Communication Complexity of XOR Functions

A Thesis

Submitted to the
Tata Institute of Fundamental Research, Mumbai
for the Degree of
Doctor of Philosophy
in Computer Science

by

Nikhil Shekhar Mande

School of Technology and Computer Science
Tata Institute of Fundamental Research
Mumbai

August, 2018
Final Version Submitted in November, 2018

Declaration

This thesis is a presentation of my original research work. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgement of collaborative research and discussions. The work was done under the guidance of Professor Arkadev Chattopadhyay, at the Tata Institute of Fundamental Research, Mumbai.

[Candidate's name and signature]

In my capacity as supervisor of the candidate's thesis, I certify that the above statements are true to the best of my knowledge.

[Guide's name and signature]

Date:

Abstract

Given a boolean function $f : \{0,1\}^n \to \{0,1\}$ define the function $f \circ \mathsf{XOR}$ on $2n$ bits by $f \circ \mathsf{XOR}(x_1, \ldots, x_n, y_1, \ldots, y_n) = f(x_1 \oplus y_1, \ldots, x_n \oplus y_n)$. Such a function is called an XOR function. A natural communication game for such a function is as follows. Alice is given $x = (x_1, \ldots, x_n)$, Bob is given $y = (y_1, \ldots, y_n)$, and they jointly wish to compute $f \circ \mathsf{XOR}(x, y)$. They have unbounded computational power individually and wish to minimize the amount of communication between them on the worst-case input.

We study the communication complexity of XOR functions in various randomized models, and resolve several open questions in the areas of communication complexity, boolean circuit complexity and analysis of boolean functions.

1) We characterize the weakly unbounded-error communication complexity of XOR functions in terms of a certain approximation theoretic property of the outer function. We use this characterization to reprove several known results. Along the way, we also resolve some open questions in the area of analysis of boolean functions.

2) We prove a strong unbounded-error communication complexity lower bound for an easily describable function. We then use this to show a boolean circuit complexity class separation that has been open since the early nineties, and first explicitly asked in 2005. This also resolves a recent open problem in communication complexity by separating two communication complexity classes. We also prove a lower bound on the size required by any decision list of linear threshold functions to compute a simple XOR function, and prove unbounded-error communication complexity lower bounds for XOR functions when the outer function is symmetric.

3) Finally, we separate two randomized communication complexity classes in the 'number-on-forehead' model of multi-party communication. This also implies boolean circuit class separations.

To my parents.

Contents

Chapter 1

Introduction

1.1 Communication Complexity

Suppose we have a computational system comprising multiple processors, and the system wishes to perform a computation when the input is distributed amongst the processors. Our focus is on how efficiently the computation can be performed *in parallel*. That is, we are not interested in the amount of time each processor takes for its own computations, but the amount of *communication* required between the processors. Such a problem and many of its variants appear in various areas of computer science - network protocols, VLSI circuit design, data structures, communication complexity and boolean circuit complexity.

In a seminal work, Andrew Yao [Yao79] introduced the area of *communication complexity*, a subarea of theoretical computer science which deals with communication as a resource rather than number of operations during computation as in the Turing machine model. In the most basic model, two parties, say Alice and Bob, are given inputs $x \in X$ and $y \in Y$, respectively, for some domains X, Y. They wish to jointly evaluate a given two-party function, defined by $f : X \times Y \to \{-1, 1\}$, on the input (x, y). We assume that Alice and Bob have unbounded computational power, and wish to minimize the number of bits communicated between them. This allows us to view communication as the main resource, and ignore the computation time of each individual party (processor). They communicate using a set of rules agreed upon in advance. In other words, they follow a *protocol* for computing f. The notion of correctness of this protocol may vary depending on our requirement. For instance, one of the most natural notions of correctness of the protocol is that Alice and Bob must output the correct answer on each input. The cost of the protocol is considered to be the number of bits communicated on the worst-case input. It is easy to see that

it is always feasible for Alice to send her whole input x to Bob, and Bob can output the correct answer. The general question we address is whether or not there exist protocols with a cheaper cost.

As in the Turing machine world, one may ask what happens if we allow access to non-determinism, randomness etc. In the standard randomized model, Alice and Bob have access to unlimited *public* random bits, and wish to compute the target function f with probability at least 90% on all inputs. Are there functions hard to compute in the former (deterministic) model, but which are easy in the randomized model? A classic witness of a positive answer is the *Equality* function. Here, Alice and Bob are given two n-bit strings, say x and y, and wish to test whether $x = y$. It is not hard to show that in the deterministic model, any protocol for this function requires $n+1$ bits. Surprisingly, there exists a cheap randomized protocol which requires just 5 bits of communication to get the right answer with probability at least 90% on all inputs. We provide a sketch of this protocol in Section 1.4.

However, there exist functions which are hard to compute even in the randomized model. Two classic instances of this are the *Set Disjointness* (DISJ) and *Inner Product Modulo 2* (IP) functions. Babai, Frankl and Simon [BFS86] defined analogues of Turing machine classes in communication complexity. While a standard notion of efficiency in the Turing machine world corresponds to computability in polynomial time, Babai et al. argued that this notion of efficiency translates to that of polylogarithmic communication (in the length of the inputs to Alice and Bob) in the communication complexity world. Functions efficiently computable in the deterministic communication model mentioned above form the communication complexity class P^{cc} [1]. Functions efficiently computable in randomized model mentioned above form the class BPP. Thus, *Equality* separates P from BPP. Several other class separations are known in the communication complexity world, which are still open in the Turing machine world, for instance $\mathsf{P} \neq \mathsf{NP}$. However, there are also several larger classes against which explicit lower bounds remain unknown. For example, it is a long-standing open question [BFS86] to prove strong lower bounds against the polynomial hierarchy, for which IP has been identified as a natural target. Unfortunately we do not have strong lower bounds against even the second level. Thus a natural program is to understand these communication complexity classes better.

[1] Henceforth, we often drop the superscript cc since we exclusively deal with communication complexity classes. We also abuse notation and let $\mathcal{C}(F)$ denote the cost of the function F under the model (class) $\mathcal{C}$.

Another reason for studying communication complexity is motivated by the fact that several known circuit complexity lower bounds are proven by arguments that show lower bounds on the communication complexity of the target function under some partition of the inputs, for a suitable communication model. We elaborate on this aspect in later chapters, and indeed, we prove several circuit complexity lower bounds using communication lower bounds.

1.2 XOR Functions

The functions we use for our lower bounds and class separations are all 'XOR' functions. This class of functions is one of interest since it captures many natural functions, *Equality* for instance. As indicated in the previous section, the communication complexity of *Equality* has been studied widely under various models. The communication complexity of XOR functions also has various connections to Fourier analysis, additive combinatorics, approximation theory and boolean circuit complexity.

Towards formally defining XOR functions, we first introduce the notion of *composed functions*. Given functions $f_n : \{-1,1\}^n \to \{-1,1\}$ and $g_m : \{-1,1\}^m \to \{-1,1\}$,[2] define the *composed* function $f_n \circ g_m : \{-1,1\}^{nm} \to \{-1,1\}$ as follows. $f_n \circ g_m(x_{11},\ldots,x_{1m},x_{21},\ldots,x_{2m},\ldots,x_{n1},\ldots,x_{nm}) = f_n(g_m(x_1),g_m(x_2),\ldots,g_m(x_n))$. We often drop the subscripts when the arities of the constituent functions are clear. When g is a function on two bits, there are only two functions that it can be (up to negation of variables): AND and XOR. The communication complexity of AND functions has been widely studied [BFS86, KS92, BVdW07, She11a, She09b, She11b, RS10, BT16, BCH+16]. The class of all functions of the form $f \circ \mathsf{XOR}$ yields the class of XOR functions. The communication complexity of XOR functions has also received considerable interest of late [MO09, ZS09, LLZ11, LZ13, Zha14, HHL18, KMSY18].

Given a function $f : \{-1,1\}^n \to \{-1,1\}$, a natural communication game for the function $f \circ \mathsf{XOR}$ is as follows. Alice gets input $x_1,\ldots,x_n$, Bob gets input $y_1,\ldots,y_n$, and they wish to compute $f \circ \mathsf{XOR}(x,y)$ (which is defined as $f \circ \mathsf{XOR}(x,y) := f(x_1 \oplus y_1,\ldots,x_n \oplus y_n)$). The class of XOR functions is a natural one to consider while aiming to prove lower bounds, since we may view the input as distributed between two parties, and neither party has any information about the output given their own input. This motivates the program of studying the communication complexity of XOR functions. Several communication complexity problems, for instance

[2] Throughout this thesis, we interchangeably view the input and output domains as $\{-1,1\}^n$ and $\{-1,1\}$, and $\{0,1\}^n$ and $\{0,1\}$ respectively. -1 is identified with 'True', and 1 with 'False'.

the fabled log-rank conjecture [LS88], are still open for the restricted class of XOR functions. In this thesis, we prove lower bounds against XOR functions under various randomized models of communication. Our results also yield some new boolean circuit class separations, communication complexity class separations, and resolve some open questions in the area of analysis of boolean functions, all of which we elaborate on in later sections.

1.3 Models of Communication

In this section, we describe some models of communication of our interest, along with some examples. As mentioned in Section 1.1, Babai et al. [BFS86] argued that Turing machine complexity classes have natural analogues in the communication complexity world, where the notion of polynomial time as efficiency translates to that of polylogarithmic communication in the length of inputs to Alice and Bob. They also argued that the Turing machine class PP has two natural analogues: PP^{cc} and UPP^{cc}. Recall that in the class BPP, the correctness requirement is that the protocol should be correct with probability at least $1/2 + 2/5$ $(= 90\%)$ on all inputs.

PP protocols are probabilistic with the requirement that the protocol be correct with probability at least $1/2 + \epsilon$ on all inputs for some $\epsilon > 0$. However an additive term of $\log(1/\epsilon)$ is added to the number of bits communicated to yield the cost of the protocol. For a two-party function F, if there exists an $\epsilon > 0$ such that there are polylogarithmic cost PP protocols computing F, then we say $F \in \mathsf{PP}$. This is the weakly unbounded-error model, and clearly $\mathsf{BPP} \subseteq \mathsf{PP}$. The function *Set Disjointness* famously separates these two classes [BFS86, KS92, Raz92b].

In the UPP model,[3] the correctness requirement is the same as that in the PP model, but there is no additive $\log(1/\epsilon)$ charged to the cost of the protocol.

Examples

Below, we provide some examples of XOR functions efficiently computable by protocols in the BPP, PP and UPP models.

[3] In the UPP model, the random coins are assumed to be *private*. That is, Alice cannot view the outcome of Bob's random coin tosses and vice versa. Indeed, it is not hard to show that all functions have UPP protocols of constant cost if allowed public randomness. Throughout this thesis, we assume that we are dealing with *private* coin protocols whenever considering the UPP model, and public coin protocols for the BPP and PP models.

1. Recall the Equality function $\mathsf{EQ} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, defined by $\mathsf{EQ}(x,y) = 1$ if and only if $x = y$. Equality is an XOR function since it can be expressed as $\mathsf{NOR} \circ \mathsf{XOR}$. Consider the following protocol: Alice samples (publicly) 4 random strings r_1, r_2, r_3, r_4 of length n each. She then sends Bob $(\langle x, r_1\rangle \pmod 2, \langle x, r_2\rangle \pmod 2, \langle x, r_3\rangle \pmod 2, \langle x, r_4\rangle \pmod 2)$. Bob computes $(\langle y, r_1\rangle \pmod 2, \langle y, r_2\rangle \pmod 2, \langle y, r_3\rangle \pmod 2, \langle y, r_4\rangle \pmod 2)$ and outputs 1 if this agrees with Alice's message.

   If $x = y$, the protocol is always correct. If $x \neq y$, then $\Pr_r[\langle x, r_1\rangle \pmod 2 = \langle y, r_1\rangle \pmod 2] = 1/2$ and hence $\Pr[\langle x, r_i\rangle \pmod 2 = \langle y, r_i\rangle \pmod 2 \text{ for all } i \in \{1,2,3,4\}] < 10\%$. Thus, the above protocol describes a BPP protocol of cost 5.

2. Earlier in this section, we pointed out that DISJ (formally defined in Chapter 2) famously separates BPP from PP. However, this separation can also witnessed by an XOR function. Consider the Majority function, denoted $\mathsf{MAJ} : \{-1,1\}^n \to \{-1,1\}$, by $\mathsf{MAJ}(x_1, \ldots, x_n) = -1$ iff $\sum_{i=1}^n x_i < 0$. The function $\mathsf{MAJ} \circ \mathsf{XOR}$ was shown to be hard for BPP in [ZS09]. The following is an efficient PP protocol for $\mathsf{MAJ} \circ \mathsf{XOR}$. Alice and Bob sample (using public randomness) an input i to the top Majority gate uniformly at random. They then output $x_i \oplus y_i$. This protocol has constant communication and its probability of success is at least $1/2 + 1/2n$ for all inputs. Thus, $\mathsf{MAJ} \circ \mathsf{XOR}$ has efficient PP protocols.

3. A function $f : \{-1,1\}^n \to \{-1,1\}$ is said to be a *linear threshold function* if there exist integers $w_0, w_1, \ldots, w_n$ such that $f(x) = \mathsf{sgn}(w_0 + \sum_{i=1}^n w_i x_i)$. Denote the class of linear threshold functions by THR. Consider any function in $\mathsf{THR} \circ \mathsf{XOR}$, and the following *private* coin protocol for it. Alice samples an input i to the top linear threshold gate with probability proportional to w_i, and sends this index to Bob along with x_i. Bob outputs $\mathsf{sgn}(x_i \oplus y_i)$. It is not hard to show that this is a valid UPP protocol and its cost is polylogarithmic in n (see Claim 2.3.10).

1.4 Weakly Unbounded-Error Communication

In this section, we describe our work on the PP complexity of XOR functions with applications to BPP lower bounds and also to analysis of boolean functions.

The separation $\mathsf{PP} \subsetneq \mathsf{UPP}$ was observed by Sherstov [She08], based on the work of Goldmann, Håstad and Razborov [GHR92]. This separation was independently

proven by Buhrman, Vereshchagin and de Wolf [BVdW07], who used a different function and technique for the separation. Summarizing,

$$\mathsf{BPP} \subsetneq \mathsf{PP} \subsetneq \mathsf{UPP}.$$

Sherstov [She11a] introduced the *pattern matrix* method, which led to several works demonstrating PP lower bounds [Cha07, CA08, She09b], and subsequently UPP lower bounds [RS10, She11b, BT16, BCH+16] of composed functions where the inner function is AND. On the other hand, we are not aware of any systematic method of analyzing even the PP complexity of XOR functions before our work.

1.4.1 Our Work

PP Complexity

We prove a general theorem tightly characterizing the PP complexity of $f \circ \mathsf{XOR}$ in terms of how well f can be approximated by low weight polynomials. It is known [Kla07] that the PP complexity of a function F is tightly related to a combinatorial measure, called the *discrepancy*, of F, denoted $\mathsf{disc}(F)$ (see Definition 2.3.7).

Define the weight of a real polynomial to be the sum of the absolute values of its coefficients. We introduce a notion called the *polynomial margin* of a boolean function f, denoted $m(f)$, which captures the error in the best uniform approximation of f by polynomials of weight 1 (see Definition 2.4.8).

We show a tight relationship between the polynomial margin of f and the discrepancy $f \circ \mathsf{XOR}$, thus giving a tight characterization of the PP complexity of $f \circ \mathsf{XOR}$ in terms of the polynomial margin of f.

For the purpose of this Chapter, we do not formally define polynomial margin or discrepancy. The interested reader may refer to Chapter 3 for a formal statement. We choose to state the follwing theorem here since it is a fundamental building block for some of the following results (Theorems 1.4.2, 1.4.4, 1.4.7).

Theorem 1.4.1 (Polynomial Margin-Discrepancy theorem)**.** Let $f \to \{-1,1\}^n \to \{-1,1\}$.

$$m(f) \leq m(f \circ \mathsf{XOR}) \leq 4\mathsf{disc}(f \circ \mathsf{XOR}) \leq 4m(f).$$

It is significantly easier to analyze approximation theoretic properties of f, rather than analyzing the PP complexity of $f \circ \mathsf{XOR}$ from first principles. As evidence of this, we provide two applications of this theorem.

The first application tightly characterizes $\mathsf{PP}(f \circ \mathsf{XOR})$ when f is a symmetric function (its value only depends on the Hamming weight of the input). For a symmetric function $f : \{-1,1\}^n \to \{-1,1\}$, define its *spectrum* or *predicate* $D_f : \{0,1,\ldots,n\} \to \{-1,1\}$ by $D_f(i) = f(x)$ where $x \in \{-1,1\}^n$ is any string such that $|x| = i$ (here, $|x|$ denotes the number of -1's in x). Note that the spectrum (predicate) of a symmetric function is well defined. Define the *odd-even degree* of a symmetric function f, which we denote by $\mathsf{deg}_{oe}(f)$, to be $|i \in \{0,1,\ldots,n-2\} : D_f(i) \neq D_f(i+2)|$. Shi and Zhang [ZS09] conjectured that for symmetric f, the UPP complexity of $f \circ \mathsf{XOR}$ is essentially $\mathsf{deg}_{oe}(f)$.

We resolve a weak form of this conjecture by showing that when f is symmetric, $\mathsf{PP}(f \circ \mathsf{XOR})$ equals $\mathsf{deg}_{oe}(f)$ up to polylogarithmic factors (see Theorem 3.1.5).

