebraic set living within that subspace. More formally, we partition the collection of variables into two pieces: E and $[m] \setminus E$. For any $x \in \mathbb{F}_2^n$, let x_E and $x_{[m]\setminus E}$ denote the portions of x indexed by E and $[m] \setminus E$ respectively. We define the algebraic set $\widetilde{Z} \subseteq \mathbb{F}_2^{m-n^h}$ where $\widetilde{Z} = \{x_{[m]\setminus E} : x \in Z\}$.

We now consider the inclusion matrix

$$\widetilde{M} = \mathcal{M}\left(\widetilde{Z}, \binom{[m-n^h]}{d(h)-n^h}\right),$$

where the rows are indexed by the $x_{[m]\setminus E} \in \widetilde{Z}$ and the columns are indexed by the monomials $x^{\beta}_{[m]\setminus E}$. The next key observation is that, in order to prove the existence of a column x^{α} of the desired form, it suffices to show $\mathrm{rank}_{\mathbf{F}_2}\widetilde{M} = |\widetilde{Z}|$, in other words, that the matrix $\widetilde{M}$ is full rank. To see this, notice that if $\widetilde{M}$ is full rank then, by definition, for every non-empty set of rows $R \subseteq \widetilde{Z}$, there is some column $x^{\beta}_{[m]\setminus E}$ such that the sum in that column over the rows R is equal to 1. In particular, there is some column $x^{\beta}_{[m]\setminus E}$ such that the sum of every row in the column $x^{\beta}_{[m]\setminus E}$ is equal to 1. Fix any such β, and define α such that $\alpha_E = M_T$ and $\alpha_{[m]\setminus E} = \beta$. By construction, the sum of the entries of $M_{d(h)}$ in column x^{α} and rows Z is equal to the sum of the entries of $\widetilde{M}$ in column $x^{\beta}_{[m]\setminus E}$ and rows $\widetilde{Z}$. Therefore, the sum of the entries of $M_{d(h)}$ in column x^{α} and rows Z is 1, as desired.

All that remains is to show $\mathrm{rank}_{\mathbf{F}_2}\widetilde{M} = |\widetilde{Z}|$. By Lemma 3, this is equivalent to showing $\mathrm{reg}(\widetilde{Z}) \leq d(h) - n^h$. Before providing the details of this regularity bound, we briefly state the main idea which is that $\widetilde{Z} \subseteq V_1 \times \cdots \times V_w$ where each V_i is (isomorphic to) either U^{MAJ}_{0,h_i} or U^{MAJ}_{1,h_i} where each $h_i < h$. The induction hypothesis bounds the regularity of each such V_i, which in turn provides the required bound on the regularity of $\widetilde{Z}$, because, by the definition of regularity, $\widetilde{Z} \subseteq V_1 \times \cdots \times V_w$ immediately implies

$$\mathrm{reg}(\widetilde{Z}) \leq \mathrm{reg}(V_1 \times \cdots \times V_w) = \sum_i \mathrm{reg}(V_i).$$

We now show the required bound on $\mathrm{reg}(\widetilde{Z})$. By construction $\widetilde{Z}$ is the algebraic set consisting of the elements of T which reside in the subspace defined by $\sigma(x) = M_T$. We now consider how the constraint $\sigma(x) = M_T$ interacts with the linearity promise of recursive Fourier sampling. Consider the recursive Fourier sampling tree. The key observation is that the constraint $\sigma(x) = M_T$ fixes the value of all of the elementary children, which in turn

fixes the value of every "sibling" of an elementary child. This, essentially, "decouples" the problem into the cartesian product of several independent, smaller instances of the recursive Fourier sampling problem.

To be precise, begin by noting that requiring $\sigma(x) = M_T$ directly forces the value (that is to say, the output) of each of the elementary *leaves* of the recursive Fourier sampling tree. By simply propagating this constraint upward through the tree, the value of all of the elementary *children* is also forced. To see this, simply notice that, by construction, if the value of all elementary children of a particular node t is forced, then the value of t itself is forced. Since each elementary child which is not an elementary leaf has its own collection of elementary children, the result immediately follows.

We therefore conclude that the constraint $\sigma(x) = M_T$ forces the value of all n^{h-1} elementary children. Of course, this is only a tiny portion of the $\theta(n2^{nh})$ nodes of the recursive Fourier sampling tree. However, the linearity constraint imposed by the promise within recursive Fourier sampling causes the constraint $\sigma(x) = M_T$ to constrain other portions of the recursive Fourier sampling tree. In particular, begin by considering the root of the recursive Fourier sampling tree. As noted above, the constraint $\sigma(x) = M_T$ directly forces the value of each of the n elementary children of the root. Moreover, due to the linearity constraint, the value of the other $2^n - n$ children of the root are also forced. In particular, if we let t denote the root of the tree, t_i denote its i^{th} child, $b(t_i)$ denote the value of node t_i, i_j denote the j^{th} bit of i, and χ_j denote the element of $\{0,1\}^n$ which has value 1 in position j and value 0 elsewhere, then

$$b(t_i) = \sum_{j:i_j=1} b(t_{\chi_j}).$$

Therefore, for any x that satisfies $\sigma(x) = M_T$, if we consider the portion of x that lies under the subtree rooted at t_i, for any t_i which is *not* an elementary child of the root node t, then this portion of x must lie within an algebraic set isomorphic to $U^{\text{MAJ}}_{b(t_i),h-1}$.

Precisely the same logic applies if we consider any elementary child that is not an elementary leaf. At l levels down from the root of the tree, there are n^l elementary children,

each of which have n elementary children and $2^n - n$ non-elementary children. For any x that satisfies $\sigma(x) = M_T$, the portion of x that lies under the subtree rooted at each of the non-elementary child t must lie within an algebraic set isomorphic to $U^{\mathrm{MAJ}}_{b(t),h-l-1}$.

Next, notice that this process completely partitions the input of the recursive Fourier sampling tree into a piece that lies beneath the elementary leaves and many other pieces which each lie beneath the subtree rooted at a non-elementary sibling of some elementary child. To see this, consider any particular input variable x_i and consider its highest ancestor (other than the root of the tree) which is not an elementary child. If such an ancestor does not exist, then this variable is an input to an elementary leaf. If such an ancestor does exist, then it must be a non-elementary child of an elementary child (or of the root of the tree), and so this ancestor will have elementary children of its parent as siblings. We therefore conclude that $\widetilde{Z} \subseteq V_1 \times \cdots \times V_w$ where each V_i is (isomorphic to) either U^{MAJ}_{0,h_i} or U^{MAJ}_{1,h_i} where each $h_i < h$, as claimed. Counting the number of copies of each $U^{\mathrm{MAJ}}_{0,j}$ and $U^{\mathrm{MAJ}}_{1,j}$, using Lemma 19 to conclude that $\mathrm{reg}(U^{\mathrm{MAJ}}_{0,j}) = \mathrm{reg}(U^{\mathrm{MAJ}}_{1,j})$, and applying the induction hypothesis to bound the regularity of $U^{\mathrm{MAJ}}_{0,j}$ and $U^{\mathrm{MAJ}}_{1,j}$ yields the following.

$$\mathrm{reg}(\widetilde{Z}) \leq \sum_{i=1}^{h-1} n^{i-1}(2^n - n)\mathrm{reg}(U^{\mathrm{MAJ}}_{0,h-i})$$

$$\leq \sum_{i=1}^{h-1} n^{i-1}(2^n - n)\left[\left(\frac{n-1}{2}\right)\sum_{j=0}^{h-i-1} 2^{jn}\left(\frac{n+1}{2}\right)^{h-i-j-1}\right]$$

$$= \left(\frac{n-1}{2}\right)\left[\left(\sum_{i=1}^{h-1} n^{i-1}\sum_{j=0}^{h-i-1} 2^{(j+1)n}\left(\frac{n+1}{2}\right)^{h-i-j-1}\right) - \left(\sum_{i=1}^{h-1} n^{i}\sum_{j=0}^{h-i-1} 2^{jn}\left(\frac{n+1}{2}\right)^{h-i-j-1}\right)\right]$$

$$= \left(\frac{n-1}{2}\right)\left[\left(\sum_{i=0}^{h-2} n^{i}\sum_{j=0}^{h-i-2} 2^{(j+1)n}\left(\frac{n+1}{2}\right)^{h-i-j-2}\right) - \left(\sum_{i=1}^{h-1} n^{i}\sum_{j=0}^{h-i-1} 2^{jn}\left(\frac{n+1}{2}\right)^{h-i-j-1}\right)\right]$$