Theorem 1.4.2. For symmetric $f : \{-1,1\}^n \to \{-1,1\}$,

$$\mathsf{PP}(f \circ \mathsf{XOR}) = \tilde{\Theta}(\mathsf{deg}_{oe}(f)).$$

Remark 1.4.3. The full conjecture was subsequently resolved independently by Hatami and Qian [HQ17], and Ada, Fawzi and Kulkarni [AFK17]. However, their proofs involve a reduction to symmetric AND functions, and use an involved result of Sherstov [She11b] in a black-box fashion and does not provide new insight about XOR functions. Moreover their proof techniques only apply for symmetric outer functions. Our proof technique on the other hand, crucially uses the Margin-Discrepancy equivalence, is from first principles and applies to non-symmetric outer functions as well.

Our next application of the Margin-Discrepancy connection involves a non-symmetric outer function. Recall that a function f is said to be a *linear threshold function* if there exist integers $w_0, w_1, \ldots, w_n$ such that $f(x) = \mathsf{sgn}(w_0 + \sum_{i=1}^{n} w_i x_i)$. It is well-known that there exists a linear threshold function on $k = O(n^2 \log n)$ variables, which we denote UTHR_k, such that any linear threshold function on n bits can be obtained by fixing inputs to UTHR_k suitably.

We prove that an XOR function has exponentially small discrepancy, where the outer function is a linear threshold function. It is not hard to show that such functions have efficient UPP protocols (see Claim 2.3.10). Thus, this gives a new proof of the separation between PP and UPP.

Theorem 1.4.4.

1. There exists a linear threshold function $f : \{-1,1\}^n \to \{-1,1\}$ such that $\mathsf{PP}(f \circ \mathsf{XOR}) = \Omega(n)$.

2. $\mathsf{PP}(\mathsf{UTHR}_n \circ \mathsf{XOR}) = \Omega(\sqrt{n})$.

Polynomial Measures of Symmetric Functions

The study of inapproximability of functions by low-degree polynomials is an active area of research. Define the ϵ-approximate degree (if not parametrized, ϵ is assumed to equal $1/3$) of a boolean function f, denoted by $\mathsf{deg}_\epsilon(f)$ to be the minimum degree of any polynomial p that approximates f uniformly to error ϵ (see Definition 2.4.3). In a seminal work, Paturi [Pat92] characterized the approximate degree of symmetric functions, building upon a celebrated result of Nisan and Szegedy [NS94], who gave tight bounds on the approximate degree of OR. Paturi's theorem has found several applications in various areas of theoretical computer science (see, for example, [Raz03, She09a, She11a, dW10, Spa08, BT15a]). Lower bounds on the sign degree (see Definition 2.4.2) of functions has also been widely studied, starting from the seminal work of Minsky and Papert [MP69]. A natural question that arises is whether or not there are analogous results when the approximating polynomials may have large degree, but have a small number of monomials or small weight. We prove such analogous results for symmetric functions, as we elaborate on later in this section.

A useful tool we develop in the course of proving our PP lower bounds is a lifting lemma (see Lemma 3.3.1), which allows us to analyze the polynomial margin of 'lifted' functions (see Equation 3.2). It is worthwile to note here that while standard approximation theory deals with hardness of approximating functions by *low degree* polynomials, the polynomials we are constrained to work with may have high degree but low *weight*. Also, it is not clear apriori how to view a symmetric (linear threshold) function as the lift of another symmetric (linear threshold) function.

Our lifting lemma, along with another technical tool (the projection lemma, see Lemma 3.3.4), allows us to overcome this hurdle. The lifting lemma translates degree-hardness properties of f to monomial-hardness properties of f^{op} (which is a certain lifted function obtained from f, formally defined in Equation 3.2). The proof of this lemma is based on ideas from [KP97]. The lifting lemma used along with the projection lemma has consequences in boolean analysis of symmetric functions as well, which we outline below.

The *signed monomial complexity* of a boolean function f, denoted by $\mathsf{mon}_{\pm}(f)$, is the minimum number of monomials required by any polynomial p to sign represent f on all inputs.

We resolve a conjecture of Zhang [Zha92] by tightly characterizing the signed monomial complexity of symmetric functions in terms of their odd-even degree.

Theorem 1.4.5 (Conjecture 1 in [Zha92])**.** For any symmetric f,

$$\mathsf{mon}_{\pm}(f) = 2^{\tilde{\Omega}(\mathsf{deg}_{oe}(f))}.$$

Ada, Fawzi and Hatami [AFH12] defined a measure of symmetric functions, denoted $r(f)$, that captures the position closest to the middle of the spectrum where there is a i-$(i+2)$ sign change (see Definition 2.4.11). Ada et al. showed that the *exact weight* of any symmetric function f is tightly characterized by $r(f)$. The main conjecture they posed was whether or not f's *approximate weight* was also bounded below by the same quantity. We resolve this conjecture in the affirmative.

Theorem 1.4.6 (Conjecture 1 in [AFH12])**.** For any symmetric f,

$$\log(\mathsf{wt}_{1/3}(f)) = \Omega(r(F)).$$

As observed by Ada et al., the resolution of this conjecture in conjunction with their main theorem yields several implications (cf. [AFH12]). One is the characterizing the *approximate monomial complexity* of symmetric functions, which is a natural analogue of Paturi's celebrated result [Pat92], which characterized the approximate degree of symmetric functions. Another application is the resolution of the log approximation-rank conjecture, which is the randomized analogue of the fabled log-rank conjecture [LS88], for symmetric XOR functions.

BPP Complexity

Using linear programming duality and the generalized discrepancy method, we give a simple alternate proof of the following result from [LS09b].

Theorem 1.4.7. For any boolean function f,

$$\mathsf{BPP}(f \circ \mathsf{XOR}) = \Omega(\log(\mathsf{wt}_{1/3}(f)).$$

Although Theorem 1.4.7 was known from [LS09b], to the best of our knowledge, ours is the first work to use the inequality to prove lower bounds for explicit functions.

Our lifting theorem in conjunction with Theorem 1.4.7 provides an alternate proof of a result of Shi and Zhang [ZS09] that states that $\mathsf{BPP}(f \circ \mathsf{XOR}) = \Omega(r(f))$ for any symmetric f.

The general framework of the proofs of our results are summarized in Figure 3.1.

1.5 Unbounded-Error Communication

In this section, we outline our results on unbounded-error communication complexity of XOR functions, with applications in boolean circuit complexity.

Sign Rank

Sign rank is a delicate but powerful notion, which has a matrix rigidity-like flavor. The sign rank of a $\{-1, 1\}$ valued matrix M is defined to be the minimum rank of a real valued matrix each of whose entries agrees in sign with the corresponding entry of M. Sign rank has found numerous applications in computer science in areas like communication complexity, boolean circuit complexity, and computational learning theory. Paturi and Simon [PS86] showed that the logarithm of the sign rank of a (communication) matrix is essentially equivalent to the UPP complexity of the underlying function.

Low-Depth Threshold Circuits

Understanding the computational power of constant-depth, unbounded fan-in threshold circuits is one of the most fundamental open problems in theoretical computer science. Despite several years of intensive research [HMP+93, HG91, GHR92, Raz92a, KP97, KP98, For02, FKL+01, AM05, RS10, HP10, HP15, KW16, CSS16], we still do not have strong lower bounds against depth-3 or depth-2 threshold circuits, depending on how we define threshold gates. The most natural definition of such a gate, denoted by $\mathsf{THR}_{\mathbf{w}}$, is one that computes a linear halfspace induced by the real weight vector $\mathbf{w} = (w_0, w_1, \ldots, w_n) \in \mathbb{R}^{n+1}$. In other words, on an input $x \in \{-1, 1\}^n$,

$$\mathsf{THR}_{\mathbf{w}}(x) = \mathsf{sgn}\left(w_0 + \sum_{i=1}^{n} w_i x_i\right).$$

We denote by THR the class of all functions expressible as a linear threshold function. Define the *majority* function $\mathsf{MAJ} : \{-1, 1\}^n \to \{-1, 1\}$ by $\mathsf{MAJ}(x) = \mathsf{sgn}(\sum_{i=1}^{n} x_i)$.

For classes of functions $\mathcal{C}$ and $\mathcal{D}$, denote by $\mathcal{C} \circ \mathcal{D}$ the class of all functions computable by polynomial-size depth-2 circuits where the top gate computes a function in $\mathcal{C}$ and the bottom gates compute functions in $\mathcal{D}$. Larger depth circuit classes are defined in a similar fashion. The seminal work of Minsky and Papert [MP69] showed that a simple function, Parity, is not in THR. While it is not hard to verify that Parity is in $\mathsf{THR} \circ \mathsf{THR}$ (in fact in $\mathsf{MAJ} \circ \mathsf{MAJ}$), an outstanding problem is to exhibit an explicit function that is not in $\mathsf{THR} \circ \mathsf{THR}$. This problem is now a well-identified frontier for research in circuit complexity. In particular, $\mathsf{THR} \circ \mathsf{THR}$ is one of the smallest known boolean circuit classes against which no strong lower bounds are known.

Goldmann, Håstad and Razborov [GHR92] proved several intersting results, yielding the following structure.

$$\mathsf{MAJ} \circ \mathsf{MAJ} = \mathsf{MAJ} \circ \mathsf{THR} \subsetneq \mathsf{THR} \circ \mathsf{MAJ} \subseteq \mathsf{THR} \circ \mathsf{THR} \subseteq \mathsf{MAJ} \circ \mathsf{MAJ} \circ \mathsf{MAJ}.$$

It raises the following two questions: how powerful is the class $\mathsf{THR} \circ \mathsf{MAJ}$ and how does one prove lower bounds on the size of such circuits?

Lower bounds against $\mathsf{MAJ} \circ \mathsf{MAJ}$ have been known since the work of Hajnal et al. [HMP+93]. Forster et al. [FKL+01] observed that strong lower bounds on the sign rank of the matrix corresponding to a convenient bi-partition of the input variables of a function f is sufficient for proving lower bounds on the size of $\mathsf{THR} \circ \mathsf{MAJ}$ circuits computing f. In a breakthrough work, Forster [For02] showed that IP has sign rank $2^{\Omega(n)}$ for the natural partition of input variables in which each part has n input variables. This, therefore, yielded an exponential separation between $\mathsf{THR} \circ \mathsf{MAJ}$ and $\mathsf{MAJ} \circ \mathsf{MAJ} \circ \mathsf{MAJ}$. This meant that at least one of the two containments $\mathsf{THR} \circ \mathsf{MAJ} \subseteq \mathsf{THR} \circ \mathsf{THR}$ and $\mathsf{THR} \circ \mathsf{THR} \subseteq \mathsf{MAJ} \circ \mathsf{MAJ} \circ \mathsf{MAJ}$ is strict. However, the question of which of these containments is strict is one towards which no progress was made until now. In particular, Amano and Maruoka [AM05] and Hansen and Podolskii [HP10] state that separating $\mathsf{THR} \circ \mathsf{MAJ}$ from $\mathsf{THR} \circ \mathsf{THR}$ would be an important step for shedding more light on the structure of depth-2 boolean circuits. However, as far as we know, there was no clear target function identified for the purpose of separating the two classes. We remark here that it is not a priori clear that these classes ought to be different, especially in light of Goldmann et al.'s result that $\mathsf{MAJ} \circ \mathsf{MAJ} = \mathsf{MAJ} \circ \mathsf{THR}$.

We, however, prove that $\mathsf{THR} \circ \mathsf{MAJ} \subsetneq \mathsf{THR} \circ \mathsf{THR}$, and elaborate on this in Section 1.5.1.

Communication Complexity Frontiers

A major goal, set by Babai et al. [BFS86], is to prove lower bounds against the polynomial hierarchy, for which the simple function IP has long been identified as a target. Unfortunately, it even remains open to exhibit a function that is not in the second level of the hierarchy.

The strongest communication complexity class against which we know how to prove explicit lower bounds is UPP (see Chapter 2 for formal definitions of communication complexity classes mentioned in this section). Razborov and Sherstov [RS10] showed that PH (in fact, $\Pi_2\mathsf{P}$) contains functions outside UPP, rendering the sign rank technique essentially useless to prove lower bounds against the second level. A natural question is to understand until where, between the first and second level, does the sign rank method suffice to prove lower bounds.

Indeed, there is a rich landscape of communication complexity classes below the second level as discussed in a recent, almost exhaustive survey by Göös, Pitassi and Watson [GPW18]. To motivate our contributions, we informally define MA protocols. Merlin, an all powerful prover, has access to Alice and Bob's inputs. He sends a (purported) proof string to Alice and Bob, who then run a randomized protocol to verify the proof. The protocol accepts an input if and only if the verification goes through. We say the protocol computes a function F if for all inputs to Alice and Bob, the probability of outputting the correct answer is at least $2/3$. The cost of the protocol on an input is the sum of the length of Merlin's proof string and the number of bits communicated between Alice and Bob. A function is said to be in the complexity class MA if there is such a protocol computing it with polylogarithmic worst-case cost (in the size of the input). For example, the function $\mathsf{OR} \circ \mathsf{EQ}$ can be seen to be in MA as follows: Merlin sends Alice and Bob the index of an input to the OR gate (if it exists) where EQ outputs -1, and Alice and Bob run an efficient randomized protocol for EQ to verify this. The class MA is a natural generalization of NP, and has received a lot of attention, starting with the work of [Kla03]. It is known that MA is strictly contained in UPP.

One could similarly define AM, but its power remains much less understood; we do not know any lower bounds against this class. A natural question to ask is for which classes does the sign rank method suffice to prove lower bounds? Can we come up with lower bound techniques for those classes for which the sign rank method fails to prove lower bounds?

We prove that P^{MA} contains functions with large sign rank, strongly resolving an open problem posed by Göös et al. [GPW18]. We define the class P^{MA} and elaborate on this result in Section 1.5.1.

1.5.1 Our Work

We consider the following easily describable function F_n: The input, of length $n = 2m\ell$, is split into two disjoint parts, $X \in \{-1,1\}^{m\ell}$ and $Y \in \{-1,1\}^{m\ell}$. X and Y are each further divided into ℓ disjoint blocks $X_1, \ldots, X_\ell, Y_1, \ldots, Y_\ell$, of length m each. The function F_n outputs -1 iff the largest index $i \in [\ell]$ for which $X_i = Y_i$ holds is an odd index. For the purpose of this thesis, we set $m = \ell^{1/3} + \log \ell$. We observe that F_n can be easily described as a decision list (see Definition 4.1.1) of Equalities. Decision lists are a natural class of functions to study and have widespread applications in learning theory, for example [Riv87, KS06, Kra06].

Another way of looking at our function is to view it as a composed function in the following way: consider a simple adaptation of the well-known $\mathsf{ODD\text{-}MAX\text{-}BIT}$ function on ℓ bits, which we denote by OMB^0_ℓ. The function OMB^0_ℓ outputs -1 precisely if the rightmost bit that is set to 1 occurs at an odd index. It is simple to observe that it is a linear threshold function:

$$\mathsf{OMB}^0_\ell(x) = -1 \iff \sum_{i=1}^{\ell} (-1)^{i+1} 2^i (1 + x_i) \geq 0.5.$$

It is not hard to verify that $F_n = \mathsf{OMB}^0_\ell \circ \mathsf{OR}_{\ell^{1/3}+\log \ell} \circ \mathsf{XOR}_2$.

We show a strong lower bound on the sign rank of M_{F_n}, where the rows of M_{F_n} are indexed by the inputs X, the columns by Y, and the (x,y)th entry is $F_n(x,y)$. We overload notation and refer to the sign rank of M_{F_n} as the sign rank of F_n.

Theorem 1.5.1. The function F_n has sign rank $2^{\Omega(n^{1/4})}$.

The fact that such a simple function has large sign rank allows us to settle two open problems, motivated earlier in this chapter. The first is a circuit complexity class separation (in this case, it helps to view F_n as an XOR function). The second is a communication complexity class separation (here it is more convenient to view F_n as a decision list).

We first observe that F_n can be computed by linear sized $\mathsf{THR} \circ \mathsf{THR}$ formulas. Next, we use the fact that sign rank lower bounds on f yield lower bounds on the size of any $\mathsf{THR} \circ \mathsf{MAJ}$ circuit computing f. Combined with Theorem 1.5.1, these observations yield the following separation.

Theorem 1.5.2. The function F_n witnesses the following exponential separation.

$$\mathsf{THR} \circ \mathsf{MAJ} \subsetneq \mathsf{THR} \circ \mathsf{THR}.$$

This resolves an open question in [AM05, HP10]. Along with Goldmann et al.'s result that $\mathsf{MAJ} \circ \mathsf{MAJ} = \mathsf{MAJ} \circ \mathsf{THR}$, Theorem 1.5.2 may be summarized in one sentence as follows: While weights at the bottom do not matter if the top is light, they do matter if the top is heavy.

Göös [Göö17] pointed out to us that F_n can be used to demonstrate another complexity class separation, this time in communication complexity. Göös et al. [GPW18] conjectured that the (potentially incomparable) classes $\mathsf{AM} \cap \mathsf{coAM}$ and $\mathsf{S_2P}$ (which we do not define here) contain functions of large sign rank. In a very recent work, Bouland et al. [BCH+16] showed that there is a *partial function* in $\mathsf{AM} \cap \mathsf{coAM}$ which has large sign rank, (partially) resolving the first conjecture.[4] We provide a strong confirmation of the second conjecture by exhibiting a *total* function in a sub-class of $\mathsf{S_2P}$ that has large sign rank.

In order to state our result, let us consider the complexity class P^{MA} that is contained in $\mathsf{S_2P}$. A function is in P^{MA} if it can be computed by deterministic protocols of polylogarithmic cost, where Alice and Bob have oracle access to any function in MA. The function F_n under the natural input partition (recall that it can be expressed as a decision list of equalities) can be efficiently solved by P^{MA} protocols by an appropriate binary search, and querying an $\mathsf{OR} \circ \mathsf{EQ}$ oracle at each step (see Protocol 1).