$$= \left(\frac{n-1}{2}\right)\left[\left(\sum_{i=0}^{h-2} n^{i}\sum_{j=1}^{h-i-1} 2^{jn}\left(\frac{n+1}{2}\right)^{h-i-j-1}\right) - \left(\sum_{i=1}^{h-1} n^{i}\sum_{j=0}^{h-i-1} 2^{jn}\left(\frac{n+1}{2}\right)^{h-i-j-1}\right)\right]$$

$$= \left(\frac{n-1}{2}\right)\left[\left(\sum_{j=1}^{h-1} 2^{jn}\left(\frac{n+1}{2}\right)^{h-j-1}\right) - n^{h-1} - \left(\sum_{i=1}^{h-2} n^i\left(\frac{n+1}{2}\right)^{h-i-1}\right)\right]$$

$$= \left(\frac{n-1}{2}\right)\left[\left(\sum_{j=1}^{h-1} 2^{jn}\left(\frac{n+1}{2}\right)^{h-j-1}\right) - \left(\sum_{i=1}^{h-1} n^i\left(\frac{n+1}{2}\right)^{h-i-1}\right)\right]$$

$$= \left(\frac{n-1}{2}\right)\left(\sum_{j=1}^{h-1} 2^{jn}\left(\frac{n+1}{2}\right)^{h-j-1}\right) - \left(\frac{n-1}{2}\right)\left(\sum_{i=1}^{h-1} n^i\left(\frac{n+1}{2}\right)^{h-i-1}\right)$$

$$= \left(\frac{n-1}{2}\right)\left(\sum_{j=1}^{h-1} 2^{jn}\left(\frac{n+1}{2}\right)^{h-j-1}\right) - \left(n - \left(\frac{n+1}{2}\right)\right)\left(\sum_{i=1}^{h-1} n^i\left(\frac{n+1}{2}\right)^{h-i-1}\right)$$

$$= \left(\frac{n-1}{2}\right)\left(\sum_{j=1}^{h-1} 2^{jn}\left(\frac{n+1}{2}\right)^{h-j-1}\right) - \left(\sum_{i=1}^{h-1} n^{i+1}\left(\frac{n+1}{2}\right)^{h-i-1}\right) + \left(\sum_{i=1}^{h-1} n^i\left(\frac{n+1}{2}\right)^{h-i}\right)$$

$$= \left(\frac{n-1}{2}\right)\left(\sum_{j=1}^{h-1} 2^{jn}\left(\frac{n+1}{2}\right)^{h-j-1}\right) - \left(\sum_{i=2}^{h} n^{i+1}\left(\frac{n+1}{2}\right)^{h-i-1}\right) + \left(\sum_{i=1}^{h-1} n^i\left(\frac{n+1}{2}\right)^{h-i}\right)$$

$$= \left(\frac{n-1}{2}\right)\left(\sum_{j=1}^{h-1} 2^{jn}\left(\frac{n+1}{2}\right)^{h-j-1}\right) + n\left(\frac{n+1}{2}\right)^{h-1} - n^h$$

$$= d(h) - n^h.$$

This immediately implies

$$\mathrm{reg}(U_{p,h}^{\mathrm{MAJ}}) \leq d(h) = n\left(\frac{n+1}{2}\right)^{h-1} + \left(\frac{n-1}{2}\right)\sum_{j=1}^{h-1} 2^{jn}\left(\frac{n+1}{2}\right)^{h-j-1}.$$

Essentially the same argument applies to bound $\mathrm{reg}(U_{0,h}^{\mathrm{MAJ}})$. More precisely, again consider the case in which $h > 1$, we will show

$$\mathrm{reg}(U_{0,h}^{\mathrm{MAJ}}) \leq \left(\frac{n-1}{2}\right)\sum_{j=0}^{h-1} 2^{jn}\left(\frac{n+1}{2}\right)^{h-j-1}$$

$$= \left(\frac{n-1}{2}\right)\left(\frac{n+1}{2}\right)^{h-1} + \left(\frac{n-1}{2}\right)\sum_{j=1}^{h-1} 2^{jn}\left(\frac{n+1}{2}\right)^{h-j-1}$$

$$= \left(n - \left(\frac{n+1}{2}\right)\right)\left(\frac{n+1}{2}\right)^{h-1} + \left(\frac{n-1}{2}\right)\sum_{j=1}^{h-1} 2^{jn}\left(\frac{n+1}{2}\right)^{h-j-1}$$

$$= n\left(\frac{n+1}{2}\right)^{h-1} - \left(\frac{n+1}{2}\right)^{h} + \left(\frac{n-1}{2}\right)\sum_{j=1}^{h-1} 2^{jn}\left(\frac{n+1}{2}\right)^{h-j-1}$$

$$= d(h) - \left(\frac{n+1}{2}\right)^{h}.$$

We perform precisely the same analysis used to bound $\mathrm{reg}(U_{p,h}^{\mathrm{MAJ}})$, with the only change being the fact that when $M_T \in \{0,1\}^{n^h}$ is now constructed, we can now conlude that $wt(M_T) \leq n^h - \left(\frac{n+1}{2}\right)^h$, where $wt(M_T)$ denotes the number of 1s (the weight) of M_T. This follows because, for any $x \in T \subseteq U_{0,h}$ the value of the root node must be 0, by definition. For any node to have value 0, the majority of the elementary children of that node must have value 0 (because the function being evaluated at each node is MAJ). Due to the fact that each node has n elementary children, this requires that any node with value 0 has at least $\frac{n+1}{2}$ (recall that, by assumption, n is odd) elementary children with value 0. In particular, the majority of the elementary children of the root node must have value 0. Moreover, for each elementary child of the root node that has value 0, the majority of its children must have value 0. Continuing in this fashion until we reach the elementary leaves, we conclude that at least $\left(\frac{n+1}{2}\right)^h$ variables that are inputs to the elementary leaves must have value 0, and so at most $n^h - \left(\frac{n+1}{2}\right)^h$ have value 1, which shows the claimed bound on $wt(M_T)$. Therefore, when we construct α by $\alpha_E = M_T$ and $\alpha_{[m]\setminus E} = \beta$, we now have

$$\mathrm{reg}(U_{0,h}^{\mathrm{MAJ}}) \leq wt(\alpha)$$

$$= wt(\alpha_E) + wt(\alpha_{[m]\setminus E})$$

$$= wt(M_T) + wt(\beta)$$

$$\leq \left(n^h - \left(\frac{n+1}{2}\right)^h\right) + \left(d(h) - n^h\right)$$

$$= d(h) - \left(\frac{n+1}{2}\right)^h.$$

Finally, to bound $\mathrm{reg}(U_{1,h}^{\mathrm{MAJ}})$, simply notice that by Lemma 19

$$\mathrm{SM}(U_{0,h}^{\mathrm{MAJ}}) = \mathrm{SM}(U_{1,h}^{\mathrm{MAJ}}) \ \forall h \geq 1.$$

Lemma 1(b) then immediately implies

$$\mathrm{reg}(U_{1,h}^{\mathrm{MAJ}}) = \mathrm{reg}(U_{0,h}^{\mathrm{MAJ}}) \ \forall h \geq 1.$$

□

We now conclude that, for appropriately chosen input size, $RFS_{n,h}^{\mathrm{MAJ}}$ is versatile.

Lemma 21. *Let $n = 2^k - 1$ for any positive integer k, then $RFS_{n,h}^{MAJ}$ is $\left(\frac{n+1}{2}\right)^h$-versatile on $U_{p,h}^{MAJ}$. Moreover, $U_{p,h}^{MAJ}$ is a critical algebraic set.*

Proof. By the assumed form of n, Lemma 17 immediately allows us to conclude

$$\mathrm{MAJ}(x) = e_{\frac{n+1}{2}}(x) \ \forall x \in \mathbb{F}_2^n.$$

Clearly, $\deg(e_{\frac{n+1}{2}}) = \frac{n+1}{2}$, and so Lemma 16 immediately implies

$$\mathrm{reg}(U_{p,h}) \geq n\left(\frac{n+1}{2}\right)^{h-1} + \left(\frac{n-1}{2}\right)\sum_{j=1}^{h-1} 2^{jn}\left(\frac{n+1}{2}\right)^{h-j-1}$$

$$\mathrm{reg}(U_{0,h}), \mathrm{reg}(U_{1,h}) \geq \left(\frac{n-1}{2}\right)\sum_{j=0}^{h-1} 2^{jn}\left(\frac{n+1}{2}\right)^{h-j-1}.$$

By Lemma 20,

$$\mathrm{reg}(U_{p,h}) \leq n\left(\frac{n+1}{2}\right)^{h-1} + \left(\frac{n-1}{2}\right)\sum_{j=1}^{h-1} 2^{jn}\left(\frac{n+1}{2}\right)^{h-j-1}$$

$$\mathrm{reg}(U_{0,h}), \mathrm{reg}(U_{1,h}) \leq \left(\frac{n-1}{2}\right) \sum_{j=0}^{h-1} 2^{jn} \left(\frac{n+1}{2}\right)^{h-j-1}.$$