We thus prove the following as a consequence of Theorem 1.5.1.

Theorem 1.5.3. The function F_n witnesses the following communication complexity class separation.

$$\mathsf{P}^{\mathsf{MA}} \not\subseteq \mathsf{UPP}.$$

Our result thus strongly confirms the second conjecture of Göös et al. by exhibiting the first *total function* in a complexity class contained, plausibly strictly, in $\Pi_2\mathsf{P}$, that has large sign rank.

On the other hand, it is known that $\mathsf{P}^{\mathsf{NP}} \subsetneq \mathsf{UPP}$ and $\mathsf{MA} \subsetneq \mathsf{PP} \subsetneq \mathsf{UPP}$. This places the class P^{MA} right on the frontier of our current knowledge of lower bounds in communication complexity.

Proof idea: We now outline the proof of Theorem 1.5.1. We are guided by the communication complexity theoretic interpretation of sign rank, due to Paturi and

[4] It still remains unknown if there are *total* functions in $\mathsf{AM} \cap \mathsf{coAM}$ that have large sign rank.

Simon [PS86], who showed that the logarithm of the sign rank of a communication matrix is essentially equal to the unbounded-error communication complexity (UPP) of the underlying function. Figure 1.1 describes a general passage from the problem of proving a lower bound on the sign rank of a function $f \circ \mathsf{XOR}$ to a sufficient problem of proving an approximation theoretic hardness property of f: namely f has no good 'mixed margin' representation by low weight polynomials. The key difference between our work and previous works [RS10, She11b, BT16, BCH+16] is in the nature of the approximation theoretic problem that we end up with. While all these previous works had to rule out good *low degree* representations, our program stipulates us to rule out good *low weight* representations of otherwise *unrestricted degree*.

Our main technical contribution is an approximation theoretic result, which shows that the function $\mathsf{OMB}^0 \circ \mathsf{OR}$ is inapproximable by low weight polynomials of *unrestricted degree* in a certain relaxed sense (see Section 4.4 for a description of our notion of approximation). We prove this by a novel combination of ideas, sketched in Figure 1.2, that differs entirely from the analysis in earlier works. One may view this result as a hardness amplification result, albeit specific to the function OMB^0. We start with the function OMB^0 which has no low weight 'worst-case margin' representation when the degree of the approximating polynomial is bounded [Bei94]. We show that on composition with large fan-in OR gates, the function $\mathsf{OMB}^0 \circ \mathsf{OR}$ becomes '*mixed margin*'-inapproximable by low weight polynomials, even with *unrestricted degree*. We believe this result to be of independent interest in the area of analysis of Boolean functions and approximation theory.

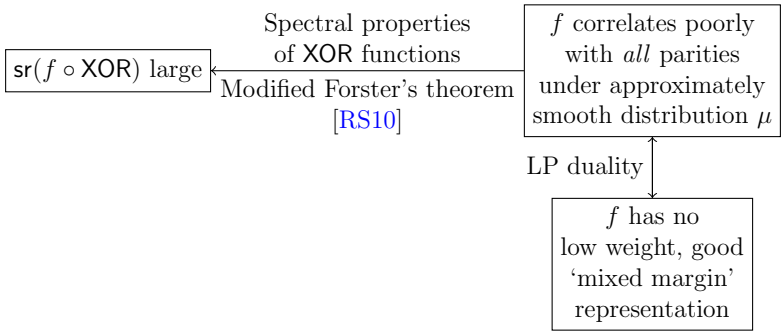

Figure 1.1: Approximation theoretic hardness of f implies large sign rank of $f \circ \mathsf{XOR}$.

Symmetric XOR Functions

We also consider the UPP complexity $f \circ \mathsf{XOR}$ when f is symmetric and periodic. Refer to Definition 2.1.11 for a formal definition of MOD functions.

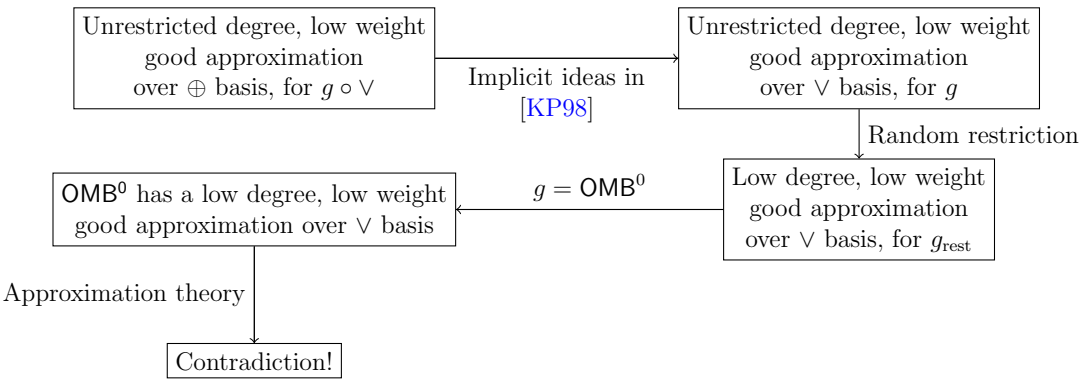

Figure 1.2: Approximation theoretic analysis

We show the following.

Theorem 1.5.4. For any integer $3 \leq m \leq n^{1/2-\epsilon}$,

$$\mathsf{UPP}(\mathsf{MOD}_m^A \circ \mathsf{XOR}) = \Omega(n)$$

if MOD_m^A does not equal a constant or Parity or the complement of Parity.

Remark 1.5.5. Two very recent results of Hatami and Qian [HQ17] and Ada, Fawzi and Kulkarni [AFK17] subsume Theorem 1.5.4. However, their results are based on a simple reduction to symmetric AND functions, whose unbounded-error complexity has been tightly characterized by Sherstov [She11b] using sophisticated tools from approximation theory. Our result, on the other hand, is based on first principles using Fourier analysis of boolean functions. Our result also shaves significant logarithmic factors off that of [HQ17, AFK17].

Interestingly, we do not invoke linear programming duality in the proof of Theorem 1.5.4, as opposed to our PP and BPP lower bounds stated in Section 1.4, or even the UPP lower bound in Theorem 1.5.1.

Proof Idea: We first recall that Forster's theorem along with a simple observation tells us that the sign rank of $f \circ \mathsf{XOR}$ is bounded below by the inverse of the maximum Fourier coefficient of f. Observe that $\mathsf{MOD}_3^{\{0\}}$ has a large principal Fourier coefficient even though the other coefficients are inverse exponentially small. We prove a generalization of Forster's theorem, allowing us to handle such cases and prove strong lower bounds for $\mathsf{MOD}_m^A \circ \mathsf{XOR}$ when m is odd. Finally, we use a shifting and XORing trick to prove hardness of $\mathsf{MOD}_m^A \circ \mathsf{XOR}$ for all $m = O(\sqrt{n})$.

1.6 Multi-Party Communication

Chandra, Furst and Lipton [CFL83] introduced the "number-on-forehead" (NOF) model of multi-party communication, over thirty years ago, to obtain lower bounds on the size of branching programs. In the NOF model, there are k players each having an input that is metaphorically held on their foreheads. Every forehead is visible to a player except her own.[5] The two features that make this model more subtle than its classical two-party counterpart, are the mutual overlap of information and the fact that as k grows, each player misses less information. Indeed, starting with the surprising work of Grolmusz [Gro94], several works (see for example [BGKL03, ACFN15, CS14]) have shown that there are very counter-intuitive protocols especially when k is larger than $\log n$. This makes proving multi-party lower bounds on the cost of protocols quite challenging. However, researchers have been well motivated to take on this challenge due to many well-known applications of such lower bounds in diverse areas like circuit complexity, proof complexity, and pseudo-random generators. More recently new applications have emerged in areas like data structures [Pat10] and distributed computing [DKO14].

The notion of communication complexity classes being analogous to Turing machine complexity classes also extends easily to the NOF model and gives rise to complexity classes $\mathsf{P}_k^{cc}, \mathsf{BPP}_k^{cc}, \mathsf{NP}_k^{cc}, \mathsf{PP}_k^{cc}$ etc. As mentioned in Section 1.1, many separations in the communication world are known when $k = 2$. However, for $k \geq 3$, things become more delicate. For instance, Beame et al. [BDPW10] separated P_k^{cc} from BPP_k^{cc} for $k \geq 3$ not too long ago, but it is still outstanding to find an explicit function witnessing this separation for even $k = 3$. A line of work [LS09a, CA08, Cha09, She16b, She14, RY15] showed that Set-Disjointness also separates BPP_k^{cc} and PP_k^{cc} for $k \leq \delta \cdot \log n$ for some constant $\delta < 1$.

Recall from Section 1.4 that the inclusion $\mathsf{PP}_2^{cc} \subsetneq \mathsf{UPP}_2^{cc}$ was shown independently (using different functions) by Buhrman, Vereshchagin and de Wolf [BVdW07] and by Sherstov [She08], and reproved by us using different techniques in this thesis. Sherstov [She08] observed that this separation was already implicit in the work of Goldmann et al. [GHR92]. However the corresponding separation question for $k \geq 3$ players remained unaddressed in the literature.

[5] The interested reader may note that this is the same model as used in the popular card game, Hanabi.

1.6.1 Our Work

We separate $\mathsf{PP}^{\mathrm{cc}}_k$ from $\mathsf{UPP}^{\mathrm{cc}}_k$ for $k \leq \delta \log N$, for any constant $\delta < 1/4$, using a simple and natural extension of the GHR function, formally defined in Definition 5.1.1. This function is an XOR function, where the outer function is a linear threshold function.

Theorem 1.6.1. The function GHR^N_k witnesses the separation

$$\mathsf{PP}^{\mathrm{cc}}_{k+1} \subsetneq \mathsf{UPP}^{\mathrm{cc}}_{k+1}$$

for $k \leq \delta \log N$ for any constant $0 < \delta < \frac{1}{4}$.

After an initial manuscript of our work was published, Sherstov [She16c] noted that a super-polynomial separation is also implicit by combining an earlier result of his [She11a] and the works of [Bei94, Tha16]. These routes use tools from approximation theory. The best of these separations yields $\mathsf{PP}^{\mathrm{cc}}_k$ lower bounds of $\Omega\left(n^{2/11}\right)$ which is quantitatively weaker than the lower bound we obtain. Sherstov [She16c] also noted that one can match our $\sqrt{n}$ bounds by combining some explicit and implicit results from previous works. Our argument, on the other hand, requires less background and proceeds via first principles, and provides one of the best known separations through an explicit function whose NOF complexity has not been analyzed before. This is also an arguably simpler function which uses composition with XOR functions.

The structure of the GHR^N_k function and Theorem 1.6.1 also yield the following boolean circuit class separations. Let ANY_k denote the class of all boolean functions that depend on only k of the input bits (this class is also popularly referred to as k-juntas).

Corollary 1.6.2. There exists a constant c and a function f computable by linear sized $\mathsf{THR} \circ \mathsf{XOR}_k$ circuits, but cannot be computed by polynomial sized $\mathsf{MAJ} \circ \mathsf{THR} \circ \mathsf{ANY}_{k-1}$ circuits, for any $k < c \log n$.

Corollary 1.6.3. There exists a constant c and a linear threshold function f such that f cannot be computed by polynomial sized $\mathsf{MAJ} \circ \mathsf{XOR} \circ \mathsf{ANY}_k$ circuits for any $k < c \log n$.

Proof idea of Theorem 1.6.1: We follow the ideas of Goldmann et al. [GHR92], and show that it is sufficient to exhibit an upper bound on the discrepancy of a function related to the GHR function under a particular product distribution. Analyzing the discrepancy of this related function on the obtained product distribution is still non-trivial, and is the main technical contribution of this work. It involves proving

independently interesting properties of integral solutions to a particular linear program when the constraint matrix is a Hadamard matrix. The upper bound is easy to prove.

1.7 Linear Decision Lists

A natural program that arises in view of Theorem 1.5.3 is to devise a strong lower bound technique against the class P^{MA}. Recall that F_n can be efficiently expressed as a decision list of equalities. Equality is a special case of an *exact threshold function*. Exact threshold functions f are those for which there exist reals $w_1, \ldots, w_n, w_0$ such that $f(x) = -1$ iff $\sum_{i=1}^{n} w_i x_i = w_0$. It is not hard to observe that any decision list of exact threshold functions can be computed efficiently using P^{MA} protocols (a simple modification of Protocol 1). Thus, a plausibly easier first step is to prove lower bounds against the circuit class defined by polynomial sized decision lists of exact thresholds.

Towards this goal, we show a simple lower bound against decision lists, where the queries are just linear threshold functions, rather than exact threshold functions. We denote decision lists with queries to linear threshold functions by *linear decision lists* (which we occasionally denote by LDL's). Lower bounds against linear decision lists and linear decision trees for IP were proved by [GT91, TV97]. Subsequently, [UT11, UT15] observed lower bounds for functions with large UPP complexity, against the classes of linear decision lists and linear decision trees when the weights of the linear threshold queries are bounded by a polynomial in the input length, by noting that functions computed in these classes can be efficiently computed by $\mathsf{THR} \circ \mathsf{MAJ}$ circuits.

LDL's also have an intricate connection with threshold circuits, as they are easily seen to be a subclass of $\mathsf{THR} \circ \mathsf{THR}$. Naturally, one might ask if they are as powerful as $\mathsf{THR} \circ \mathsf{THR}$, or if they are too weak to even compute all functions in $\mathsf{MAJ} \circ \mathsf{MAJ}$. Turán and Vatan [TV97] asked the question of how linear decision lists compare to $\mathsf{MAJ} \circ \mathsf{MAJ}$. The results of Buhrman et al. [BVdW07] and Sherstov [She11a] implied that there is an LDL which cannot be efficiently computed by $\mathsf{MAJ} \circ \mathsf{MAJ}$ circuits.

Our main theorem regarding linear decision lists is as follows, resolving the aforementioned open question posed by Turán and Vatan.

Theorem 1.7.1. There exists a function that can be computed by polynomial sized $\mathsf{MAJ} \circ \mathsf{MAJ}$ circuits, but any linear decision list computing it requires exponential size.

In order to prove this, we first observe that the lower bound argument of [TV97] shows that functions efficiently computable by linear decision lists (with no restrictions on the weights of the queried linear threshold functions) must have large monochromatic rectangles.

Next, we show, using Harper's theorem, that any monochromatic rectangle in the communication matrix of $\mathsf{MAJ} \circ \mathsf{XOR}$ must have size at most $2^{0.82n}$. Theorem 1.7.1 is now proven, since it is not hard to show that $\mathsf{MAJ} \circ \mathsf{XOR}$ can be computed by linear sized $\mathsf{MAJ} \circ \mathsf{MAJ}$ circuits.

1.8 Organization of Thesis

1. **Chapter 2**: We define various functions, communication complexity classes and approximation-theoretic notions of interest. We also provide some basic preliminaries that shall be used throughout this thesis.

2. **Chapter 3**: We describe how the *weakly unbounded-error* communication complexity of XOR functions is tightly characterized by an approximation theoretic quantity, the *polynomial margin* of the outer function. We also prove a lifting theorem, translating degree-hardness properties of f to monomial-hardness properties of a certain lifted version of f. We list several applications of these results, some of which resolve open problems in the area of analysis of boolean functions.

3. **Chapter 4**: We describe our results on the *unbounded-error* communication complexity of XOR functions. We prove a lower bound on the *sign rank* of an explicit function, proving a communication complexity class separation and resolving a longstanding open problem in boolean circuit complexity. Finally, we also show an unbounded-error lower bound for $\mathsf{MOD}_m^A \circ \mathsf{XOR}$ using a different approach from our other unbounded error lower bound.

4. **Chapter 5**: We study a communication lower bound against an XOR function in the *multi-party number-on-forehead* model of communication, yielding a multi-party communication complexity class separation and boolean circuit class separations.

5. **Chapter 6**: We take a short excursion to study *linear decision lists*.

6. **Chapter 7**: We conclude with a summary of our contributions to this thesis, a discussion on possible future directions, and list some open problems.

Chapter 2

Definitions and Preliminaries

In this chapter, we define functions, communication complexity classes and approximation-theoretic notions of interest to us in this thesis. The purpose of this chapter is purely for ease of reference, and several definitions here are restated elsewhere in this thesis. We also state some basic preliminaries of use throughout this thesis.

2.1 Functions

Recall that we interchangeably view the input variables and/or outputs as $\{0,1\}$ and $\{-1,1\}$ valued. In $\{-1,1\}$, -1 is to be interpreted as logical TRUE and 1 as logical FALSE. We consider the $\{-1,1\}$ view unless mentioned otherwise. The Hamming weight of a string $x \in \{-1,1\}$, denoted $|x|$, is defined to be the number of -1's in x. In this section, we define some functions and classes of functions of interest to us in this thesis.

Definition 2.1.1 (XOR)**.** $\mathsf{XOR} : \{-1,1\}^k \to \{-1,1\}$ is defined as

$$\mathsf{XOR}(x_1, \ldots, x_k) = x_1 \oplus \cdots \oplus x_k,$$

where

$$x \oplus y = \begin{cases} -1 & \text{if } x \neq y \\ 1 & \text{otherwise.} \end{cases}$$

Definition 2.1.2 (Sign)**.** Define the sign function, denoted $\mathsf{sgn} : \mathbb{R} \to \{-1,1\}$, by $\mathsf{sgn}(x) = -1$ iff $x < 0$.

Definition 2.1.3 (Majority)**.** Define the Majority function, denoted MAJ : $\{-1,1\}^n \to \{-1,1\}$, by

$$\mathsf{MAJ}(x_1,\ldots,x_n) = \mathsf{sgn}\left(\sum_{i=1}^{k} x_i\right).$$

Definition 2.1.4 (Linear threshold functions)**.** A function $f : \{-1,1\}^n \to \{-1,1\}$ is said to be a linear threshold function if there exist reals $w_0, w_1, \ldots, w_n$ such that $f(x) = \mathsf{sgn}(w_0 + \sum_{i=1}^{n} w_i x_i)$. Let THR denote the class of all linear threshold functions.

Definition 2.1.5 (Exact threshold functions)**.** A function $f : \{-1,1\}^n \to \{-1,1\}$ is said to be an exact threshold function if there exist reals $w_0, w_1, \ldots, w_n$ such that $f(x) = -1$ iff $\sum_{i=1}^{n} w_i x_i = w_0$. Let ETHR denote the class of all exact threshold functions.

Definition 2.1.6 (Decision lists)**.** A decision list of length k, is a sequence $D = (L_1, a_1), (L_2, a_2), \ldots, (L_k, a_k)$, where each $a_i \in \{-1,1\}$, and L_k is the constant -1 function. The decision list computes a function $f : \{-1,1\}^n \to \{-1,1\}$ as follows. If $L_1(x) = -1$, then $f(x) = a_1$; elseif $L_2(x) = -1$, then $f(x) = a_2$, elseif $\ldots$, elseif $L_k(x) = -1$, then $f(x) = a_k$. That is,

$$f(x) = \bigvee_{i=1}^{k} \left(a_i \bigwedge L_i(x) \bigwedge_{j<i} \neg L_j(x) \right).$$

Definition 2.1.7 (OMB)**.** Define the ODD-MAX-BIT function, denoted OMB : $\{-1,1\}^n \to \{-1,1\}$, by

$$\mathsf{OMB}(x) = -1 \text{ iff } \max\{i \in [n] : x_i = -1\} \text{ is odd.}$$

Definition 2.1.8 (OMB^0)**.** Define a simple variant of OMB function, which we denote by $\mathsf{OMB}^0 : \{-1,1\}^n \to \{-1,1\}$, by

$$\mathsf{OMB}^0(x) = -1 \text{ iff } \max\{i \in [n] : x_i = 1\} \text{ is odd.}$$

For classes of functions $\mathcal{C}$ and $\mathcal{D}$, denote by $\mathcal{C} \circ \mathcal{D}$ the class of all functions computable by polynomial-size (in the input length) depth-2 circuits where the top gate computes a function in $\mathcal{C}$ and the bottom gates compute functions in $\mathcal{D}$. Larger depth circuit classes are defined in a similar fashion.

Definition 2.1.9 (Symmetric functions)**.** A function $f : \{-1,1\}^n \to \{-1,1\}$ is said to be symmetric if $f(x_1,\ldots,x_n) = f(x_{\sigma(1)},\ldots,x_{\sigma(n)})$ for all $\sigma \in S_n$ where S_n denotes the set of all permutations on n elements. In other words, a function is symmetric if its value on an input only depends on the Hamming weight of the input. Let SYM denote the class of all symmetric functions.

Definition 2.1.10 (Spectrum)**.** We denote the spectrum (or predicate) of a symmetric function $f : \{-1,1\}^n \to \{-1,1\}$ by $D_f : \{0,1,\ldots n\} \to \{-1,1\}$ and define it as follows.

$$D_f(i) = -1 \text{ iff } f(x) = -1 \text{ for } |x| = i.$$

Note that the spectrum of a symmetric function is well-defined.

Definition 2.1.11 (MOD functions)**.** A function $f : \{-1,1\}^n \to \{-1,1\}$ is called a MOD function if there exists a positive integer $m < n$ and an 'accepting' set $A \subseteq [m]$ such that

$$f(x) = \begin{cases} -1 & |x| \equiv k \bmod m \text{ for some } k \in A \\ 1 & \text{otherwise.} \end{cases}$$

We write $f = \mathsf{MOD}_m^A$.

We now introduce function composition. Given functions $f_n : \{-1,1\}^n \to \{-1,1\}$ and $g_m : \{-1,1\}^m \to \{-1,1\}$, define the *composed* function $f_n \circ g_m : \{-1,1\}^{mn} \to \{-1,1\}$ as $f_n \circ g_m(x_{11},\ldots,x_{1m},x_{21},\ldots,x_{2m},\ldots,x_{n1},\ldots,x_{nm}) = f_n(g_m(x_1),g_m(x_2),\ldots,g_m(x_n))$. We often drop the subscripts when the arities of the constituent functions are clear.

Definition 2.1.12 (Equality)**.** The Equality function, denoted $\mathsf{EQ} : \{-1,1\}^n \times \{-1,1\}^n$ is defined by

$$\mathsf{EQ}(x,y) = \begin{cases} -1 & \text{if } x_i = y_i \text{ for all } i \in [n] \\ 1 & \text{otherwise.} \end{cases}$$

Note that $\mathsf{EQ} = \mathsf{NOR} \circ \mathsf{XOR}_2$.

Definition 2.1.13 (Greater Than)**.** The Greater-Than function, denoted $\mathsf{GT} : \{-1,1\}^n \times \{-1,1\}^n$ is defined by

$$\mathsf{GT}(x,y) = \begin{cases} -1 & \text{if } n(x) > n(y) \\ 1 & \text{otherwise} \end{cases}$$

where $n(x)$ and $n(y)$ are the integers corresponding to the binary representations x and y, respectively.

Definition 2.1.14 (Set Disjointness)**.** The Set-Disjointess function, denoted $\mathsf{DISJ} : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ is defined by

$$\mathsf{DISJ}(x,y) = \begin{cases} -1 & \text{if } (x_i, y_i) \neq (-1,-1) \text{ for all } i \in [n] \\ 1 & \text{otherwise.} \end{cases}$$

Note that $\mathsf{DISJ} = \mathsf{NOR} \circ \mathsf{AND}_2$.

Definition 2.1.15 (Inner Product Modulo 2)**.** The Inner Product Modulo 2 function, denoted $\mathsf{IP} : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ is defined by

$$\mathsf{IP}(x,y) = \mathsf{XOR}((x_1 \wedge y_1), \ldots, (x_n \wedge y_n)).$$

Note that $\mathsf{IP} = \mathsf{XOR} \circ \mathsf{AND}_2$.

2.2 Fourier Analysis

Consider the vector space of functions from $\{-1,1\}^n$ to $\mathbb{R}$, equipped with the following inner product.

$$\langle f, g \rangle = \mathbb{E}_{x \in \{-1,1\}^n} f(x)g(x) = \frac{1}{2^n} \sum_{x \in \{-1,1\}^n} f(x)g(x).$$

Define characters χ_S for every $S \subseteq [n]$ by $\chi_S(x) = \prod_{i \in S} x_i$. The set $\{\chi_S : S \subseteq [n]\}$ forms an orthonormal basis for this vector space. Thus, every $f : \{-1,1\}^n \to \mathbb{R}$ can be uniquely written as $f = \sum_{S \subseteq [n]} \widehat{f}(S)\chi_S$ where

$$\widehat{f}(S) = \langle f, \chi_S \rangle = \mathbb{E}_{x \in \{-1,1\}^n} f(x)\chi_S(x). \tag{2.1}$$

Lemma 2.2.1 (Folklore)**.** For any function $f : \{-1,1\}^n \to \mathbb{R}$,

$$\mathbb{E}_{x \in \{-1,1\}^n} [|f(x)|] \geq \max_{S \subseteq [n]} \left| \widehat{f}(S) \right|.$$

Fact 2.2.2 (Plancherel's identity)**.** For any functions $f, g : \{-1,1\}^n \to \mathbb{R}$,

$$\mathbb{E}_{x\in\{-1,1\}^n}[f(x)g(x)] = \sum_{S\subseteq[n]} \widehat{f}(S)\widehat{g}(S).$$

Lemma 2.2.3 (Folklore)**.** Let $f : \{-1,1\}^n \to \mathbb{R}$ be any real valued function and let M denote the communication matrix of $f \circ \mathsf{XOR}$. Then, the eigenvalues of M are $\{2^n\widehat{f}(S) : S \subseteq [n]\}$. Thus,

$$||M|| = 2^n \cdot \max_{S\subseteq[n]} \left|\widehat{f}(S)\right|.$$

Although this is fairly well known, we supply a proof below for completeness.

Proof. Let M denote the communcation matrix of $f \circ \mathsf{XOR}$. That is, $M_{x,y} = f(x \oplus y)$. Corresponding to each $T \subseteq [n]$, consider the vector $\chi_T \in \{-1,1\}^{2^n}$ (which is defined by $(\chi_T)_y = \chi_T(y)$).

Note that

$$M_{x,y} = \sum_{S\subseteq[n]} \widehat{f}(S)\chi_S(x \oplus y) = \sum_{S\subseteq[n]} \widehat{f}(S)\chi_S(x)\chi_S(y).$$

Fix any $T \subseteq [n]$. We now show χ_T is an eigenvector of M with eigenvalue $2^n\widehat{f}(T)$. Consider the xth coordinate of $M\chi_T$.

$$\begin{aligned}(M\chi_T)_x &= \sum_{y\in\{-1,1\}^n}\sum_{S\subseteq[n]} \widehat{f}(S)\chi_S(x)\chi_S(y)\chi_T(y)\\ &= \sum_S \widehat{f}(S)\chi_S(x)\sum_{y\in\{-1,1\}^n} \chi_{S\triangle T}(y)\\ &= \widehat{f}(T)\chi_T(x)2^n.\end{aligned}$$

Hence the eigenvalues of M are precisely $\{2^n\widehat{f}(S) : S \subseteq [n]\}$. Now, the singular values of M are just the square root of the eigenvalues of M^TM, which are the absolute values of the eigenvalues of M since M is symmetric. The lemma now follows.

□

2.3 Communication Complexity

In this section, we first define the general framework of communication complexity, and then define communication models of interest to us in this thesis.

In the general problem of two-party communication complexity [Yao79], two parties, say Alice and Bob, are individually given $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ for some finite input sets $\mathcal{X}, \mathcal{Y}$,[1] and wish to compute a given function $f : \mathcal{X} \times \mathcal{Y} \to \{-1, 1\}$ with as little communication between them as possible. They communicate according to a protocol which has been fixed in advance. The cost of a protocol is the maximum number of bits communicated on the worst case input. A formal introduction to Yao's model of communication complexity can be found in [KN97], for example.

Access to oracles, non-determinization and randomization are restricted by the model of communication under consideration. In the randomized models of our interest, Alice and Bob have access to public randomness (except for the UPP model where they only have access to private randomness). A probabilistic protocol Π computes f with advantage ϵ if the probability that f and Π agree is at least $1/2 + \epsilon$ for all inputs. Denote the cost of the best such protocol to be $R_\epsilon(f)$. Note that we deviate from the notation used in [KN97], for example.

For a communication model $\mathcal{C}$ and a function $f : \{-1, 1\}^n \times \{-1, 1\}^n \to \{-1, 1\}$, we denote by $\mathcal{C}(f)$ the minimum (or infimum) 'cost' of a 'correct' protocol for f.

2.3.1 Models of Communication

Below, we define some models of communication of interest to us in this thesis along with the definitions of their cost and correctness. In the following definitions, we denote inputs by (x, y), protocols by Π and functions by f.

Definition 2.3.1 (P)**.**

- Syntax: In a P protocol, a player's message can depend only upon previous messages and the player's input.
- Correctness: $\Pi(x, y) = f(x, y)$ for all (x, y).
- Cost: Maximum number of bits communicated on the worst-case input.

Definition 2.3.2 (NP)**.**

- Syntax: In an NP protocol, the players are given access to a 'certificate' c. A player's message can depend upon the certificate, previous messages, and the player's input.

[1] Unless mentioned otherwise, we use $\mathcal{X} = \mathcal{Y} = \{-1, 1\}^n$.

- Correctness: If $f(x,y) = 1$, then there exists a $c_{x,y}$ such that $\Pi(x,y,c_{x,y}) = 1$. If $f(x,y) = 0$, then for all c, $\Pi(x,y,c) = 0$.
- Cost: $\max_{x,y} \min_{c_{x,y}} (|c_{x,y}| + \text{ number of bits communicated})$.

Randomized Models

In randomized models, both parties have access to unlimited random strings, which we denote by r. Recall that in all models except UPP, the randomness is public, whereas the randomness in private to each party in the UPP model.

Definition 2.3.3 (BPP)**.**

- Syntax: Same as P, but players have additional access to unlimited public random bits, upon which the messages can depend.
- Correctness: $\Pr_r[\Pi(x,y,r) = f(x,y)] > 1/2 + 2/5$ for all (x,y).
- Cost ($R_{2/5}(f)$): Maximum number of bits communicated on the worst-case input.

Definition 2.3.4 (MA)**.**

- Syntax: Same as NP, but players have additional access to unlimited random bits, upon which the messages can depend.
- Correctness: If $f(x,y) = 1$, then there exists $c_{x,y}$ such that $\Pr_r[\Pi(x,y,c_{x,y}) = 1] > 1/2 + 2/5$. If $f(x,y) = 0$, then for all $c, \Pr_r[\Pi(x,y,c) = 0] > 1/2 + 2/5$.
- Cost: $\max_{x,y} \min_{c_{x,y}} (|c_{x,y}| + \text{ number of bits communicated})$.

Definition 2.3.5 (PP)**.**

- Syntax: Same as BPP.
- Correctness: $\Pr_r[\Pi(x,y,r) = f(x,y)] > 1/2 + \epsilon(n)$ for all (x,y), for some $\epsilon(n) > 0$.
- Cost: $\inf_{\epsilon(n)>0} \left(R_{\epsilon(n)}(f) + \log\left(\frac{1}{\epsilon(n)}\right)\right)$.

Definition 2.3.6 (UPP)**.**

- Syntax: Same as PP.[2]

[2]Recall that the randomness is *private* in the UPP model.

- Correctness: $\Pr_r[\Pi(x, y, r) = f(x, y)] > 1/2 + \epsilon(n)$ for all (x, y), for some $\epsilon(n) > 0$.
- Cost: $\inf_{\epsilon(n)>0} \left(R_{\epsilon(n)}(f)\right)$.

Models with Access to Oracles

For classes $\mathcal{C}$ and $\mathcal{D}$, define the model $\mathcal{C}^{\mathcal{D}}$ to have:

- Syntax: Same as $\mathcal{C}$, with the following additional property. At each step of the protocol, Alice and Bob either send a message, or they 'invoke an oracle' to compute a function $g(x', y')$ where x' and y' are strings computed by Alice and Bob from the history of the protocol until that point.
- Correctness: Same as in $\mathcal{C}$.
- Cost: Maximum number of bits communicated + the $\mathcal{D}$-cost of all the functions queried during the protocol.

We also denote by $\mathcal{C}$, or occassionally $\mathcal{C}^{cc}$, the class of all functions which admit protocols of cost (in the model $\mathcal{C}$) polylogarithmic in n.

2.3.2 Preliminaries

Definition 2.3.7 (Discrepancy)**.** For a function $F : \{-1, 1\}^n \times \{-1, 1\}^n \to \{-1, 1\}$, define the discrepancy of a rectangle $S \times T$ under a distribution λ on $\{-1, 1\}^n \times \{-1, 1\}^n$ as follows.

$$\mathsf{disc}_\lambda(S \times T, F) = \left| \sum_{(x,y)\in S\times T} F(x, y)\lambda(x, y) \right|.$$

The discrepancy of F under a distribution λ is defined as

$$\mathsf{disc}_\lambda(F) = \max_{S\subseteq\{-1,1\}^n, T\subseteq\{-1,1\}^n} \mathsf{disc}_\lambda(S \times T, F)$$

and the discrepancy of F is defined to be

$$\mathsf{disc}(F) = \min_\lambda \mathsf{disc}_\lambda(F).$$

Klauck [Kla07] proved that discrepancy and PP complexity are equivalent notions.

Theorem 2.3.8 (Klauck [Kla07])**.** For any function $f : \{-1,1\}^n \to \{-1,1\}$,

$$\mathsf{PP}(f) = \Theta\left(\log\left(\frac{1}{\mathsf{disc}(f)}\right)\right).$$

Define the *sign rank* of a real valued matrix $A = [A_{ij}]$, denoted by $\mathsf{sr}(A)$ to be the least rank of a real matrix $B = [B_{ij}]$ such that $A_{ij}B_{ij} > 0$ for all (i,j) such that $A_{ij} \neq 0$. For the purpose of this thesis, we abuse notation, and use $\mathsf{sr}(F)$ and $\mathsf{sr}(M_F)$ interchangeably, to denote the sign rank of M_F where M_F denotes the communication matrix of the function F.

Paturi and Simon [PS86] showed an equivalence between $\mathsf{UPP}(F)$ and the sign rank of M_F.

Theorem 2.3.9 (Paturi and Simon [PS86])**.** For any function $F : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$,

$$\mathsf{UPP}(F) = \log \mathsf{sr}(M_F) \pm O(1).$$

We now show a basic upper bound that is of use to us throughout this thesis.

Claim 2.3.10. For any linear threshold function $f : \{-1,1\}^n \to \{-1,1\}$, $\mathsf{UPP}(f \circ \mathsf{XOR}) = O(\log n)$.