Therefore,

$$\mathrm{reg}(U_{p,h}) = n\left(\frac{n+1}{2}\right)^{h-1} + \left(\frac{n-1}{2}\right) \sum_{j=1}^{h-1} 2^{jn} \left(\frac{n+1}{2}\right)^{h-j-1}$$

$$\mathrm{reg}(U_{0,h}), \mathrm{reg}(U_{1,h}) = \left(\frac{n-1}{2}\right) \sum_{j=0}^{h-1} 2^{jn} \left(\frac{n+1}{2}\right)^{h-j-1}.$$

Finally,

$$\mathrm{reg}(U_{p,h}) - \mathrm{reg}(U_{0,h}) = \mathrm{reg}(U_{p,h}) - \mathrm{reg}(U_{0,h})$$

$$= n\left(\frac{n+1}{2}\right)^{h-1} + \left(\frac{n-1}{2}\right) \sum_{j=1}^{h-1} 2^{jn} \left(\frac{n+1}{2}\right)^{h-j-1} - \left(\frac{n-1}{2}\right) \sum_{j=0}^{h-1} 2^{jn} \left(\frac{n+1}{2}\right)^{h-j-1}$$

$$= n\left(\frac{n+1}{2}\right)^{h-1} - \left(\frac{n-1}{2}\right)\left(\frac{n+1}{2}\right)^{h-1}$$

$$= \left(n - \left(\frac{n-1}{2}\right)\right)\left(\frac{n+1}{2}\right)^{h-1}$$

$$= \left(\frac{n+1}{2}\right)^{h}.$$

Therefore, $RFS_{n,h}^{\mathrm{MAJ}}$ is $\left(\frac{n+1}{2}\right)^h$-versatile on $U_{p,h}^{\mathrm{MAJ}}$. To see that $U_{p,h}^{\mathrm{MAJ}}$ is a critical algebraic set, simply notice that, as shown in the proof of Lemma 16,

$$\deg(r_{U_{p,h}^{\mathrm{MAJ}}}) \leq (2^n - n) \sum_{j=1}^{h-1} 2^{(j-1)n} d^{h-j},$$

where $d = \frac{n+1}{2}$.

By the above,

$$\deg(r_{U_{p,h}^{\mathrm{MAJ}}}) \geq (2^n - n) \sum_{j=1}^{h-1} 2^{(j-1)n} d^{h-j} = a(\overline{U_{p,h}^{\mathrm{MAJ}}}),$$

and so

$$\deg(r_{U_{p,h}^{\mathrm{MAJ}}}) = a(\overline{U_{p,h}^{\mathrm{MAJ}}}),$$

which, by definition implies that $U_{p,h}^{\mathrm{MAJ}}$ is a critical algebraic set.

□

Next, we exhibit another class of functions such that the lower bound on regularity in Lemma 16 is tight. Consider any $g \in \mathbb{F}_2[x_1, \ldots, x_n]$ and let $d = \deg(g)$. $V_0 = g^{-1}(0)$ and $V_1 = g^{-1}(1)$ denote the preimages of 0 and 1, respectively. For any $k \times n$ matrix A with entries in $\mathbb{F}_2$, let $\phi_A : \mathbb{F}_2^n \to \mathbb{F}_2^k$ denote the linear map defined by A. We say a function g is *well-mixed* if, for every $n-d+1 \times n$ matrix A, $\frac{V_0}{\ker\phi_A} \not\cong \mathbb{F}_2^{n-d+1}$ and $\frac{V_1}{\ker\phi_A} \not\cong \mathbb{F}_2^{n-d+1}$. We then have the following.

Lemma 22. *For any positive integers n, h, let $g \in \mathbb{F}_2[x_1, \ldots, x_n]$ be well-mixed. Let $d = deg(g)$ and let $RFS_{n,h}^g : \mathbb{F}_2^m \to \mathbb{F}_2$ denote the recursive Fourier sampling function with g. Then*

$$reg(U_{p,h}^g) \leq nd^{h-1} + (n-d)\sum_{j=1}^{h-1} 2^{jn} d^{h-j-1}$$

$$reg(U_{0,h}^g), reg(U_{1,h}^g) \leq (n-d)\sum_{j=0}^{h-1} 2^{jn} d^{h-j-1}.$$

Proof. Before proceeding with the proof, we briefly remark that this Lemma could be proven by use of the inclusion matrix, in a similar manner to the proof of Lemma 20, shown above. We provide an different proof to illustrate an alternate method of bounding regularity.

We show this claim by induction on h. First, consider the case in which $h = 1$. Clearly, $U_{p,1}^g = \mathbb{F}_2^n$, and so $\mathrm{reg}(U_{p,1}^g) = n$. We now show $\mathrm{reg}(U_{0,1}^g), \mathrm{reg}(U_{1,1}^g) \leq n - d$. First, consider $\mathrm{reg}(U_{1,1}^g)$. Begin by noticing that, by Lemma 1(b), this is equivalent to showing that

$$x^\alpha \in \mathrm{LM}(U_{1,1}^g)\ \forall \alpha \text{ such that } \deg(x^\alpha) > n - d.$$

Due to the fact that, for any algebraic set V, $\mathrm{LM}(V)$ is an ideal (of the semigroup of

monomials) and because, for any $V \subseteq \mathbb{F}_2^n$, $x_j^2 \in \mathrm{LM}(V)\ \forall j$, the above is equivalent to showing

$$x^\alpha \in \mathrm{LM}(U_{1,1}^g)\ \forall \alpha \text{ such that } \deg(x^\alpha) = n - d + 1 \text{ and } x^\alpha \text{ is multilinear.}$$

To see this, consider any multilinear monomial x^α where $\deg(x^\alpha) = n - d + 1$. Let $J = \{j : \alpha_j = 1\}$. For any $x \in \mathbb{F}_2^n$, let $x_J \in \mathbb{F}_2^{n-d+1}$ denote the substring at positions indexed by J. The key observation is that, because g is well-mixed, there is a $b \in \mathbb{F}_2^{n-d+1}$ such that for every $x \in U_{1,1}^g$, $x_J \neq b$. To see this, let $\phi_A : \mathbb{F}_2^n \to \mathbb{F}_2^{n-d+1}$ denote the unique linear map such that $\phi_A(x) = x_J\ \forall x \in \mathbb{F}_2^n$. Then, because g is well-mixed, we have, by definition, that

$$\frac{U_{1,1}^g}{\ker(\phi_A)} \ncong \mathbb{F}_2^{n-d+1}$$

and so $\exists b \in \mathbb{F}_2^{n-d+1}$ such that, for every $x \in U_{1,1}^g$, $\phi_A(x) \neq b$, as claimed. Fix any such b.

For $k \in [n - d + 1]$, let J_k denote the k^{th} element of J (in the natural order), and consider the polynomial $f_\alpha \in \mathbb{F}_2[x_1, \ldots, x_n]$, where $f_\alpha = \prod_{k=1}^{n-d+1}(x_{J_k} + b_k + 1)$. We then have $f_\alpha \in I(U_{1,1}^g)$. This holds because, for any $x \in U_{1,1}^g$, $x_J = \phi_A(x) \neq b$, and so $\exists k \in [n - d + 1]$ such that $x_{J_k} \neq b_k$. For each k, we have $x_{J_k}, b_k \in \mathbb{F}_2$ and so if $x_{J_k} \neq b_k$, then $x_{J_k} = b_k + 1$. Therefore, for any $x \in U_{1,1}^g$, $\exists k \in [n - d + 1]$ such that $x_{J_k} = b_k + 1$, and so f_α vanishes on $U_{1,1}^g$. Clearly, $\mathrm{lm}(f_\alpha) = x^\alpha$, and so

$$x^\alpha \in \mathrm{LM}(U_{1,1}^g)\ \forall \alpha \text{ such that } \deg(x^\alpha) = n - d + 1 \text{ and } x^\alpha \text{ is multilinear,}$$

as desired. Therefore, $\mathrm{reg}(U_{1,1}^g) \leq n - d$. By a precisely symmetric argument, $\mathrm{reg}(U_{0,1}^g) \leq n - d$.