Proof. Since f is a linear threshold function, it can be expressed as $f(z) = \mathsf{sgn}(w_0 + \sum_{i=1}^n w_i z_i)$. Without loss of generality, assume that $w_0 + \sum_{i=1}^n w_i z_i$ never takes the value 0 (else one can tweak the weights suitably). Denote $w = \sum_{i=0}^n |w_i|$.

We demonstrate a communication protocol for $f \circ \mathsf{XOR}$, and then analyze its correctness in the UPP model. Alice samples an input gate to the top linear threshold function with probability proportional to its weight. That is, Alice samples the ith gate of the linear threshold with probability $|w_i|/w$. She then sends Bob the value $\mathsf{sgn}(w_i) \times x_i$ along with the index i (the index needs to be sent since we are assuming that the randomness is private). This takes $\log n + 1$ bits of communication. Bob outputs $\mathsf{sgn}(w_i) \times x_i \times y_i$.

In order to analyze the success probability of this protocol, let us assume that $f(x \oplus y) = 1$ (the analysis of the case when $f(x \oplus y) = -1$ is analogous). Observe that the probability of success equals the probability of sampling an input to the threshold that satisfies $w_i x_i y_i > 0$. This probability equals $\sum w_i^+/w$, where the w_i^+'s are the weights of the inputs to the threshold for which $w_i x_i y_i > 0$. Since $\sum w_i^+ > \sum w_i^-$ (the weights of the inputs where $w_i x_i y_i < 0$), the probability of success is strictly greater than $1/2$. □

Remark 2.3.11. A more general form of the claim above can be stated as follows. If a function g has deterministic communication complexity c, then for any linear threshold function f_n, one has

$$\mathsf{UPP}(f_n \circ g) = c + O(\log n).$$

Another simple, yet powerful lemma regarding unbounded-error communication is as follows.

Lemma 2.3.12 (Folklore)**.** For any functions $F, G : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$,

$$\mathsf{UPP}(F \oplus G) \leq \mathsf{UPP}(F) + \mathsf{UPP}(G).$$

Proof. Let Π_1, Π_2 be unbounded-error protocols for F and G, respectively. Consider the protocol $\Pi = \Pi_1 \oplus \Pi_2$, that is Alice and Bob run both Π_1 and Π_2 and output the XOR of the two outputs. It remains to verify the correctness of this protocol.

Say Π_1 computed F with success probability $1/2 + \epsilon$ and Π_2 computed G with success probability $1/2 + \delta$. Since Π agrees with $F \oplus G$ precisely when either Π_1, Π_2 both succeed or both fail, the probability of Π agreeing with $F \oplus G$ is at least $(1/2 + \epsilon)(1/2 + \delta) + (1/2 - \epsilon)(1/2 - \delta)$. Thus, the success probability of Π is at least $1/2 + 2\epsilon\delta > 1/2$. □

2.4 Approximation Theory

In this section, we list some approximation-theoretic notions and some measures of functions of interest to us in this thesis.

Definition 2.4.1 (Degree)**.** The degree of a function $f : \{-1,1\}^n \to \{-1,1\}$, denoted $\mathsf{deg}(f)$, is the maximum degree of a monomial in the unique multilinear polynomial representing f.

Definition 2.4.2 (Sign degree)**.** The sign degree of a function $f : \{-1,1\}^n \to \{-1,1\}$, denoted $\mathsf{deg}_{\pm}(f)$, is the minimum degree of a polynomial p satisfying $p(x)f(x) > 0$ for all $x \in \{-1,1\}^n$.

Definition 2.4.3 (Approximate degree)**.** The approximate degree of a function $f : \{-1,1\}^n \to \{-1,1\}$, denoted $\widetilde{\mathsf{deg}}(f)$ or $\mathsf{deg}_{1/3}(f)$, is the minimum degree of a polynomial p satisfying $|p(x) - f(x)| < 1/3$ for all $x \in \{-1,1\}^n$.

Definition 2.4.4 (Sparsity)**.** The sparsity of a function $f : \{-1,1\}^n \to \{-1,1\}$, denoted $\mathsf{mon}(f)$, is the number of monomials with non-zero coefficients in the unique multilinear polynomial representing f.

Definition 2.4.5 (Signed monomial complexity)**.** The signed monomial complexity of a function $f : \{-1,1\}^n \to \{-1,1\}$, denoted $\mathsf{mon}_{\pm}(f)$, is the minimum sparsity of a polynomial p satisfying $p(x)f(x) > 0$ for all $x \in \{-1,1\}^n$.

Definition 2.4.6 (Approximate monomial complexity)**.** The approximate monomial complexity of a function $f : \{-1,1\}^n \to \{-1,1\}$, denoted $\mathsf{mon}_{1/3}(f)$, is the minimum sparsity of a polynomial p satisfying $|p(x) - f(x)| < 1/3$ for all $x \in \{-1,1\}^n$.

Definition 2.4.7 (Weight)**.** The weight of a function $f : \{-1,1\}^n \to \{-1,1\}$, denoted $\mathsf{wt}(f)$, is the sum of absolute values of the coefficients of the unique multilinear polynomial representing f.[3] The weight of a real polynomial is defined analogously.

Definition 2.4.8 (Polynomial margin)**.** The polynomial margin of a function $f : \{-1,1\}^n \to \{-1,1\}$, denoted $m(f)$, is the error in the best uniform approximation of f by a weight 1 polynomial. More precisely,

$$m(f) = \max_{p:\mathsf{wt}(p)=1} \min_{x \in \{-1,1\}^n} p(x)f(x).$$

Definition 2.4.9 (Approximate weight)**.** The approximate weight of a function $f : \{-1,1\}^n \to \{-1,1\}$, denoted $\mathsf{wt}_{1/3}(f)$, is the minimum weight of a polynomial p satisfying $|p(x) - f(x)| < 1/3$ for all $x \in \{-1,1\}^n$.[4]

2.4.1 Measures of Symmetric Functions

In this section, we define some measures of symmetric functions of interest.

It is well-known that the sign degree of symmetric functions equals the number of sign changes in the underlying predicate. Formally,

Fact 2.4.10 (Folklore)**.** For any symmetric function $f : \{-1,1\}^n \to \{-1,1\}$,

$$\mathsf{deg}_{\pm}(f) = |\{i \in \{0,1,\ldots,n-1\} : D_f(i) \neq D_f(i+1)\}|.$$

[3] Note that this notion coincides with $||\hat{f}||_1$, the spectral norm of f. However, for convenience, we shall use the former notation.

[4] This notion coincides with the notion of the ϵ-*approximate spectral norm* of f, denoted by $||\hat{f}||_{1,\epsilon}$, as defined in [AFH12].

Definition 2.4.11. Let $F : \{-1,1\}^n \to \{-1,1\}$ be a symmetric function. Define $r_0 = r_0(F), r_1 = r_1(F)$ to be the minimum integers r'_0 and r'_1 respectively, such that $r'_0, r'_1 \leq n/2$ and $D_F(i) = D_F(i+2)$ for all $i \in [r'_0, n - r'_1)$. Define $r = r(F) = \max\{r_0, r_1\}$.

Definition 2.4.12 (Odd-even degree)**.** The odd-even degree of a symmetric function $f : \{-1,1\}^n \to \{-1,1\}$ is defined as follows.

$$\mathsf{deg}_{oe}(f) = |i \in \{0,1,\ldots,n-2\} : D_f(i) \neq D_f(i+2)|.$$

Chapter 3

Weakly Unbounded-Error Communication

3.1 Introduction

In this chapter, we describe our results regarding the weakly unbounded-error (PP) and bounded-error (BPP) models of communication.

As mentioned in Section 1.4, $\mathsf{BPP} \subsetneq \mathsf{PP} \subsetneq \mathsf{UPP}$. There are fewer known strong lower bounds for the PP model than the BPP model. This is partly explained by the fact that while techniques based on corruption and information theory yield lower bounds for the bounded-error model, the PP model is exactly characterized by the stronger measure of discrepancy [Kla07]. Luckily, the situation is not entirely bleak with respect to the PP model; several lower bounds are known in this setting. For instance, it is not hard to show that IP has $\Omega(n)$ PP cost. Proving lower bounds for the unbounded-error model, on the other hand, is even more difficult. We elaborate on this in Chapter 4.

An important step towards understanding the randomized communication complexity of block-composed functions was taken in the works of Sherstov [She09b, She11a] and Shi and Zhu [SZ09]. These papers considered the communication complexity of $f \circ g$, where g has suitably nice properties, which XOR does not possess. They reduced the task of proving lower bounds on the cost of both (quantum) bounded-error and weakly unbounded-error protocols for such functions to that of analyzing the approximability of f by low degree real polynomials. This passage was achieved by making very elegant use of linear programming duality. This method spawned further progress in proving lower bounds against AND functions. One area

of progress was the adaptation of the technique to multi-party communication complexity in [Cha07, CA08, LS09a, Cha09], resulting in the first polynomial lower bounds for DISJ in the hard NOF model.[1] Using even more powerful approximation theoretic tools for polynomials, Sherstov [She14] significantly improved these bounds. In another direction, the power of these approximation-theoretic techniques were further demonstrated by [RS10, She11b, BT16, BCH+16], where unbounded-error lower bounds were shown against functions of the form $f \circ \mathsf{AND}$ for specific f. In short, approximation-theoretic techniques provide a systematic way of analyzing the communication complexity of AND functions. Besides these impressive developments, this approach relates to research on approximation theory, that are of independent interest (see for example [BT17, Tha16]).

There are essentially two inner functions of block length 1, AND and XOR. A natural example of an XOR function is $\mathsf{AND} \circ \mathsf{XOR}$, better known as Equality. However, even its bounded-error complexity is just $O(1)$. In fact, in some contexts as discussed later in this chapter, proving even PP lower bounds for XOR functions seems more challenging than proving lower bounds for AND functions. Interestingly, Sherstov [She08] used an XOR function introduced by Goldmann, Håstad and Razborov [GHR92], to separate PP from UPP. Zhang and Shi [ZS09] characterized the bounded-error and quantum complexity of all symmetric XOR functions. Recently, Hatami, Hosseini and Lovett [HHL18] nearly characterized the deterministic complexity of all XOR functions. Even more recently, after an initial version of a manuscript of ours containing weaker results was submitted, Hatami and Qian [HQ17] and Ada, Fawzi and Kulkarni [AFK17] independently reported settling a conjecture of Zhang and Shi [ZS09] on the unbounded-error complexity of symmetric XOR functions. Both papers analyze XOR functions by finding simple reductions to appropriate AND functions. While such arguments are short, as commented by Ada et al. [AFH12], it seems they do not provide new insights and techniques that can be applied more broadly to XOR functions.

In this chapter, we develop an approximation-theoretic technique for analyzing XOR functions with several applications. Along the way, we discover an independently interesting general connection between the discrepancy of functions of the form $f \circ \mathsf{XOR}$ and the polynomial margin complexity of f. Using this and other tools, we characterize the PP complexity of symmetric XOR functions and provide a new proof of the exponential separation between PP and UPP via an XOR function. We further

[1] We also prove some lower bounds in multi-party communication complexity in Chapter 5 in this thesis. However, we do not make use of linear programming duality.

provide a new proof of the characterization of Zhang and Shi [ZS09] of the bounded-error complexity of symmetric XOR functions. Our argument, unlike theirs, is based on a connection between the approximate spectral norm of f and the bounded-error communication complexity of $f \circ \mathsf{XOR}$. While this connection seems to have been first reported in the survey by Lee and Shraibman [LS09b], as far we know, and as expressed in Ada et al. [AFH12], ours is the first work to use it to derive explicit BPP lower bounds.

In the course of proving lower bounds on communication complexity, we obtain new results on two complexity measures of symmetric functions that are of independent interest. First, we characterize symmetric functions computable by quasi-polynomial size depth-2 boolean circuits of the form Threshold of Parity, resolving an old conjecture of Zhang [Zha92]. Further, we characterize the approximate spectral norm of symmetric functions, confirming the main conjecture of Ada et al. [AFH12], which has several consequences (cf. [AFH12]). We feel that these developments exhibit the potential of our approximation-theoretic technique for proving lower bounds against general XOR functions.

3.1.1 Our Results

In this section, we first outline our (non-communication complexity) results regarding analysis of symmetric boolean functions. These resolve open questions posed by Ada et al. [AFH12] and Zhang [Zha92], and are an independently interesting aspect of our work. Later, we list our results regarding the PP and BPP complexity of XOR functions.

Polynomial Complexity Measures of Symmetric Functions

We first outline results we obtain by amplifying hardness of functions using the method of lifting functions as defined in Krause and Pudlák. Next, we list applications of this 'hardness amplification' to symmetric functions.

In a seminal work, Paturi [Pat92] characterized the approximate degree of symmetric functions, building upon a celebrated result of Nisan and Szegedy [NS94]. Paturi's theorem has found various applications and spawned several lines of research [Raz03, She09a, She11a, dW10, Spa08, BT15a]. The sign degree of symmetric functions is well-known to be characterized by the number of sign changes in the underlying predicate. A natural analogous question that arises is whether or not there are similar results on the hardness of approximation of symmetric functions by low

monomial/weight polynomials, which might have unrestricted degree. To the best of our knowledge, such characterizations were unkown until our work.

We prove the following theorem which gives us lower bound tools against approximate weight, signed monomial complexity, and polynomial margin of symmetric functions.

Theorem 3.1.1. There exists a universal constant $c > 0$ such that for any symmetric $F : \{-1,1\}^n \to \{-1,1\}^n$, the following hold.

1. $r(F) \geq 5 \implies \log(\mathsf{wt}_{1/3}(F)) \geq c \cdot r(F)$.
2. $k = \mathsf{deg}_{oe}(F) \geq 16 \implies \mathrm{mon}_{\pm}(F) \geq 2^{c \cdot k/\log(n/k)}$.
3. $k = \mathsf{deg}_{oe}(F) \geq 16 \implies m(F) \leq \frac{1}{2^{c \cdot k/\log(n/k)}}$.

We also use Part 1 of Theorem 3.1.1, to prove the following theorem, posed as a conjecture by Ada et al. [AFH12].

Theorem 3.1.2 (Conjecture 1 in [AFH12])**.** There exist universal constants $c_0, c_1 > 0$ such that for any symmetric function $F : \{-1,1\}^n \to \{-1,1\}$,

$$c_0 \cdot r(F) \log\left(\frac{n}{r(F)}\right) \geq \log \mathsf{wt}(F) \geq \log \mathsf{wt}_{1/3}(F) \geq c_1 \cdot r(F).$$

One can view Theorem 3.1.2 as a weight-hardness analogue of Paturi's theorem. It has several other consequences, which we do not elaborate on. The interested reader may refer to Section 4 in [AFH12] for details.

We also resolve the following conjecture by Zhang [Zha92].

Theorem 3.1.3 (Conjecture 1 in [Zha92])**.** A symmetric function $f : \{-1,1\}^n \to \{-1,1\}$ is computable by a quasi-polynomial size Threshold of Parity circuit if and only if its odd-even degree is $\log^{O(1)} n$.

PP Complexity

In this section, we list our results regarding the PP complexity of XOR functions.

Recall from Theorem 2.3.8 that $\mathsf{PP}(F)$ is equivalent (up to constants) to $\log(1/\mathsf{disc}(F))$ for any boolean function F. Thus, it suffices to prove strong upper bounds on discrepancy in order to prove strong PP lower bounds. Our main tool for analyzing the discrepancy of XOR functions is a tight relationship (upto constant factors) between $\mathsf{disc}(f \circ \mathsf{XOR})$ and $m(f)$. We derive this using linear programming duality.

Theorem 3.1.4 (Polynomial Margin-Discrepancy theorem)**.** Let $f \to \{-1,1\}^n \to \{-1,1\}$.

$$m(f) \leq m(f \circ \mathsf{XOR}) \leq 4\mathsf{disc}(f \circ \mathsf{XOR}) \leq 4m(f).$$

The proof of Theorem 3.1.4 shows that the discrepancy of every XOR function is attained on a lifted distribution. Indeed, our Margin-Discrepancy Theorem is a lifting theorem for XOR functions that primarily reduces the task of proving a lower bound on the discrepancy of $f \circ \mathsf{XOR}$ to that of establishing bounds on the polynomial margin of f. The second task is likely easier using tools from approximation theory. There is a compelling parallel here with the Degree-Discrepancy Theorem of Sherstov [She09b]. That theorem has yielded a methodical way of proving discrepancy bounds for $f \circ \mathsf{PM}$ by showing a lower bound on the *sign degree* of f, where PM denotes the pattern matrix gadget, and is defined formally in Section 3.2. This has led to much progress in understanding the communication complexity of AND functions (for example, [Cha07, CA08, She11a, She11b]). We believe our polynomial Margin-Discrepancy Theorem will yield a unified approach in making similar progress for XOR functions. As evidence of this, we provide two applications of this theorem.

Zhang and Shi [ZS09] conjectured that for any symmetric f, $\mathsf{UPP}(f \circ \mathsf{XOR})$ is essentially the odd-even degree of f. Our first application of the Margin-Discrepancy Theorem shows that the PP complexity of functions of the form $f \circ \mathsf{XOR}$ for symmetric f is essentially the odd-even degree of f (upto polylogarithmic factors) as predicted by the conjecture of Zhang and Shi [ZS09].

Theorem 3.1.5. There exists universal constants $c_1, c_2 > 0$ such that for any symmetric $f : \{-1,1\}^{4n} \to \{-1,1\}$ with $r \geq 4$ denoting its odd-even degree,

$$c_1 r \log n \leq \mathsf{PP}(f \circ \mathsf{XOR}) \geq c_2 r / \log(n/r).$$

In Section 4.8, we describe our further progress towards the full Zhang-Shi conjecture, proving it for the case when f is symmetric and its spectrum is periodic. The full conjecture was subsequently resolved independently by Ada et al. [AFK17] and Hatami and Qian [HQ17], who used completely different techniques from ours. Their approaches, among other things, make use of the characterization of the UPP complexity of symmetric AND functions by Sherstov [She11b]. Our approaches, on the other hand develop tools of independent interest that contribute to the theory of XOR functions.

To prove the Theorem 3.1.5, Theorem 3.1.4 sets the goal of establishing a bound on the margin complexity of symmetric functions with large odd-even degree. We do

this by showing that symmetric functions with large odd-even degree can be projected onto a certain lift of symmetric functions with high sign degree. This enables us to work with the more convenient notion of sign degree rather than odd-even degree of symmetric functions.

As another application of our Margin-Discrepancy connection, we provide a new proof of the separation of PP from UPP. We do this by proving that an XOR function, almost identical to the GHR function (cf. [GHR92]) has exponentially small discrepancy. It is well-known that this function has very efficient UPP protocols (see Claim 2.3.10). We define the GHR function formally in Section 3.2.

Theorem 3.1.6.

1. There exists an absolute constant $c > 0$ and a linear threshold function $f : \{-1,1\}^n \to \{-1,1\}$ such that $\mathsf{PP}(f \circ \mathsf{XOR}) \geq cn$.
2. $\mathsf{PP}(\mathsf{GHR}) \geq \Omega(\sqrt{n})$.

BPP Complexity

Using linear programming duality and the generalized discrepancy method (Theorem 3.2.13), we give a simple alternate proof of the following result due to Lee and Shraibman [LS09b].

Theorem 3.1.7. For any function $f : \{-1,1\}^n \to \{-1,1\}$,

$$\mathsf{BPP}(f \circ \mathsf{XOR}) \geq \log \mathsf{wt}_{1/3}(f) - 4.$$

Remark 3.1.8. In fact, Lee and Shraibman proved that lower bounds on $\mathsf{wt}_{1/3}(f)$ also yield lower bounds on the bounded-error quantum communication complexity of $f \circ \mathsf{XOR}$. Our proof also implies this.

Using Part 1 of Theorem 3.1.1 in conjunction with Theorem 3.1.7 provides an alternate proof of the following result of Zhang and Shi [ZS09].

Theorem 3.1.9 ([ZS09])**.** Let $F : \{-1,1\}^n \to \{-1,1\}$ be any symmetric function. Then, $\mathsf{BPP}(F \circ \mathsf{XOR}) = \Omega(r(F))$.

Remark 3.1.10. Blais et al. [BBG14] also provided an alternate proof to Theorem 3.1.9 by showing a lower bound on the information complexity of symmetric XOR functions (it is known however that information complexity lower bounds need not imply quantum lower bounds [KLL+15]).

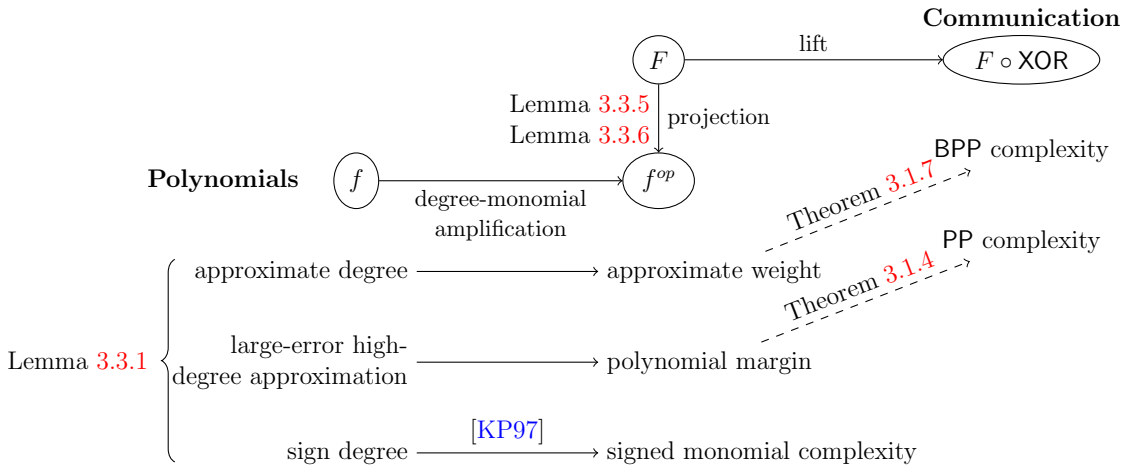

Figure 3.1: General framework

3.1.2 Proof Outline

Our proof strategy is depicted in Figure 3.1. First, we use an idea due to Krause and Pudlák [KP97], who showed that if a function f has high sign degree, then a certain *lift* of that function, denoted by f^{op} has high sign monomial complexity. We observe that their argument can be adapted to show a more general result. In particular, our Lemma 3.3.1 shows that the hardness of f for *low degree* polynomials, with respect to natural notions like uniform approximation and sign representation, gets amplified to corresponding hardness of f^{op} for *sparse (low weight)* polynomials. Next, we observe that LP duality implies, via Theorems 3.1.4 and 3.1.7, that such hardness of a function F against sparse polynomials translates to the hardness of $F \circ \mathsf{XOR}$ for the appropriate randomized $(\mathsf{BPP}, \mathsf{PP})$ communication model. The main problem at this point is to understand how F relates to an appropriately hard f^{op}. In particular, our interest is when F is a symmetric function or a linear halfspace. These functions do not seem to have the structure of a lifted function f^{op}.

At this point, inspired by the work of Krause [Kra06], we make a simple but somewhat counter-intuitve observation that turns out to be crucial. A function g is called a monomial projection of h, if g can be obtained by substituting each input variable of h with a monomial in variables of g. What is nice about such projections is that for the polynomial sparsity measures (Lemma 3.3.4) that are relevant for us, the complexity of g is bounded above by that of h. We observe (Lemma 3.3.6 and Lemma 3.3.5) that if f is a symmetric (linear threshold) function, then there exists a symmetric (linear threshold) function F such that f^{op} is a monomial projection

of F. Moreover, the combinatorial parameters of f that caused its hardness against low-degree polynomials, nicely translate to combinatorial parameters of F that have been conjectured to cause hardness of F against sparse (low weight) polynomials. By our LP duality theorems, these result in the hardness of $F \circ \mathsf{XOR}$ against randomized communication protocols as well.

The above describes the general framework of our passage from polynomials to communication protocols. We describe below the particular instantiations of this framework for each of the lower bounds that we prove.

PP Complexity

We prove two main results regarding PP complexity by showing explicit upper bounds on polynomial margin of certain functions. The first is to reprove an exponential separation between PP and UPP, making use of the above framework. For this, it is natural to prove a strong PP lower bound against a function of the type $F \circ \mathsf{XOR}$ where F is a linear threshold function. Proving a polylogarithmic UPP upper bound for such a function is straightforward (see Claim 2.3.10). However, precisely this feature of F makes it difficult to prove a strong PP lower bound. Goldmann et al. [GHR92] used an ingenious specialized argument directly establishing that the discrepancy is small.[2] We, on the other hand, use Theorem 3.1.4 which directs us in proving that F must have small polynomial margin. The challenge here is to prove a strong unrestricted degree polynomial margin lower bound against a function with sign degree just 1. We use a variety of techniques to prove this. First, we use a result of Sherstov, Theorem 3.2.3, which states that there exists a linear threshold function f which requires linear degree to approximate uniformly, even with error inverse exponentially close to 1. Second, we use lifting as depicted in Figure 3.1 to show that the polynomial margin (unrestricted degree) of f^{op} is exponentially small. We then use our monomial projection lemma for threshold functions, Lemma 3.3.5, to embed such a lifted function in a linear threshold function F without blowing up the weights too much. Finally, we exploit the fact that the Universal Threshold function (UTHR) embeds any other threshold function with at most a quadratic loss in number of variables. The last step of considering UTHR is needed only to match the currently best known *explicit* exponential separation of PP and UPP.

As a second application of our framework to PP complexity, we prove Theorem 3.1.5, which states that for symmetric F, the PP complexity of $F \circ \mathsf{XOR}$ is essentially the odd-even degree of F. The main challenge here is to work with the notion of

[2] In Chapter 5, we extend the argument of Goldmann et al. to the multi-party NOF model.

odd-even degree, which has no immediate algebraic interpretation as opposed to sign degree. Lemma 3.3.6 solves this by essentially showing that there exists a symmetric f whose sign degree corresponds to the odd-even degree of F, such that f^{op} is a monomial projection of F. Finally, our lifting lemma, Lemma 3.3.1, shows that the margin of f^{op} must be small if the base function f has large sign degree.

BPP Complexity and Approximate Weight

We first make a simple observation that the polynomial margin of a function F equals its *threshold weight*, as defined in Definition 3.2.7. Just as the notion of threshold degree inspires the natural notion of approximate degree, threshold weight inspires the definition of approximate weight as in Definition 2.4.9. In Section 3.5, we consider a linear program capturing the $(1/3)$-approximate weight of a symmetric function $F : \{-1,1\}^n \to \{-1,1\}$. Using linear programming duality and the generalized discrepancy method, we show in Theorem 3.1.7 that $\log \mathsf{wt}_{1/3}(F)$ is a lower bound on $\mathsf{BPP}(F \circ \mathsf{XOR})$.

The general framework of Figure 3.1 then prescribes us to find a suitable symmetric f such that f^{op} has large approximate weight and is a monomial projection of F. Lemma 3.3.6 provides such a monomial projection in which the combinatorial quantity $r(F)$ corresponds to another combinatorial quantity $\Gamma(f)$, which is defined in Section 3.2. Paturi [Pat92] showed that $\Gamma(f)$ characterizes the approximate degree of f. The polynomial hardness amplification of Figure 3.1, via Lemma 3.3.1, implies that f^{op}, and therefore F, has large approximate weight. This already proves Theorem 3.1.2 which was conjectured by Ada et al. [AFH12]. Moreover, Theorem 3.1.7 implies the hardness of $F \circ \mathsf{XOR}$ against bounded-error protocols.

3.2 Preliminaries

We provide the necessary preliminaries in this section.

Definition 3.2.1 (Universal threshold)**.** Define a class of threshold functions, $U_{l,k} : \{\{-1,1\}^k\}^l \to \{-1,1\}$ defined by

$$U_{l,k}(x_{1,1},\ldots,x_{1,k},\ldots,x_{l,1},\ldots,x_{l,k}) = sgn\left(\sum_{i=1}^{k}\sum_{j=1}^{l} 2^i x_{i,j} + \frac{1}{2}\right).$$

The constant term $\frac{1}{2}$ is added to ensure that the sum inside the brackets is never 0.

Fact 3.2.2 (Minsky and Papert [MP69])**.** $U_{l,k}$ is universal in the sense that any linear threshold function on n variables occurs as a subfunction of $U_{l,k}$ for some $l, k \in O(n \log n)$.

We use the notation UTHR to denote such a function.

Recall that the weight of a polynomial p, denoted $\mathsf{wt}(p)$, is defined to be the sum of absolute values of its coefficients in the unique multilinear expression for p. For a polynomial of weight 1, say p, which sign represents a function f, we say that p represents f with a margin of value $\min_{x \in \{-1,1\}^n} f(x)p(x)$. Let us also define a notion of the error in a pointwise approximation of a function by low degree polynomials. This notion is studied widely in classical approximation theory, see [SZ09, She11a, Tha16] for example. Note that we do not restrict the weight of the approximating polynomial in this case.

$$\varepsilon_d(f) \triangleq \min_{p:deg(p) \leq d} \left(\max_{x \in \{-1,1\}^n} |p(x) - f(x)| \right). \tag{3.1}$$

Sherstov [She16c] proved that there exists a linear threshold function which cannot be approximated well, even by large degree polynomials.

Theorem 3.2.3 ([She16c], Cor 3.3)**.** There exists a linear threshold function $f : \{-1,1\}^n \to \{-1,1\}$ and an absolute constant $c > 0$ such that

$$\varepsilon_{cn}(f) > 1 - 2^{-cn}.$$

Moreover, the weights of the coefficients in the function have magnitude at most 2^n.

Note that the signed monomial complexity (see Definition 2.4.5) of a function f exactly corresponds to the minimum size Threshold of Parity circuit computing it, since monomials are just parities.

Theorem 3.2.4 ([Zha92])**.** Let $f : \{-1,1\}^n \to \{-1,1\}$ be a symmetric boolean function such that $\mathsf{deg}_{oe}(f) = \log^{O(1)} n$. Then, f can be computed by a quasi-polynomial size Threshold of Parity circuit.

The following is a result by Paturi [Pat92] which gives us tight bounds on the approximate degree of symmetric functions.

Theorem 3.2.5 ([Pat92])**.** For any symmetric function $f : \{0,1\}^n \to \{-1,1\}$, define the quantity $\Gamma(f) = \min\{|2k - n + 1| : D_f(k) \neq D_f(k+1) \text{ and } 0 \leq k \leq n-1\}$. Then,

$$\widetilde{\mathsf{deg}}_{2/3}(f) = \Theta(\sqrt{n(n - \Gamma(f))}).$$

Definition 3.2.6. For functions $f, g : \{-1,1\}^n \to \{-1,1\}$ and a distribution ν on $\{-1,1\}^n$, define the correlation between f and g under the distribution ν to be

$$\mathrm{corr}_\nu(f,g) = \mathbb{E}_\nu[f(x)g(x)].$$

Definition 3.2.7 (Threshold weight)**.** Define the *threshold weight* of a function $f : \{-1,1\}^n \to \{-1,1\}$, denoted by $\mathsf{wt}_\pm(f)$ to be the weight of a minimum weight real polynomial p such that $p(x)f(x) \geq 1$ for all $x \in \{-1,1\}^n$.

Note that this definition differs from the notion of more widely studied notion of threshold weight (see for example [Kra06], [She11a], [BT15b]), where the coefficients of p are restricted to be integer valued. It is convenient for us to work with the notion as defined in Definition 3.2.7 because of its following relationship with the polynomial margin, which can be easily verified.

Lemma 3.2.8. For any function $f : \{-1,1\}^n \to \{-1,1\}$,

$$m(f) = \frac{1}{\mathsf{wt}_\pm(f)}.$$

The following theorem is due to Ada et al. [AFH12], which characterizes the weight of a symmetric function.

Theorem 3.2.9 ([AFH12])**.** For any symmetric function $f : \{-1,1\}^n \to \{-1,1\}$,

$$\log(\mathsf{wt}(f)) = \Theta\left(r(f)\log\left(\frac{n}{r(f)}\right)\right).$$

Goldmann et al. [GHR92] exhibited a distribution under which the one way communication complexity of $U_{4n,n} \circ \mathsf{XOR}$ is large. Sherstov [She08] noted that the same proof can be used to show that $\mathsf{PP}(U_{4n,n} \circ \mathsf{XOR})$ is large as well.

Remark 3.2.10. We remark here that the function considered by Goldmann et al. was not exactly $U_{4n,n} \circ \mathsf{XOR}$, because the variables feeding to the XOR gates had a mild dependence on each other. Thus the discrepancy bound they obtained was slightly stronger than as stated above. However, we will refer to $\mathsf{UTHR} \circ \mathsf{XOR}$ as the GHR function.

Sherstov defined the notion of a pattern matrix communication game in [She11a]. Let n be a positive integer and $f : \{0,1\}^n \to \{-1,1\}$. Alice is given $2n$ bits $x_{1,1}, x_{1,2}, x_{2,1}, x_{2,2}, \ldots, x_{n,1}, x_{n,2}$. Bob is given $2n$ bits $z_1, z_2, \ldots, z_n, w_1, w_2, \ldots w_n$. Define PM to be the function on 4 bits defined as $\mathsf{PM}(x_0, x_1, z, w) = x_z \oplus w$. In the

pattern matrix game corresponding to f, the PM gadget is applied on each tuple $\{x_{i,1}, x_{i_2}, z_i, w_i\}$, and the resultant n bit string is fed as input to f. This is the composed function, $f \circ \mathsf{PM}$. Notice that this is similar to the lifting as defined in Equation 3.2.

Theorem 3.2.11 ([She11a] Thm 1.5)**.** Let $F = f \circ \mathsf{PM}$ for a given function $f : \{0,1\}^n \to \{-1,1\}$. Then

$$\mathsf{disc}(F) \leq \min_{d=1,\ldots,n} \max \left\{ \left(\frac{n}{W(f, d-1)} \right)^{1/2}, \left(\frac{1}{2} \right)^{d/2} \right\}.$$

In the above theorem, $W(f, d-1)$ corresponds to the minimum weight of a polynomial of degree $d-1$ with integer weights which sign represents f.

Remark 3.2.12. Sherstov defined pattern matrices in a more general fashion, where n bits could be split into t blocks containing n/t elements each. However, for the purposes of this thesis, we only consider the case when each block is of size 2.

The following theorem, first proposed by Klauck [Kla07], provides a tool for proving bounded-error communication lower bounds for functions. Its proof may be found in [Cha09, CA08], for example.

Theorem 3.2.13 (Generalized discrepancy)**.** Let $F, G : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ and ν be a distribution over $\{-1,1\}^n \times \{-1,1\}^n$ such that $\mathrm{corr}_\nu(F, G) \geq \delta$. Then.

$$R_\epsilon(F) \geq \log \left(\frac{\delta - 1 + 2\epsilon}{\mathsf{disc}_\nu(G)} \right).$$

3.3 Lifting Functions

In this section we first show how we 'lift' functions as introduced by Krause and Pudlák [KP97]. We then show how certain hardness properties of the base function translate to related hardness properties of the lifted function. Then, we show how lifted functions can be embedded in certain 'simple' functions, if the base function was 'simple' itself. Finally, we list the consequences we obtain for lifting symmetric functions, which include resolving conjectures posed by Ada et al. [AFH12] and Zhang [Zha92].