Next, we consider the case in which $h > 1$. Consider $U_{1,h}^g$. Let $r(h) = (n -$

$d) \sum_{j=0}^{h-1} 2^{jn} d^{h-j-1}$. We wish to show

$$\mathrm{reg}(U_{1,h}^{g}) \leq r(h).$$

For the same reason as above, it is equivalent to show

$$x^{\alpha} \in \mathrm{LM}(U_{1,h}^{g}) \ \forall \alpha \text{ such that } \deg(x^{\alpha}) = r(h) + 1 \text{ and } x^{\alpha} \text{ is multilinear.}$$

Consider any multilinear monomial x^{α}, where $\deg(x^{\alpha}) = r(h) + 1$. Let $J = \{j : \alpha_j = 1\}$. Consider the recursive Fourier sampling tree. For each child t of the root of the tree, say that t is *heavy* if at least $r(h-1)+1$ of the variables in the subtree rooted at t appear in J (that is to say, there are at least $r(h-1)$ variables x_j, such that $j \in J$ and x_j is a variable that appears at one of the leaves of the subtree rooted at t). Moreover, say that t is *very heavy*, if at least $r(h-1) + d^{h-1} + 1$ of the variables in the subtree rooted at t appear in J. Due to the fact that $\deg(x^{\alpha}) = r(h)+1$, it must be the case that at least one of the following two statements is true:

(1):At least one of the children of the root is very-heavy.

(2):At least $n - d + 1$ of the children of the root are heavy.

To see this, assume, for contradiction, that neither of these statements are true. Then at most $n - d$ of the children of the root are heavy, and none of the children of the root are very heavy. We then have

$$\deg(x^{\alpha}) \leq (n-d)(r(h-1) + d^{h-1}) + (2^n - (n-d))(r(h-1))$$

$$= (n-d)d^{h-1} + 2^n r(h-1)$$

$$= (n-d)d^{h-1} + 2^n(n-d) \sum_{j=0}^{(h-1)-1} 2^{jn} d^{(h-1)-j-1}$$

$$= (n-d)\left(d^{h-1} + \sum_{j=0}^{h-2} 2^{(j+1)n} d^{h-(j+1)-1}\right)$$

$$= (n-d)\left(d^{h-1} + \sum_{j=1}^{h-1} 2^{jn} d^{h-j-1}\right)$$

$$= (n-d)\left(\sum_{j=0}^{h-1} 2^{jn} d^{h-j-1}\right)$$

$$= r(h)$$

$$< r(h) + 1$$

$$= \deg(x^\alpha).$$

This contradiction immediately allows us to conclude that at least one of the above statements are true.

We now conclude that $x^\alpha \in \mathrm{LM}(U^g_{1,h})$. We first consider the case in which statement (1) holds. Let t denote an arbitrary very-heavy child of the root of the recursive Fourier sampling tree. Let x^β denote the multilinear monomial consisting of the product of all variables that are in the subtree rooted at t that appear in x^α. Clearly, $x^\beta | x^\alpha$, and so it suffices to show that $x^\beta \in \mathrm{LM}(U^g_{1,h})$. Due to the fact that t is very-heavy, we have,

$$\deg(x^\beta) \geq r(h-1) + d^{h-1} + 1 \geq \mathrm{reg}(U_{p,h-1}) + 1,$$

where the first inequality follows from the definition of a very-heavy child, and the second inequality follows from the induction hypothesis. Let $\widetilde{x}$ denote the portion of the input x within the subtree rooted at t. The key observation is that, since the subtree rooted at t corresponds to an instance of the recursive Fourier sampling problem of height $h-1$, we must have $\widetilde{x} \in \widetilde{V} \cong U_{p,h-1}$ (where $\widetilde{V}$ is simply $U_{p,h-1}$ with variables renamed $\widetilde{x}$). Since $\deg(x^\beta) > \mathrm{reg}(U_{p,h-1}) = \mathrm{reg}(\widetilde{V})$, we have, by the definition of regularity, that $x^\beta \in \mathrm{LM}(\widetilde{V})$, and so $\exists f_\beta$ (which only contains variables in $\widetilde{x}$) such that $x^\beta = \mathrm{lm}(f_\beta)$ and $f_\beta \in I(\widetilde{V})$.

Therefore, f_β vanishes on every $\widetilde{x} \in \widetilde{V}$, and so it must also vanish on every $x \in U^g_{1,h}$ because, by construction, if $x \in U^g_{1,h}$, then $\widetilde{x} \in \widetilde{V}$ and f_β only consists of variables in $\widetilde{x}$. Therefore, $x^\beta \in \mathrm{LM}(U^g_{1,h})$, which implies that $x^\alpha \in \mathrm{LM}(U^g_{1,h})$, as desired.

Next, we consider the case in which statement (2) holds. Let $\sigma : U^g_{1,h} \to \mathbb{F}_2^n$ be defined such that, for any $x \in U^g_{1,h}$, $\sigma(x)$ is defined such that the i^{th} position of $\sigma(x)$ is equal to the value of the i^{th} elementary child of the root when the input to the recursive Fourier sampling problem is x. Let $t_1, \ldots, t_{n-d+1}$ denote an arbitrary collection of (distinct) heavy children of the root of the recursive Fourier sampling tree. Let $\widetilde{\sigma} : U^g_{1,h} \to \mathbb{F}_2^{n-d+1}$ be defined such that, for any $x \in U^g_{1,h}$, $\widetilde{\sigma}(x)$ is defined such that the i^{th} position of $\widetilde{\sigma}(x)$ is equal to the value of t_i. In other words, the function σ simply encodes the values of all elementary children and $\widetilde{\sigma}$ encodes the values of the heavy children of interest. The key observation is that, because g is well-mixed, there is a $b \in \mathbb{F}_2^{n-d+1}$, such that, for every $x \in U^g_{1,h}$, $\widetilde{\sigma}(x) \neq b$. To see this, notice that, by definition, if $x \in U^g_{1,h}$, then $\sigma(x) \in U^g_{1,1}$. Moreover, due to the linear structure of the promise, there is a linear map $\phi : \mathbb{F}_2^n \to \mathbb{F}_2^{n-d+1}$ such that

$$\phi(\sigma(x)) = \widetilde{\sigma}(x) \ \forall x \in U^g_{1,h}.$$

Due to the fact that g is well-mixed,

$$\frac{U^g_{1,1}}{\ker(\phi_A)} \ncong \mathbb{F}_2^{n-d+1},$$

and so the existence of b follows from an identical argument as in the $h = 1$ case above. Fix such a b.

For $i \in [n-d+1]$, let x^{β_i} denote the monomial consisting of all variables in the subtree rooted at t_i that appears in x^α. Let $x^\beta = \prod_i x^{\beta_i}$. Clearly $x^\beta | x^\alpha$ and so it suffices to show $x^\beta \in \mathrm{LM}(U^g_{1,h})$. Notice that, for each i,

$$\deg(x^{\beta_i}) \geq r(h-1) + 1 \geq \mathrm{reg}(U_{1,h-1}) + 1, \mathrm{reg}(U_{0,h-1}) + 1,$$

where the first inequality follows from the fact that t_i is heavy, and the second inequality follows from the induction hypothesis. By the same argument that applied in case (1) above, we conclude that, for each i, there is a polynomial f_{β_i} such that $\mathrm{lm}(f_{\beta_i}) = x^{\beta_i}$ and $f_{\beta_i} \in I(U^g_{b_i+1,h-1})$. To be clear, the bound on the degree of x^{β_i} implies that x^{β_i} is a leading monomial of both the algebraic set isomorphic to $U^g_{1,h-1}$ and the algebraic set isomorphic to $U^g_{0,h-1}$ (where the isomorphism is simply the trivial renaming of variables), we choose $f_{\beta_i} \in I(U^g_{b_i+1,h-1})$ specifically to make the next stage of the construction work.

We now consider the polynomial $f_\beta = \prod_i f_{\beta_i}$. Clearly, $\mathrm{lm}(f_\beta) = x^\beta$. Moreover, we have $f_\beta \in I(U^g_{1,h})$. To see this, notice that, for every $x \in U^g_{1,h}$, $\widetilde{\sigma}(x) \neq b$, and so, for every $x \in U^g_{1,h}$, there must be at least one i such that $\widetilde{\sigma}(x)_i \neq b_i$. Since $\widetilde{\sigma}(x)_i, b_i \in \mathbb{F}_2$, if $\widetilde{\sigma}(x)_i \neq b_i$, then $\widetilde{\sigma}(x)_i = b_i + 1$. Therefore, for every $x \in U^g_{1,h}$, there must be at least one i such that f_{β_i} vanishes at x (because f_{β_i} vanishes whenever the portion of x in the subtree rooted at t_i has value $b_i + 1$ at node t_i). Due to the fact that f_β is the product of the f_{β_i}, if at least one of the f_{β_i} vanish, then f_β vanishes. This implies that $f_\beta \in I(U^g_{1,h})$, which in turn implies that $x^\beta \in \mathrm{LM}(U^g_{1,h})$ which in turn implies that $x^\alpha \in \mathrm{LM}(U^g_{1,h})$.