3.3.1 Lifting Functions by the Krause-Pudlák Selector

In this section, we show how certain hardness properties of a function f can be amplified into other hardness properties of a particular lifted function obtained from f.

For any $f : \{-1,1\}^n \to \{-1,1\}$, define a function $f^{op} : \{-1,1\}^{3n} \to \{-1,1\}$ as follows.

$$f^{op}(x_1,\ldots,x_n,y_1,\ldots,y_n,z_1,\ldots,z_n) = f(u_1,\ldots,u_n) \tag{3.2}$$

where for all i, $u_i = (x_i \wedge z_i) \vee (y_i \wedge \bar{z_i})$.[3] Intuitively speaking, the value of z_i decides whether to feed x_i or y_i as the ith input to f. This method of lifting f was introduced by Krause and Pudlák [KP97]. The following lemma translates degree-hardness properties of f into other monomial/weight-hardness properties of f^{op}. The proof of this lemma is based on ideas from [KP97].

Lemma 3.3.1. Let $f : \{-1,1\}^n \to \{-1,1\}$ be any function.

1. $\varepsilon_d(f) > 1 - 2^{-d}$ for some $d \geq 2 \implies m(f^{op}) \leq 2^{-d+1}$.
2. $\mathrm{mon}_{\pm}(f^{op}) \geq 2^{\mathsf{deg}_{\pm}(f)}$.
3. $\mathsf{wt}_{1/3}(f^{op}) \geq 2^{\widetilde{\mathsf{deg}}_{2/3}(f)-1}$.

Proof. We first prove part 1.

Let p be a polynomial of weight 1 representing f^{op} with margin at least $\frac{1}{2^{d-1}}$, and say $p = \sum_{S \subseteq [n]\times[n]\times[n]} w_S \chi_S$. Recall that f^{op} (and also p) has $3n$ input variables. For this proof, we view the input variables as $\{x_{j,1}, x_{j,2}, z_j | j \in \{1,\ldots,n\}\}$, where z_i's are the 'selector' variables.

For any fixing of the z variables, define a relevant variable to be one that is 'selected' by z. Thus, for each $j \in \{1,\ldots,n\}$, exactly one of $\{x_{j,1}, x_{j,2}\}$ is relevant. Analogously, define a relevant monomial to be one that contains only those variables selected by z. For a uniformly random fixing of z and any subset $S \subseteq [n]$ such that $|S| \geq d$,

$$\Pr_z[\chi_S \text{ is relevant}] \leq \frac{1}{2^d}.$$

[3] The interested reader may note that f^{op} is exactly the same function as f composed with the Indexing gadget.

Now since $\mathsf{wt}(p) = 1$, we have

$$\mathbb{E}_z[\text{weight of relevant monomials in } p\big|_z \text{ of degree at least } d] = \sum_{|S|\geq d} |w_S| \cdot \Pr_z[\chi_S \text{ is relevant}] \leq \frac{1}{2^d} \sum_{|S|\geq d} |w_S| \leq \frac{1}{2^d}.$$

Thus, there exists a fixing of the z variables such that the weight of the relevant monomials of degree at least d in $p\big|_z$ is at most $\frac{1}{2^d}$. Select this fixing of z.

- Note that $p\big|_z$ is a polynomial on only the variables $\{x_{i,1}, x_{i,2} | i \in \{1, \ldots, n\}\}$. Drop the relevant monomials of degree at least d from $p\big|_z$ to obtain a polynomial p_1.
- Observe that p_1 sign represents $f^{op}\big|_z$ with margin at least $\frac{1}{2^{d-1}} - \frac{1}{2^d} = \frac{1}{2^d}$.
- For each $j \in \{1, \ldots, n\}$, denote the irrelevant variable by x_{j,i_j}. Consider the polynomial p_2 on n variables defined by $p_2 = \mathbb{E}_{x_{1,i_1},\ldots,x_{n,i_n}}[p_1]$, where the expectation is over each irrelevant variable being sampled uniformly and independently from $\{-1, 1\}$.
- It is easy to see that any monomial containing an irrelevant variable in p_1 vanishes in p_2. Also note that p_2 is a polynomial of degree at most d, and it must sign represent f with margin at least $\frac{1}{2^d}$. This leads to a contradiction since we assumed that $\varepsilon_d(f) > 1 - \frac{1}{2^d}$.

Part 2 was proved in [KP97]. Its proof, and the proof of Part 3 follow along extremely similar lines as the above proof, and we omit them. □

3.3.2 Lifts as Projections of Simpler Functions

In this section, we show how lifts of threshold (and symmetric) functions can be viewed as the projections of threshold (symmetric) functions.

Definition 3.3.2 (Monomial projection)**.** We call a function $g : \{-1, 1\}^m \to \{-1, 1\}$ a *monomial projection* of a function $f : \{-1, 1\}^n \to \{-1, 1\}$ if $g(x_1, \ldots, x_m) = f(M_1, \ldots, M_n)$, where each M_i is a monomial in the variables $x_1, \ldots, x_m$.

Remark 3.3.3. Note that since our input domain is $\{-1, 1\}^m$, a monomial $M = \Pi_{i\in S} x_i$ is the same as the XOR of the variables in the monomial, that is $M = \bigoplus_{i\in S} x_i$.

The following lemma is an easy consequence of definitions.

Lemma 3.3.4. For any functions $f : \{-1,1\}^n \to \{-1,1\}$ and $g : \{-1,1\}^m \to \{-1,1\}$ such that g is a monomial projection of f, and any $\epsilon > 0$, we have

$$\begin{aligned} m(f) &\leq m(g), \\ \mathrm{mon}_{\pm}(g) &\leq \mathrm{mon}_{\pm}(f), \\ \mathsf{wt}(g) &\leq \mathsf{wt}(f), \\ \mathsf{wt}_{\epsilon}(g) &\leq \mathsf{wt}_{\epsilon}(f). \end{aligned}$$

We first show that any lifted threshold function can be viewed as a monomial projection of a threshold function with a similar number of input variables. This proof is based on methods of [Kra06].

Lemma 3.3.5. Given any linear threshold function $f : \{-1,1\}^n \to \{-1,1\}$, there exists a linear threshold function $f' : \{-1,1\}^{4n} \to \{-1,1\}$ such that f^{op} is a monomial projection of f.

Proof. Let $f : \{-1,1\}^n \to \{-1,1\}$ be a linear threshold function such that $m(f^{op}) \leq \delta$. Fix a threshold representation for f, that is $f(x) = \mathsf{sgn}\left(\sum_{i=1}^{n} w_i x_i\right)$. Note that

$$\begin{aligned} f^{op}(x,y,z) &= \mathsf{sgn}\left(\sum_{i=1}^{n} w_i \left(\frac{x_i(1-z_i)}{2} + \frac{y_i(1+z_i)}{2}\right)\right) \\ &= \mathsf{sgn}\left(\sum_{i=1}^{n} w_i (x_i + y_i - x_i z_i + y_i z_i)\right). \end{aligned}$$

Consider a linear threshold function $f' : \{-1,1\}^{4n} \to \{-1,1\}$ defined as

$$f'(x,y,u,v) = \mathsf{sgn}\left(\sum_{i=1}^{n} w_i (x_i + y_i - u_i + v_i)\right).$$

Clearly, f^{op} is a monomial projection of f'. □

Lemma 3.3.6. Given a symmetric function $F : \{-1,1\}^{4n} \to \{-1,1\}$, defined by the predicate $D_F : [n] \to \{-1,1\}$, define a symmetric function $f : \{-1,1\}^n \to \{-1,1\}$ defined by the predicate $D_f(b) = D_F(2b+n)$ for all $b \in \{0,1,\ldots,n\}$. Then, f^{op} is a monomial projection of F.

Proof. Let $g : \{-1,1\}^{3n} \to \{-1,1\}$ be defined as follows.

$$g(x_1, \dots x_n, y_1, \dots, y_n, z_1, \dots, z_n) \\ = F(x_1, \dots, x_n, y_1, \dots, y_n, -x_1z_1, \dots, -x_nz_n, y_1z_1, \dots, y_nz_n).$$

Clearly, g is a monomial projection of F. We show now that $g = f^{op}$.

For every input to g and each $i \in [n]$, define the i'th *relevant* variable to be x_i if $z_i = -1$ (define y_i to be the *irrelevant* variable in this case), and y_i if $z_1 = 1$ (x_i is irrelevant in this case). Suppose there are b many relevant variables with value -1 on a fixed input $x_1, \dots, x_n, y_1, \dots, y_n, z_1, \dots, z_n$ and $n - b$ relevant variables with value 1. Say $(x_1, \dots, x_n, y_1, \dots, y_n, -x_1z_1, \dots, -x_nz_n, y_1z_1, \dots, y_nz_n)$ contains a many -1's. Then,

$$4n - 2a = \sum_{i=1}^{n} x_i + y_i - x_iz_i + y_iz_i = \sum_{i=1}^{n} x_i(1 - z_i) + y_i(1 + z_i) = 2n - 4b \\ \implies a = 2b + n.$$

Thus,

$$g(x_1, \dots, x_n, y_1, \dots, y_n, z_1, \dots, z_n) = D_F(2b + n) = D_f(b) \\ = f^{op}(x_1, \dots, x_n, y_1, \dots, y_n, z_1, \dots, z_n).$$

The last equality follows from Equation 3.2. □

In fact, the proof of Lemma 3.3.6 can be seen to imply the following lemma.

Lemma 3.3.7. Given a symmetric function $f : \{-1,1\}^n \to \{-1,1\}$ defined by the predicate $D_f(b)$, define a function $F : \{-1,1\}^{4n} \to \{-1,1\}$ such that on inputs of Hamming weight $2b + n$ for some $b \in \{0, 1, \dots, n\}$, F takes the value $D_f(b)$, and F takes arbitrary values on inputs of Hamming weight not in $\{2b+n : b \in \{0, 1, \dots, n\}\}$. Then, f^{op} is a monomial projection of F.

3.3.3 Consequences for Symmetric Functions

In this section, we show consequences of hardness amplification of lifted symmetric functions.

We first prove Theorem 3.1.1.

Proof of Theorem 3.1.1.

- Assume that n is even and that $r-1$ is a multiple of 4. (If not, we can fix a constant number of input bits). Note that $D_F(r-1) \neq D_F(r+1)$. Further assume $r_0(F) > r_1(F)$. Define $F' : \{0,1\}^{2r} \to \{-1,1\}$ by $D_{F'}(i) = D_F(i)$. It suffices to show $\log \mathsf{wt}_{1/3}(F') \geq c'r$ for some universal constant $c' > 0$. (If $r_1(F) \geq r_0(F)$, define $F' : \{0,1\}^{2r} \to \{-1,1\}$ by $D_{F'}(i) = D_F(4n-2r+i)$, and an analogous argument to the one that follows can be carried out. Define $f : \{0,1\}^{(r-1)/2} \to \{-1,1\}$ by $D_f(i) = D_{F'}(2i + (r-1)/2)$. By Lemma 3.3.6, f^{op} is a monomial projection of F'. Note that $D_f\left(\frac{r-1}{4}\right) \neq D_f\left(\frac{r-1}{4}+1\right)$, and thus $\Gamma(f) \leq 1$. By Theorem 3.2.5, $\widetilde{\mathsf{deg}}_{2/3}(f) = \Theta(r)$.

  Using Lemma 3.3.1 and Lemma 3.3.4, we conclude that there exists a universal constant $c_1 > 0$ such that

$$\log(\mathsf{wt}_{1/3}(F)) \geq \log(\mathsf{wt}_{1/3}(F')) \geq \log(\mathsf{wt}_{1/3}(f^{op})) \geq c_1 r. \tag{3.3}$$

- Consider any symmetric function $F : \{-1,1\}^{4n} \to \{-1,1\}$ such that $\mathsf{deg}_{oe}(F) \geq 4j$ where $j \geq 4$. Assume that there are at least $2j$ many $(i, i+2)$ sign changes in $[0, 3n]$. Further assume that at least j of them occur when i's are even integers (if not, set one variable to -1). Define a family of symmetric functions $\{f_i : \{-1,1\}^{\frac{4n}{3^i}} \to \{-1,1\} : i \in \{0,1,\ldots,\lceil \frac{1}{\log 3} \log\left(\frac{2n}{j}\right) \rceil\}\}$ as follows.

$$\forall b \in \left[\frac{4n}{3^i}\right],\ D_{f_i}(b) = D_F\left(2b + \frac{4n}{3^i}\right).$$

  (If there were less than j many $(i, i+2)$ sign changes in $[0, 3n]$ for even integers i, then there must be at least j many $(i, i+2)$ sign changes in $[n, 4n]$. In this case, define $D_{f_i}(b) = D_F\left(4n - 2b - \frac{4n}{3^i}\right)$, and an argument similar to the one that follows can be carried out).

  Note that the sign degree of f_i equals the number of $(k, k+2)$ sign changes in the spectrum of F in the interval $[\frac{n}{3^i}, \frac{n}{3^{i-1}}]$. Since D_F has at least $\lfloor j/2 \rfloor$ many $(k, k+2)$ sign changes in the interval $[\lfloor j/2 \rfloor, 3n]$, this implies that at least one of the f_i's has at least $\frac{\lfloor j \rfloor/2}{\lceil \frac{1}{\log 3} \log\left(\frac{2n}{j}\right) \rceil}$ many $(k, k+1)$ sign changes (sign degree). Using Lemma 3.3.6, Lemma 3.3.1 and Lemma 3.3.4, we obtain that there exists a constant $c_2 > 0$ such that

$$\mathrm{mon}_{\pm}(F) \geq 2^{c_2 j}.$$

- The proof of the Part 3 follows along extremely similar lines as that of Part 2, and we omit it.

□

We next prove Theorem 3.1.2, resolving a conjecture of Ada et al. [AFH12].

Proof of Theorem 3.1.2. It follows as a direct consequence of Part 1 of Theorem 3.1.1 and the upper bound in Theorem 3.2.9. □

Finally, we prove Theorem 3.1.3 here, settling a conjecture of Zhang [Zha92].

Proof of Theorem 3.1.3. The upper bound follows from Theorem 3.2.4. It suffices to show a lower bound for when $\mathsf{deg}_{oe}(f) \geq 16$. The lower bound follows from Part 2 of Theorem 3.1.1 in this case.

□

3.4 Discrepancy of XOR Functions

In this section, we analyze the discrepancy of XOR functions.

3.4.1 Margin-Discrepancy Equivalence

In this section, we prove Theorem 3.1.4, which is a necessary and sufficient approximation theoretic condition of f in order for $f \circ \mathsf{XOR}$ to have small discrepancy.

Proof of Theorem 3.1.4. We first show that $m(f) \leq m(f \circ \mathsf{XOR})$. For notational convenience, let us denote $f \circ \mathsf{XOR}$ by F. View f's inputs as $x_1, \ldots x_n$, and F's inputs as $y_1, \ldots, y_n, z_1, \ldots, z_n$, where f is fed $y_1 \oplus z_1, \ldots, y_n \oplus z_n$. Let p be any polynomial of weight 1 sign representing f. Replace every variable x_i in p by $y_i z_i$. Clearly, the new polynomial obtained sign represents F with the same margin as p represented f, and the weight remains unchanged. Thus, $m(f) \leq m(F)$.

Next, we show that $m(F) \leq 4\mathsf{disc}(F)$. Let λ denote a distribution under which $\mathsf{disc}_\lambda(F) = \mathsf{disc}(F)$, and let $P(x, y) = \sum_{S \subseteq [2n]} c_S \chi_S(x, y)$ be a polynomial of weight 1, which sign represents F.

$$
\begin{aligned}
m(F) &\leq \mathbb{E}_\lambda[F(x,y)P(x,y)] \\
&\leq \mathbb{E}_\lambda\left[F(x,y)\sum_{S\subseteq[2n]} c_S\chi_S(x,y)\right] \\
&\leq \left(\sum_{S\subseteq[2n]}|c_S|\right)\cdot \max_{S\subseteq[2n]}\left(|\mathbb{E}_\lambda[F(x,y)\chi_S(x,y)]|\right) \\
&\leq \left|\sum_{\substack{\chi_S(x)=1\\ \chi_S(y)=1}} F(x,y)\lambda(x,y)\right| + \left|\sum_{\substack{\chi_S(x)=1\\ \chi_S(y)=-1}} F(x,y)\lambda(x,y)\right| + \left|\sum_{\substack{\chi_S(x)=-1\\ \chi_S(y)=1}} F(x,y)\lambda(x,y)\right| \\
&\quad + \left|\sum_{\substack{\chi_S(x)=-1\\ \chi_S(y)=-1}} F(x,y)\lambda(x,y)\right| \\
&\leq 4\mathsf{disc}(F).
\end{aligned}
$$

Thus, $m(F) \leq 4\mathsf{disc}(F)$.

Now we show that $\mathsf{disc}(F) \leq m(f)$.

Let us first write a program whose optimal value corresponds to the margin of f.

Variables	$\Delta, \{\alpha_S : S \subseteq [n]\}$		
Maximize	Δ		
s.t.	$f(x)\sum_{S\subseteq[n]} \alpha_S\chi_S(x)$	$\geq \Delta$	$\forall x \in \{-1,1\}^n$
	$\sum_{S\subseteq[n]} \lvert\alpha_S\rvert$	≤ 1	
	$\Delta \in \mathbb{R}$		
	$\alpha_S \in \mathbb{R}$		$\forall S \subseteq [n]$

We write another linear program, which is easier to work with.