The above argument shows that $\mathrm{reg}(U^g_{1,h}) \leq r(h)$ for any $h > 1$, given the induction hypothesis. It is easy to see that this argument is precisely symmetric with respect to $U^g_{1,h}$ and $U^g_{0,h}$ and so we immediately also conclude $\mathrm{reg}(U^g_{1,h}) \leq r(h)$. An essentially identical argument shows $\mathrm{reg}(U^g_{p,h}) \leq r(h) + d^h$, with the only changes being the fact that statement (2) now becomes "At Least $n+1$ children of the root are heavy", and the analysis of the case in which statement (2) holds no longer relies on the fact that g is well-mixed, but instead the fact that, due to the linearity constraint, given any collection of $n + 1$ children of the root, there is at least one tuple of values that violates the promise.

□

This immediately allows us to conclude that, for any well-mixed g, $RFS^g_{n,h}$ is versatile, as shown in the following lemma.

Lemma 23. *For any positive integers n, h, let $g \in \mathbb{F}_2[x_1, \ldots, x_n]$ be well-mixed. Let $d =$*

deg(g) and let $RFS^g_{n,h} : \mathbb{F}_2^m \to \mathbb{F}_2$ *denote the recursive Fourier sampling function with g. Then* $RFS^g_{n,h}$ *is* d^h*-versatile on* $U^g_{p,h}$ *and* $U^g_{p,h}$ *is a critical algebraic set.*

Proof. Combining the bounds from Lemma 16 and Lemma 22, we have

$$\mathrm{reg}(U^g_{p,h}) = nd^{h-1} + (n-d)\sum_{j=1}^{h-1} 2^{jn} d^{h-j-1}$$

$$\mathrm{reg}(U^g_{0,h}), \mathrm{reg}(U^g_{1,h}) = (n-d)\sum_{j=0}^{h-1} 2^{jn} d^{h-j-1}.$$

Therefore,

$$\mathrm{reg}(U^g_{p,h}) - \mathrm{reg}(U^g_{0,h}) = \mathrm{reg}(U^g_{p,h}) - \mathrm{reg}(U^g_{0,h}) = nd^{h-1} + (n-d)\sum_{j=1}^{h-1} 2^{jn} d^{h-j-1} - (n-d)\sum_{j=0}^{h-1} 2^{jn} d^{h-j-1}$$

$$= nd^{h-1} + (n-d)d^{h-1}$$

$$= d^h.$$

To see that $U^g_{p,h}$ is a critical algebraic set, simply notice that, as shown in the proof of Lemma 16,

$$\deg(r_{U^g_{p,h}}) \leq (2^n - n)\sum_{j=1}^{h-1} 2^{(j-1)n} d^{h-j}.$$

By the above,

$$\deg(r_{U^g_{p,h}}) \geq (2^n - n)\sum_{j=1}^{h-1} 2^{(j-1)n} d^{h-j} = a(\overline{U^g_{p,h}}),$$

and so

$$\deg(r_{U^g_{p,h}}) = a(\overline{U^g_{p,h}}),$$

which, by definition implies that $U^g_{p,h}$ is a critical algebraic set.

□

We now show that a certain natural function, the generalized inner product function, is well-mixed, and therefore the corresponding version of recursive Fourier sampling is versatile. For any positive integer n and any $d|n$, let $\mathrm{GIP}_{n,d} : \mathbb{F}_2^n \to \mathbb{F}_2$ be defined such that

$$\mathrm{GIP}_{n,d} = x_1 \cdots x_d + x_{d+1} \cdots x_{2d} + \ldots + x_{n-d+1} \cdots x_n.$$

Notice that the ordinary inner product function simply corresponds to the case in which $d = 2$.

Lemma 24. *For any positive integers d, n such that $d|n$, and $n \geq d(2^{d^2} + d - 1)$, the function $GIP_{n,d} : \mathbb{F}_2^n \to \mathbb{F}_2$ is well-mixed. Moreover, the function $RFS_{n,h}^{GIP_{n,d}}$ is d^h-versatile on $U_{p,h}^{GIP_{n,d}}$ and $U_{p,h}^{GIP_{n,d}}$ is a critical algebraic set.*

Proof. We begin by showing that, for any positive integers d, n that satisfy the above requirements, the function $\mathrm{GIP}_{n,d} : \mathbb{F}_2^n \to \mathbb{F}_2$ is well-mixed. To do this, it clearly suffices to show that, for any $(n - d + 1) \times n$ matrix A, $\exists t^0, t^1 \in \mathbb{F}_2^{n-d+1}$ such that, for $x \in \mathbb{F}_2^n$, $Ax = t^0 \Rightarrow \mathrm{GIP}_{n,d}(x) = 0$ and $Ax = t^1 \Rightarrow \mathrm{GIP}_{n,d}(x) = 1$. We begin by noting that it suffices to show this claim only for A of a certain very special form. Let $\phi_A : \mathbb{F}_2^n \to \mathbb{F}_2^{n-d+1}$ denote the linear map corresponding to multiplication by the matrix A. Begin by noting that this claim trivially holds when A is not full rank (simply set t^0 and t^1 to be any element not in the image of ϕ_A) and so it suffices to consider only the case in which A is full rank. Next, it suffices to only consider the case in which A is in reduced row echelon form, because, for any invertible $(n - d + 1) \times (n - d + 1)$ matrix L, $Ax = t$ if and only if $(LA)x = Lt$, and so if the claim holds for every A in reduced row echelon form, then it holds for every A. Divide the n input variables $x_1, \ldots, x_n$ into blocks of size d, where each block consists of the d variables that appear in a single term of the $\mathrm{GIP}_{n,d}$ polynomial. Due to the fact that $\mathrm{rank}(A) = n - d + 1$ and that A is in reduced row echelon form, there are precisely $d - 1$ columns of A that do not have a leading 1. It suffices to only consider the case in which each of these $d - 1$ columns appear as one of the rightmost $d(d - 1)$ columns of A, because,

due to the symmetry of the generalized inner product function, the variables can be relabled such that these columns always correspond to variables that appear in the rightmost $d-1$ blocks, and hence rightmost $d(d-1)$ columns.

Therefore, in order to show that $\mathrm{GIP}_{n,d}$ is well-mixed, it suffices to show that, for any $(n-d+1) \times n$ matrix A, where $\mathrm{rank}(A) = n-d+1$, A is in reduced row echelon form, and the $d-1$ columns of A that do not contain a leading 1 appear within the rightmost $d(d-1)$ columns, $\exists t^0, t^1 \in \mathbb{F}_2^{n-d+1}$ such that, for $x \in \mathbb{F}_2^n$, $Ax = t^0 \Rightarrow \mathrm{GIP}_{n,d}(x) = 0$ and $Ax = t^1 \Rightarrow \mathrm{GIP}_{n,d}(x) = 1$.

Consider such a matrix A. We now construct t^0 and t^1 with the required properties. Let $y_1, \ldots, y_{d-1}$ denote the x_i that correspond to columns of A that do not have leading 1s, in the natural order. Let $r \leq d$ denote the value such that y_1 is in the r^{th} block from the right; that is to say, r is the minimal value such that all y_m are in the rightmost r blocks. For $i \in [n-d+1]$, and $j \in \{0,1\}$, let t_i^j denote the value of the i^{th} position of t^j.

Begin by noticing that there is a setting of $t_{n-dr}^j, \ldots, t_{n-d+1}^j$ such that, for any such t^j and any $x \in \mathbb{F}_2^n$, $Ax = t^j \Rightarrow x_{n-dr+1} \cdots x_{n-d(r-1)} + \ldots + x_{n-d+1} \cdots x_n = 0$. In other words, there is a way to set the last $dr-d+1$ values of t^j such that, for any x that satisfies $Ax = t^j$, it must be the case that the sum of the rightmost r terms of $\mathrm{GIP}_{n,d}$ is 0. To show this, we will construct the setting of the last $dr-d+1$ values of t^j in a collection of stages, where the values set in the l^{th} stage will force the l^{th} block (from the right) to evaluate to 0. Begin by considering the rightmost block of variables. Let k denote the number of y_m such that y_m correspond to columns in the rightmost block of variables; that is to say, $y_{d-k}, \ldots, y_{d-1}$ are the variables that correspond to the columns within the rightmost block that do not have leading 1s. There is a setting of the last $d-k$ values of t^j such that, for any x that satisfies $Ax = t^j$, we have $x_{n-d+1} \cdots x_n = 0$. To see this, notice that, due to the form of A, the only non-zero entries of A in the last $d-k$ rows are in the last d columns, which correspond precisely to the variables in the rightmost block. Therefore, the last $d-k$ values of Ax are completely determined by the last d values of x. In order to have $x_{n-d+1} \cdots x_n = 1$, it must

be the case that $x_{n-d+1} = \ldots = x_n = 1$, and so there is only 1 setting of these rightmost d variables such that $x_{n-d+1} \cdots x_n = 1$. On the other hand, there are $2^{d-k} \geq 2^{d-(d-1)} \geq 2 > 1$ distinct choices of the last $d-k$ values of t^j, from which it immediately follows that there is at least some setting of the last $d-k$ values of t^j such that, for any x that satisfies $Ax = t^j$, we do not have $x_{n-d+1} \cdots x_n = 1$, which then implies $x_{n-d+1} \cdots x_n = 0$. Fix any such setting of the last $d-k$ values of t^j.

In general, in the l^{th} stage, for each l such that $1 < l \leq r$, we consider the l^{th} block of variables (counting from the right). Within the first $l-1$ stages, we have set every t_i^j such that row i of matrix A has a leading 1 in a column corresponding to a variable in one of the rightmost $l-1$ blocks. This setting forces each of these $l-1$ blocks to evaluate to 0. We now force the l^{th} rightmost block to evaluate to 0 by appropriately setting all t_i^j such that row i of matrix A has a leading 1 in a column corresponding to a variable in block l. To be precise, let k denote the number of y_m that correspond to variables in block l, and let k' denote the number of y_m that appear in the rightmost $l-1$ blocks. Again, due to the form of A, the only non-zero entries in the $d-k$ rows in question are in the last dl columns, and so the corresponding $d-k$ values of Ax are completely determined by the last dl values of x. Again, there is only a single setting of the d values of x in block l such that block l evaluates to 1. Moreover, there are only $2^{k'}$ settings of the $d(l-1)$ values of x in the rightmost $l-1$ blocks which satisfy the constraint imposed by the t_i^j fixed in earlier stages. This follows from the fact that the $(d(l-1)-k') \times (d(l-1))$ submatrix of A corresponding to these constraints has rank $d(l-1)-k'$ and hence nullity k'. Therefore, there are precisely $2^{k'}$ distinct settings of the last dl values of x that both satisfy all earlier constraints and cause block l to evaluate to 1. Moreover, there are 2^{d-k} choices of the portion of t_j currently being set. Due to the fact that $k + k' \leq d-1$ (as there are only a total of $d-1$ variables y_m), we again conclude that there is a setting of the relevant portion of t^j such that, for any x that satisfies $Ax = t^j$, the l^{th} block evaluates to 0, as required.

The above argument shows that all of the rightmost r blocks can be forced to evaluate

to 0, by an appropriate setting of a portion of t^j. Next, we show that, similarly, for any $l > r$, the l^{th} rightmost block can be forced to evaluate to 0 by an appropriate setting of another portion of t^j. To be precise, consider the l^{th} rightmost block of variables, for any $l > r$. Due to the fact that every column of A that does not have a leading 1 appears among the rightmost r blocks, we conclude that every column corresponding to the l^{th} block has a leading 1. Consider the submatrix of A consiting of the d for which the leading 1 of that row appears in one of the columns corresponding to block l. The only non-zero entries in this submatrix appear in two parts. First, in the columns corresponding to block l, the submatrix is simply the $d \times d$ identity matrix. Secondly, there are non-zero entries in certain columns indexed by the y_m. In other words, this submatrix expresses the constraint that the values of x in block l are some affine combination of the y_m. To be precise, let $z_1, \ldots, z_d$ denote the d values of x that appear in block l, and let v denote the d values of t^j that correspond to rows in the submatrix in question. Then there is a $d \times (d-1)$ matrix B such that $z = By + v$. Let $\phi_B : \mathbb{F}_2^{d-1} \to \mathbb{F}_2^d$ denote the linear map corresponding to multiplication by B. As before, the key observation is that there is only a single setting of z such that block l evaluates to 0; however, there are 2^d choices of v, and $|\text{Im}(\phi_B)| \leq 2^{d-1}$, from which it immediately follows that there is a choice of v such that, for any z that satisfies $z = By + v$, it must be the case that block l evaluates to 0.

Therefore, to produce t^0, we simply use the first construction above to set the last $dr - d + 1$ values of t^0 in such a way as to force the last r blocks to evaluate to 0, and then use the second construction above to set the remaining values of t^0 in such a way as to force all other blocks to evaluate to 0.

To produce t^1, slightly more work is needed. We next show that, given a collection of 2^{d^2} blocks, all of which are not among the rightmost r blocks, it is possible to set the appropriate values of t^1 is such a way as to assure that exactly one of these blocks evaluates to 1, and all other blocks evaluate to 0. To see this, simply consider, as above, the constraint imposed by A on the variables in each block l. To be precise, let $z^l = (z_1^l, \ldots, z_d^l)$ denote

the d values of x that appear in block d, $v^l = (v^l_1, \ldots, v^l_d)$ denote the d values of t^1 that correspond to the rows of A that have a leading one in a column corresponding to block l and B^l denote the $d \times (d-1)$ matrix such that $z^l = B^l y + v^l$. As there are 2^{d^2} blocks in question, but only $2^{d(d-1)}$ distinct $d \times (d-1)$ matrices (with entries in $\mathbb{F}_2$), there must be some particular $d \times (d-1)$ matrix B such that at least 2^d blocks l have $B^l = B$. Fix any such B and let L denote a collection of precisely 2^d blocks l such that $B^l = B$. The key observation is that the portion of t^1 corresponding to the collection of blocks L can be set in such a way so that exactly one block in L evaluates to 1. This can be accomplished by setting the collection of v^l such that $l \in L$ to the 2^d elements of $\mathbb{F}_2^d$. This works because, for any setting of y, the collection of z^l, for $l \in L$ will all be distinct (as each $z^l = By + v^l$ and the v^l are distinct) and exactly one of the z^l will be all 1s (as there are 2^d possible setting of each z^l, so each appears exactly once).

Therefore, to produce t^1, we then simply use the first construction above to set the last $dr - d + 1$ values of t^0 in such a way as to force the last r blocks to evaluate to 0, then the second construction above to set the portion of t^1 that corresponds to every block not in L to force all such blocks to evaluate to 0, and finally the third construction above to set the portion of t^1 that corresponds to the blocks in L to force exactly one block in L to evaluate to 1. Due to the fact that, by assumption, $n \geq d(2^{d^2} + d - 1)$, there are at least $2^{d^2} + d - 1$ blocks, and so this construction is possible.

We have thus shown that, for any positive integers d, n that satisfy the above requirements, the function $\mathrm{GIP}_{n,d} : \mathbb{F}_2^n \to \mathbb{F}_2$ is well-mixed. By Lemma 23, it immediately follows that the function $RFS^{\mathrm{GIP}_{n,d}}_{n,h}$ is d^h-versatile on $U^{\mathrm{GIP}_{n,d}}_{p,h}$ and $U^{\mathrm{GIP}_{n,d}}_{p,h}$ is a critical algebraic set.

□

4.1.3 Polynomial Degree

Using the results of the previous section, we now prove very strong statements about the degree of any polynomial that computes, or even one-sided agrees with, the recursive Fourier

sampling problem.

Theorem 7. *For any positive integers k, h, Let $n = 2^k - 1$ and let $RFS^{MAJ}_{n,h}$ denote the recursive Fourier sampling function with majority. Then $\nexists g \in \mathbb{F}_2[x_1, \ldots, x_m]$ such that $deg(g) < \left(\frac{n+1}{2}\right)^h$ and $g(x) = RFS^{MAJ}_{n,h}(x)\ \forall x \in U^{MAJ}_{p,h}$. Moreover, if any $g \in \mathbb{F}_2[x_1, \ldots, x_m]$ such that $deg(g) < \left(\frac{n+1}{2}\right)^h$ vanishes everywhere on $U^{MAJ}_{0,h}$, it vanishes everywhere on $U^{MAJ}_{1,h}$.*

Proof. By Lemma 21, $RFS^{\mathrm{MAJ}}_{n,h}$ is $\left(\frac{n+1}{2}\right)^h$-versatile on $U^{\mathrm{MAJ}}_{p,h}$ and $U^{\mathrm{MAJ}}_{p,h}$ is a critical algebraic set. The first claim of the theorem is an immediate consequence of Lemma 6 and the second claim is an immediate consequence of Lemma 7.

□

Theorem 8. *For any positive integers d, n, h such that $d|n$, and $n \geq d(2^{d^2} + d - 1)$, Let $RFS^{GIP_{n,d}}_{n,h}$ denote the recursive Fourier sampling function with generalized inner product. Then $\nexists g \in \mathbb{F}_2[x_1, \ldots, x_m]$ such that $deg(g) < d^h$ and $g(x) = RFS^{GIP_{n,d}}_{n,h}(x)\ \forall x \in U^{GIP_{n,d}}_{p,h}$. Moreover, if any $g \in \mathbb{F}_2[x_1, \ldots, x_m]$ such that $deg(g) < d^h$ vanishes everywhere on $U^{GIP_{n,d}}_{0,h}$, it vanishes everywhere on $U^{GIP_{n,d}}_{1,h}$.*

Proof. By Lemma 24, $RFS^{\mathrm{GIP}_{n,d}}_{n,h}$ is d^h-versatile on $U^{\mathrm{GIP}_{n,d}}_{p,h}$ and $U^{\mathrm{GIP}_{n,d}}_{p,h}$ is a critical algebraic set. The first claim of the theorem is an immediate consequence of Lemma 6 and the second claim is an immediate consequence of Lemma 7.

□

4.1.4 Towards a Circuit Lower Bound

In the previous section, an extremely strong lower bound was given on the lowest degree polynomial over $\mathbb{F}_2$ that computes (or even non-trivially one-sided agrees with) the recursive Fourier sampling function on the promise. In this section, we discuss partial results towards a lower bound on the size of a constant depth circuit that computes the recursive Fourier sampling function, as well as what sort of additional results would allow these partial results to be extended to prove such a lower bound. We begin by defining the circuit class of

interest. Let n denote, as before, the size of the secret at each node of the recursive Fourier sampling tree, and h denote the height of the recursive Fourier sampling tree. We consider circuits that consist of *AND*, *OR*, and *NOT* gates, where the fan-in of the *AND* and *OR* gates is unbounded, the size of the circuit (the total number of gates) is at most $2^{O(poly(n))}$, and the depth of the circuit (the number of gates on the longest path from the input to the output) is a constant (independent of n and h). This circuit class is of interest due to the fact that proving a lower bound against it (that is to say, proving that no circuit of this form can compute the recursive Fourier sampling function on its promise), would immediately imply the existence of an oracle A such that $\mathrm{BQP}^A \not\subset \mathrm{PH}^A$. This follows due to the relationship between such circuits and the polynomial hierarchy ([FSS84],[Yao85]) and the fact that there is an efficient quantum algorithm for the recursive Fourier sampling problem ([BV93],[Aar03],[Joh08]), when $h = O(\log n)$. Such a bound is at least plausable as the trivial circuit (which simply computes the recursive Fourier sampling in the brute force, level-by-level way, in which each subproblem is solved by solving n subproblem one level down) has size $2^{\theta(n^h)}$, which, when $h = \theta(\log n)$ is, of course, not $2^{O(poly(n))}$.

One reasonable approach to proving such a lower bound would be to apply a variant of the Razborov-Smolensky method [Raz87],[Smo87]. We begin by briefly sketching the main idea of the Razborov-Smolensky method, specialized to $\mathbb{F}_2$. We consider a (total) function $g : \mathbb{F}_2^m \to \mathbb{F}_2$, where $m = 2^{O(poly(n))}$. We wish to show that no circuit C of the above form, of size at most $2^{O(poly(\log m))} = 2^{O(poly(n))}$, can compute the function g. The key observation is that there is an $f \in \mathbb{F}_2[x_1, \ldots, x_m]$ where $\deg(f) = O(poly(n))$ such that f agrees with C almost everywhere, and so if it can be shown that g is not well approximated by a low degree polynomial, it immediately follows that g is not actually computed by C. To show that a particular g cannot agree almost everywhere with a low degree polynomial, it suffices to show that g has the property that, on any set $R \subseteq \mathbb{F}_2^n$, if g is represented on R by a polynomial of degree at most d, then every function $q : R \to \mathbb{F}_2$ is represented on R by a polynomial of degree not much higher than d. This suffices because if g agrees almost

everywhere with a low degree polynomial f, then there is a very large set R on which every function $q : R \to \mathbb{F}_2$ is represented by a low-degree polynomial; a straightforward counting of the number of functions of that form and the number of low-degree polynomials shows that this is impossible.

The main idea behind the lower bound on the polynomial degree of recursive Fourier sampling, shown in the previous sections, is that there are functions g such that $RFS^g_{n,h}$ has the property that there is a large gap between the regularity of the promise, $\mathrm{reg}(U^g_{p,h})$, and the regularities of the preimages of 0 and 1, $\mathrm{reg}(U^g_{0,h})$ and $\mathrm{reg}(U^g_{1,h})$. In other words, there are functions on $U^g_{p,h}$ which can only be computed by relatively high degree polynomials, whereas every function on $U^g_{0,h}$ and $U^g_{1,h}$ can be computed by relatively low degree polynomials. It then follows that $RFS^g_{n,h}$ itself cannot be computed on $U^g_{p,h}$ by a low degree polynomial, because if it were, then every function on $U^g_{p,h}$ would be computable by a low degree polynomial.

While this is very similar to the observation made in the Razborov-Smolensky method, there is one crucial difference: due to the fact that the promise $U^g_{p,h}$ is extremely small, one cannot conclude, via a straightforward counting argument, that there is a function on $U^g_{p,h}$ that requires a high degree polynomial; instead, this fact was shown via an analysis of the structure of this algebraic set. It is the very fact that such an analysis is possible that gives hope that this technique could be extended to prove the desired circuit lower bound. To be precise, to prove the desired circuit lower bound, it would suffice to show that, not merely is it the case that $RFS^g_{n,h}$ is $\omega(poly(n))$-versatile on $U^g_{p,h}$, as already shown, but in fact $RFS^g_{n,h}$ is $\omega(poly(n))$-versatile on R for any sufficiently large $R \subseteq U^g_{p,h}$. This would suffice because, if $RFS^g_{n,h}$ had this property, then it could not be the case that $RFS^g_{n,h}$ is well approximated by a low degree polynomial on $U^g_{p,h}$, from which it would then follow that $RFS^g_{n,h}$ is not computed by a small circuit on $U^g_{p,h}$. In fact, something substantially weaker would suffice: one only needs to consider the case in which R is of the form $U^g_{p,h} \cap V(f_1, \ldots, f_k)$ where each $f_i \in \mathbb{F}_2[x_1, \ldots, x_m]$ satisfies $\deg(f_i) = O(poly(n))$. In other words, one only needs to consider the case in which R is a large subset of $U^g_{p,h}$ such that R is the intersection

of $U^g_{p,h}$ with an algebraic set that is the set of common zeros of a collection of low degree polynomials. This suffices because, much as was done in Braverman's proof of the Linial-Nisan conjecture [Bra09], one can consider the structure of the set of points on which a small circuit agrees with the low degree polynomial produced by the Razborov-Smolensky method. To be precise, consider applying the Razborov-Smolensky method to a *AND* of a collection of polynomials $p_1, \ldots, p_k \in \mathbb{F}_2[x_1, \ldots, x_m]$ where $\deg(p_i) = O(poly(n))\ \forall i$. This *AND* of low degree polynomials is well approximated by a single $p' \in \mathbb{F}_2[x_1, \ldots, x_m]$, given by the product of a collection of a small number of randomly chosen sums of the p_i. Moreover, the output of the *AND* of $p_1, \ldots, p_k$ agrees with p' precisely on $V(p'(1+p_1), \ldots, p'(1+p_k))$. Repeating this process for every gate in the circuit, from the bottom up, yields an algebraic set of the form $V = V(f_1, \ldots, f_k)$ where $\deg(f_i) = O(poly(n))\ \forall i$, where, on V, each gate individually agrees with its approximating polynomial. To be clear, this algebraic set V is a (possibly proper) subset of the set of points on which the circuit agrees with the overall approximating polynomial, due to the fact that a local mistake (that is to say, a point at which an individual gate disagreeing with its approximating polynomial) may not propagate through the entire circuit to yield a global mistake (that is to say, a point at which the circuit disagrees with the approximating polynomial); however, the extremely simple form of V makes it a natural choice for performing the required analysis of regularity.

While the current analysis falls short of being able to prove the type of circuit lower bound needed for the desired relatived separation result, it does produce some interesting partial results. For example, consider any circuit C consisting of an *OR* of a collection $p_1, \ldots, p_k \in \mathbb{F}_2[x_1, \ldots, x_m]$ where $\deg(p_i) \leq d = O(poly(n))\ \forall i$. Circuits of this type are interesting as it can easily be seen that if one can prove that such a circuit cannot be a good approximator with one-sided error of the recursive Fourier sampling problem on its promise (where we say C is a good approximator with one-sided error if C outputs 1 everywhere on $U^g_{1,h}$ and outputs 0 almost everywhere on $U^g_{0,h}$) this would immediately yield the existance of an A such that $\mathrm{BQP}^A \not\subset \mathrm{AM}^A$. The existing analysis does provide some insight into the

behavior of any such circuit on the promise, though it, unfortunately, falls short of proving the required lower bound. To be precise, by noting that the set of points on which C outputs one is given by $V = \cup_i V(1+p_i)$, and applying Lemma 8, one can immediately conclude that, for any g such that $RFS^g_{n,h}$ is δ-versatile on $U^g_{p,h}$,

$$\mathrm{SM}(U^g_{p,h} \cap V, j) = \mathrm{SM}(U^g_{0,h} \cap V, j) = \mathrm{SM}(U^g_{1,h} \cap V, j)\ \forall j \leq \delta - d.$$

This is, by itself, a very strong statement about the structure of the set of points on which any such circuit C evaluates to 1. Moreover, due to the fact that, by Lemma 1(c), the size of any algebraic set is equal to the size of the set of standard monomials of that set, the above claim also yields a (weak) statement about the relationship between the sizes of $U^g_{0,h} \cap V$ and $U^g_{1,h} \cap V$.

4.2 VC Dimension

In this section, we answer an open question posed in [MR15]. We begin with a few definitions. We begin by recalling several key results from that paper.

Lemma 25. *[MR15](Thm.2.2) For any $C \subseteq \{0,1\}^n$, $reg(C) \leq VC(C)$.*

It was shown that, if C_{i_j} denotes the value of the i^{th} element of C in the j^{th} position, then

Lemma 26. *[MR15](Prop.6.1) For any $C \subseteq \{0,1\}^n$, $reg(C) = 1$ precisely when the matrix*

$$\begin{pmatrix} 1 & C_{1,1} & \cdots & C_{1,n} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ 1 & C_{m,1} & \cdots & C_{m,n} \end{pmatrix}$$

has rank $m = |C|$.

They then asked if there was a similar simple characterization of when $\text{reg}(C) = r$, for $r > 1$, which would be highly desirable as any such characterization would, by the above lemma, provide a characterization of sets with VC dimension at least r. We show the following.

Theorem 9. *A set* $C \subseteq \{0,1\}^n$ *has* $reg(C) = r$ *if and only if* r *is the smallest positive integer such that* $rank_{\mathbb{F}_2}\mathcal{M}(C, \binom{[n]}{\leq r}) = |C|$.

Proof. By Lemma 3,

$$h^a(C,d) = \text{rank}_{\mathbb{F}_2}\mathcal{M}(C, \binom{[n]}{\leq d}).$$

By definition, $\text{reg}(C)$ is the minimum r such that $h^a(C,r) = |C|$.

□

Remark 2. *It is straightforward to see that [MR15](Prop.6.1) is a special case of the above theorem.*

Bibliography

[Aar03] S. Aaronson, *Quantum lower bound for recursive Fourier sampling*, Quantum Information and Computation (2003), 3(2):165-174.

[Aar10] S. Aaronson, *BQP and the polynomial hierarchy*, In Stoc '10: : Proceedings of the forty-second annual ACM symposium of Theory of computing (2010), 141-150.

[Ajt83] M. Ajtai, Σ_1^1-*Formulae on finite structure*, APAL (1983).

[AGM02] N. Alon, O. Goldreich, and Y. Mansour, *Almost k-wise independence versus k-wise independence*, Electronic Colloquium on Computational Complexity, Report TR02-048 (2002).

[BBC+98] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf, *Quantum lower bounds by polynomials*, In IEEE Symposium on Foundations of Computer Science (1998), 352-361.

[BV93] E. Bernstein and U. Vazirani, *Quantum complexity theory*, In STOC '93: Proceedings of the twenty-fifth annual ACM symposium of Theory of computing (1993), 11-20.

[BV97] E. Bernstein and U. Vazirani, *Quantum complexity theory*, SIAM J. Comput. (1997), 26(5):1411-1473.

[Bou07] J. Bourgain, *On the construction of affine extractors*, Geometric and Functional Analysis (2007), 17(1):33-57.

[Bra09] M. Braverman, *Poly-logarithmic independence fools AC0 circuits*, IEEE Conference on Computational Complexity (2009), 3-8.

[CSV84] A. K. Chandra, L. Stockmeyer, and U. Vishkin, *Constant depth reducibility*, SIAM Journal on Computing (1984), 13:423-439.

[CT13] G. Cohen and A. Tal, *Two structural results for low degree polynomials and applications*, ECCC (2013), TR. No. 145.

[DGW07] Z. Dvir, A. Gabizon, and A. Wigderson, *Extractors and rank extractors for polynomial sources*, In FOCS '07 (2007).

[Dvi12] Z. Dvir, *Extractors for varieties*, Computational Complexity (2012), 21(4):515-572.

[Eis02] D. Eisenbud, *The Geometry of Syzygies*, (2002).

[Fel07] B. Felszeghy, *Grobner Theory of Zero Dimensional Ideals with a View Towards Combinatorics*, Budapest University (2007), Ph. D. thesis.

[FSS84] M. Furst, J. Saxe, and M. Sipser, *Parity, circuits, and the polynomial time hierarchy*, Mathematical Systems Theory (1984), 17:13-27.

[GR05] A. Gabizon and R. Raz, *Deterministic extractors for affine sources over large fields*, In Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (2005), 407-418.

[GRS04] A. Gabizon, R. Raz, and R. Shaltiel, *Deterministic extractors for bit-fixing sources by obtaining an independent seed*, In Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science (2004), 394-403.

[Gre12] B. Green, *On (not) computing the Möbius function using bounded depth circuits*, http://arxiv.org/pdf/1103.4991.pdf (2012).

[Has86] J. Hastad, *Computational limitations for small depth circuits*, MIT Press (1986), Ph. D. thesis.

[Joh08] B. Johnson, *Upper and lower bounds for recursive Fourier sampling*, University of California at Berkeley (2008), Ph. D. thesis.

[Joh11] B. Johnson, *The polynomial degree of recursive Fourier sampling*, Theory of Quantum Computation, Communication, and Cryptography in Lecture Notes in Computer Science (2011), 6519:104-112.

[KRVZ06] J. Kamp, A. Rao, S. Vadhan, and D. Zuckerman, *Deterministic extractors for small-space sources*, In Proceedings of the thirty-eighth annual ACM symposium on Theory of computing (2006), 691-700.

[KZ03] J. Kamp and D. Zuckerman, *Deterministic extractors for bit-fixing sources and exposure-resilient cryptography*, In Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science (2003).

[KLL84] R. Kannan, A. K. Lenstra, L. Lovasz, *Polynomial factorization and nonrandomness of bits of algebraic and some transcendental numbers*, STOC (1984), 191-200.

[Kat86] I. Katai, *Distribution of digits of primes in q-ary canonical form*, Acta Math (1986), 47(3-4):341-359.

[Kop11] S. Kopparty, *On the complexity of powering in finite fields*, In Stoc '11: : Proceedings of the forty-third annual ACM symposium of Theory of computing (2011), 489-498.

[LMN93] N. Linial, Y. Mansour, and N. Nisan, *Constant depth circuits, Fourier transform, and learnability*, J. ACM (1993), 40(3):607-620.

[LN90] N. Linial and N. Nisan, *Approximate inclusion-exclusion*, Cominatorica (1990), 10(4):349-365.

[Lov79] L. Lovász, *Combinatorial problems and exercises*, North-Holland, Amsterdam (1979), 13.31.

[MR15] S. Moran and C. Rashtchian, *Shattered sets and the Hilbert function*, ECCC (2015), TR. No. 15-189.

[Nis91] N. Nisan, *Pseudorandom bits for constant depth circuits*, Combinatorica (1991), 11(1):63-70.

[NW94] N. Nisan and A. Wigderson, *Hardness vs Randomness*, Journal of Computer and Systems Sciences (1994), 49(2):149-167.

[OSB99] A. Oppenheim, R. Schafer, and J. Buck, *Discrete-Time Signal Processing*, Prentice Hall (1999).

[PR08] D. Pintér and L. Rónyai, *On the Hilbert Function of Complementary Set Families*, Annales Univ. Sci. Budapest Sect. Comp. (2008), 29:175-198.

[Raz87] A. A. Razborov, *Lower bounds on the size of bounded depth circuits over a complete basis with logical addition*, Mathematical Notes (1987), 41(4):333-338.

[Smo87] R. Smolensky, *Algebraic methods in the theory of lower bounds for Boolean circuit complexity*, Proceedings of the nineteenth annual ACM symposium on Theory of computing (1987), 77-82.

[Smo93] R. Smolensky, *On Representations by Low-degree Polynomials*, FOCS (1993).

[TV00] L. Trevisan and S. Vadhan, *Extracting randomness from samplable distributions*, In Proceedings of the 41st Annual Symposium of Foundations of COmputer Science (2000), 32.

[Vaz86] U. V. Vazirani, *Randomness, Adversaries and Computation*, Ph.D. Thesis, EECS, UC Berkeley (1986).

[Yao82] A. C. Yao, *Theory and application of trapdoor functions*, IEEE Symposium on Foundations of Computer Science (1982), 80-91.

[Yao85] A. Yao, *Separating the polynomial-time hierarchy by oracles (preliminary version)*, In Proc. IEEE FOCS (1985), 1-10.