Variables	$\Delta, \{\alpha'_S : S \subseteq [n]\}, \{\alpha''_S : S \subseteq [n]\}$		
Maximize	Δ		
s.t.	$f(x) \sum_{S \subseteq [n]} \chi_S(x)(\alpha''_S - \alpha'_S)$	$\geq \Delta$	$\forall x \in \{-1,1\}^n$
	$\sum_{S \subseteq [n]} (\alpha'_S + \alpha''_S)$	≤ 1	
	$\Delta \in \mathbb{R}$		
	$\alpha'_S, \alpha''_S \geq 0$		$\forall S \subseteq [n]$

Note that any solution to the first program is a valid solution to the second one, by setting one of α'_S or α''_S to 0, and the other to $|\alpha_S|$ for each $S \subseteq [n]$. We can also assume that a solution to the second program must have $\alpha'_S = 0$ or $\alpha''_S = 0$ for each $S \subseteq [n]$. If this was not the case, one could reduce the values of α'_S and α''_S by the same amount, thus not changing the value of $\alpha''_S - \alpha'_S$, and not violating any constraints. This gives us a solution to the first program by setting $\alpha_S = \alpha''_S$ if $\alpha''_S \neq 0$, and $\alpha_S = \alpha'_S$ otherwise. Thus, the optima of the two programs above are equal.

Let us now look at the corresponding dual to the above linear program. Notice that the program looks like a minimization problem with the objective to minimize $\max_{S \subseteq [n]} \left|\widehat{f\mu}(S)\right|$ under a variable distribution μ on $\{-1,1\}^n$.

Variables	$\epsilon, \{\mu(x) : x \in \{-1,1\}^n\}$		
Minimize	ϵ		
s.t.	$\lvert \sum_x \mu(x) f(x) \chi_S(x) \rvert$	$\leq \epsilon$	$\forall S \subseteq [n]$
	$\sum_x \mu(x)$	$= 1$	
	$\epsilon \geq 0$		
	$\mu(x) \geq 0$		$\forall x \in \{-1,1\}^n$

Thus, if f has margin at most δ, there exists a distribution μ on $\{-1,1\}^n$ such that $\left|\widehat{f\mu}(S)\right| \leq \frac{\delta}{2^n}$ for all $S \subseteq [n]$. Let $\mu^{\oplus}$ be a distribution denoting the lift of μ on $\{-1,1\}^n \times \{-1,1\}^n$. That is, $\mu^{\oplus}(x,y) = \frac{1}{2^n}\mu(x \oplus y)$. We now show that the discrepancy of F is small under $\mu^{\oplus}$. For matrices A, B, let $A \circ_H B$ denote the Hadamard (entrywise) product of A and B. Note that under the distribution $\mu^{\oplus}$, the

discrepancy of F is

$$\begin{aligned}\mathsf{disc}_{\mu^{\oplus}}(F) &= \max_{S\subseteq[n],T\subseteq[n]} 1_S^T(\mu^{\oplus} \circ_H F)1_T \\ &\leq ||\mu^{\oplus} \circ_H F|| \cdot 2^n. && \text{Cauchy-Schwarz}\end{aligned}$$

Thus,

$$\mathsf{disc}_{\mu^{\oplus}}(F) \leq \frac{||f\mu \circ \mathsf{XOR}||}{2^n} \cdot 2^n = 2^n \cdot ||\widehat{f\mu}||_\infty \leq \delta.$$

Here, the first inequality follows from the definition of $\mu^{\oplus}$, and the following equality follows from Lemma 2.2.3. This proves the claim. □

We remark here that Linial and Shraibman [LS09c] had shown a similar equivalence between the discrepancy of a matrix (the communication matrix of the target function) and its margin. This margin refers to the margin of the matrix, and not the polynomial margin of the base function. However, since we do not use this notion in the rest of this thesis, we overload notation and use $m(A)$ to denote the margin of the matrix A. Define the margin of an $m \times n$ sign matrix A as

$$m(A) = \sup \min_{i,j} \frac{|\langle x_i, y_j\rangle|}{||x_i||_2||y_j||_2}$$

where the supremum is over all choices of $x_1, \ldots, x_m, y_1, \ldots, y_n \in \mathbb{R}^{m+n}$ such that $sgn(\langle x_i, y_j\rangle) = a_{i,j}$ for all i, j. Linial and Shraibman [LS09c] showed that the margin of a sign matrix is equivalent to its discrepancy up to a constant factor.

Theorem 3.4.1 ([LS09c] Thm 3.1)**.** For every sign matrix A,

$$\mathsf{disc}(A) \leq m(A) \leq 8\mathsf{disc}(A).$$

We now note that Theorem 3.1.4 implies the first inequality of Theorem 3.4.1 for the special case of XOR functions.

Claim 3.4.2. Let $f : \{-1, 1\}^n \to \{-1, 1\}$. Then,

$$m(f) \leq m(M_{f\circ\mathsf{XOR}}).$$

Proof. Let $p = \sum_{S\subseteq[n]} c_S\chi_S$ be a polynomial which sign represents f with margin δ. This implies $p'(x, y) = \sum_{S\subseteq[n]} c_S\chi_S(x)\chi_S(y)$ sign represents $f \circ \mathsf{XOR}$ with margin δ.

We will exhibit 2^{n+1} vectors, $\{u_T : T \subseteq [n]\}$ and $\{v_T : T \subseteq [n]\}$ in $\mathbb{R}^{2^n}$ such that $m(M_{f\circ\mathsf{XOR}}) \geq \delta$. Index the coordinates by characteristic sets, $T \subseteq [n]$. For a set

$T \subseteq [n]$, we use w_T to denote the corresponding characteristic vector in $\mathbb{R}^{2^n}$. Define $u_T(S) = v_T(S) = \sqrt{c_S}\chi_S(w_T)$.

Since $\mathsf{wt}(p') = 1$, $||u_T||_2 = ||v_T||_2 = 1$. Also, $\langle u_{T_1}, v_{T_2}\rangle = \sum_{S\subseteq[n]} c_S\chi_S(w_{T_1} \oplus w_{T_2}) \geq \delta$ since p' sign represents $f \circ \mathsf{XOR}$ with margin δ.

Thus,

$$m(M_{f\circ\mathsf{XOR}}) = \sup \min_{T_1,T_2} \frac{|\langle u_{T_1}, v_{T_2}\rangle|}{||u_{T_1}||_2||v_{T_2}||_2} \geq \delta.$$

□

3.4.2 A New Separation of PP from UPP

In this section, we show here how to obtain an alternate proof that the GHR function has large PP complexity. Since $\mathsf{GHR} \in \mathsf{THR} \circ \mathsf{XOR}$, Claim 2.3.10 implies $\mathsf{GHR} \in \mathsf{UPP}$.

Proof of Theorem 3.1.6. Theorem 3.2.3 and Lemma 3.3.1 show the existence of a linear threshold function $f : \{-1,1\}^n \to \{-1,1\}$ such that $m(f^{op}) \leq 2^{-cn}$ for some absolute constant $c > 0$. Lemma 3.3.4 and Lemma 3.3.5 then show existence of a linear threshold function $f' : \{-1,1\}^{4n} \to \{-1,1\}$ such that $m(f') \leq 2^{-cn}$. Using Theorem 3.1.4 and Theorem 2.3.8, we already obtain the existence of a linear threshold function $f' : \{-1,1\}^{4n} \to \{-1,1\}$ such that $\mathsf{PP}(f' \circ \mathsf{XOR}) \geq c'n$ for some absolute constant $c' > 0$.

By Fact 3.2.2, one can embed f' in the universal threshold function by blowing up the number of variables by a quadratic factor (note that we do not lose a logarithmic factor as stated in Fact 3.2.2, because it can be verified that the weights of f' are at most $2^{\alpha n}$ for an absolute constant $\alpha > 0$). Thus, $m(\mathsf{UTHR}) \leq 2^{-\Omega(\sqrt{n})}$. By Theorem 3.1.4 and Theorem 2.3.8, we have

$$\mathsf{PP}(\mathsf{GHR}) = \Omega(\sqrt{n}).$$

□

3.4.3 PM is Harder than XOR

In this section, we observe that if $f \circ \mathsf{XOR}$ has small discrepancy, then so does $f \circ \mathsf{PM}$. Note that the converse is not true, since the inner product function is a large subfunction of $\oplus \circ \mathsf{PM}$, which has exponentially small discrepancy, but $\oplus \circ \mathsf{XOR}$ has extremely large discrepancy, since it is just the Parity function.

Theorem 3.4.3. Let $f : \{-1,1\}^n \to \{-1,1\}$. Then,

$$\mathsf{disc}(f \circ \mathsf{XOR}) < \delta \implies \mathsf{disc}(f \circ \mathsf{PM}) \leq \sqrt{4\delta n}.$$

Proof. Consider $f \circ \mathsf{PM}$ and substitute $d = n$ in Theorem 3.2.11 to obtain

$$\mathsf{disc}(f \circ \mathsf{PM}) \leq \left(\frac{n}{W(f, d-1)}\right)^{1/2}.$$

By Theorem 3.1.4, $\mathsf{disc}(f \circ \mathsf{XOR}) < \delta \implies m(f) < 4\delta$. Suppose $W(f, n-1) \leq \frac{1}{4\delta}$. This would show existence of a polynomial with integer weights, say $\sum_{S \subseteq [n]} \lambda_S \chi_S$, sign representing f, and with total weight at most $1/4\delta$. This in turn implies existence of a polynomial of weight 1, $p = \frac{\sum_{S \subseteq [n]} \lambda_S \chi_S}{\sum_{S \subseteq [n]} |\lambda_S|}$, which sign represents f with margin at least 4δ, which is a contradiction. Thus,

$$\mathsf{disc}(f \circ \mathsf{PM}) \leq \sqrt{4\delta n}.$$

□

Using the equivalence between PP and discrepancy (Theorem 2.3.8), we obtain the following lower bound for the PP complexity of $f \circ \mathsf{PM}$ in terms on the PP complexity of $f \circ \mathsf{XOR}$.

Corollary 3.4.4. For any $f : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$,

$$\mathsf{PP}(f \circ \mathsf{PM}) = \Omega(\mathsf{PP}(f \circ \mathsf{XOR}) - \log n).$$

3.4.4 Symmetric Functions with Large Odd-Even Degree

We show that for any symmetric function F, $\mathsf{PP}(F \circ \mathsf{XOR})$ is bounded below by $\mathsf{deg}_{oe}(F)$ (up to a logarithmic factor in the input size).

Proof of Theorem 3.1.5. Using Theorem 3.1.4 and Part 3 of Theorem 3.1.1, we obtain that there exists a universal constant $c > 0$ such that $\mathsf{PP}(F \circ \mathsf{XOR}) \geq cr/\log(n/r)$, which proves Theorem 3.1.5. □

3.4.5 An Upper Bound

In this section, we show that for any symmetric function $f : \{-1,1\}^n \to \{-1,1\}$, the PP complexity of $f \circ \mathsf{XOR}$ is bounded above by essentially $\mathsf{deg}_{oe}(f)$. Our proof follows

along the lines of Zhang [Zha92] who showed that a symmetric function with small odd-even degree has a small Threshold of Parity circuit representation.

Theorem 3.4.5. Suppose $f : \{-1,1\}^n \to \{-1,1\}$ is a symmetric function defined by the predicate $D_f : \{0,1,\ldots,n\} \to \{-1,1\}$. Say the odd-even degree of f equals k and n is even. Then,

$$\mathsf{PP}(f \circ \mathsf{XOR}) = O(k \log n).$$

Proof. Define $S_{even} = \{i \in \{0,2,\ldots,n\} : D_f(i) \neq D_f(i+2)\}$, and define $S_{odd} = \{i \in \{1,3,\ldots,n-1\} : D_f(i) \neq D_f(i+2)\}$. By our assumption, $|S_{even}|, |S_{odd}| \leq k$.

Consider the polynomials $p_{even}, p_{odd} : \{-1,1\}^n \to \mathbb{R}$ defined by

$$p_{even}(x) = D_f(0) \cdot \prod_{i \in S_{even}} \left(n - 2i + 1 - \left(\sum_{j=1}^{n} x_j \right) \right)$$

and

$$p_{odd}(x) = D_f(1) \cdot \prod_{i \in S_{odd}} \left(n - 2i + 1 - \left(\sum_{j=1}^{n} x_j \right) \right).$$

The polynomial $p : \{-1,1\}^n \to \mathbb{R}$ defined by

$$p(x) = (1 + \chi_{[n]}(x))p_{even}(x) + (1 - \chi_{[n]}(x))p_{odd}(x)$$

sign represents f on $\{-1,1\}^n$.

We now use the simple observations that $\mathsf{wt}(q_1 \cdot q_2) \leq \mathsf{wt}(q_1) \cdot \mathsf{wt}(q_2)$ and $\mathsf{wt}(q_1 + q_2) \leq \mathsf{wt}(q_1) + \mathsf{wt}(q_2)$. Thus,

$$\begin{aligned} \mathsf{wt}(p) &\leq 2\mathsf{wt}(p_{even}) + 2\mathsf{wt}(p_{odd}) \\ &\leq 2(2n)^k + 2(2n)^k \\ &\leq 4(2n)^k. \end{aligned}$$

Note that all the coefficients of p are integer valued. Thus, the polynomial $p' = \frac{p}{\mathsf{wt}(p)}$ is a polynomial of weight 1, which sign represents f with margin at least $\frac{1}{\mathsf{wt}(p)}$. By Theorem 3.1.4 and Theorem 2.3.8,

$$\mathsf{PP}(f \circ \mathsf{XOR}) \leq O(\log(\mathsf{wt}(p))) \leq O(k \log n).$$

□

3.5 Bounded-Error Communication Complexity of XOR Functions

In this section, we analyze the BPP complexity of XOR functions.

Proof of Theorem 3.1.7. We first write a program which captures the best error a weight w polynomial can achieve in approximating a given function f.

$$
\begin{array}{lll}
\text{Variables} & \epsilon, \{\alpha_S : S \subseteq [n]\} & \\
\text{Minimize} & \epsilon & \\
\text{s.t.} & \left| f(x) - \sum_{S \subseteq [n]} \alpha_S \chi_S(x) \right| \leq \epsilon & \forall x \in \{-1,1\}^n \\
& \sum_{S \subseteq [n]} |\alpha_S| \leq w & \\
& \epsilon \geq 0 & \\
& \alpha_S \in \mathbb{R} & \forall S \subseteq [n]
\end{array}
$$

By manipulations similar to those in Section 3.4.1, we obtain the following dual program.

$$
\begin{array}{lll}
\text{Variables} & \Delta, \{\mu(x) : x \in \{-1,1\}^n\} & \\
\text{Maximize} & \sum_x f(x)\mu(x) - \Delta w & \\
\text{s.t.} & |\sum_x \mu(x)\chi_S(x)| \leq \Delta & \forall S \subseteq [n] \\
& \sum_x \mu(x) \leq 1 & \\
& \Delta \geq 0 & \\
& \mu(x) \geq 0 & \forall x \in \{-1,1\}^n
\end{array}
$$

By strong linear programming duality, the optima of the two programs above are equal. Let us call the optimal value OPT, which is clearly non-negative. Note that in any feasible solution to the dual, $1 - \Delta w \geq \sum_x f(x)\mu(x) - \Delta w \geq 0$. This implies $\Delta \leq \frac{1}{w}$. Suppose a function $f : \{-1,1\}^n \to \{-1,1\}$ satisfied $\mathsf{wt}_{1/3}(f) = w'$. This means if we fix $w = w'$ in the programs, then $\mathsf{OPT} = 1/3$, which implies $\sum_x f(x)\mu(x) \geq 1/3$ since Δ is non-negative. Thus, any optimum solution to the dual

must satisfy $\sum_x \mu(x) \geq 1/3$. Define a distribution μ' by $\mu'(x) = \frac{\mu(x)}{\sum_{x \in \{-1,1\}^n} \mu(x)}$, and we obtain $|\sum_x \mu'(x)\chi_S(x)| \leq \frac{3}{w'}$ (hence, setting $\Delta = \frac{3}{w'}$ gives us a feasible solution).

Write $\mu' = g \cdot \nu$ uniquely, where $g : \{-1,1\}^n \to \{-1,1\}$ is a boolean function and $\nu : \{-1,1\}^n \to [0,1]$ is a distribution on the inputs. Thus, $\mathrm{corr}_\nu(f,g) \geq 1/3$ (which implies $\mathrm{corr}_{\nu^\oplus}(f \circ \mathsf{XOR}, g \circ \mathsf{XOR}) \geq 1/3$), and

$$\mathsf{disc}_{\nu^\oplus}(g \circ \mathsf{XOR}) \leq \frac{||g\nu \circ \mathsf{XOR}||}{2^n} \cdot 2^n = 2^n \cdot ||\widehat{g\nu}||_\infty \leq \Delta \leq \frac{3}{w'}.$$

This, along with Theorem 3.2.13 implies

$$R_{2/5}(f \circ \mathsf{XOR}) \geq \log w' - 4,$$

proving Theorem 3.1.7. □

Using Part 1 of Theorem 3.1.1 and Theorem 3.1.7, we obtain a new proof of Theorem 3.1.9.

3.6 References

The results presented in this chapter are based on joint work with Arkadev Chattopadhyay ([CM17b] and some additional results from [CM17a]).

Chapter 4

Unbounded-Error Communication

4.1 Introduction

4.1.1 Sign Rank

Recall that the sign rank of a $\{-1, 1\}$ valued matrix M is defined to be the minimum rank of a real valued matrix each of whose entries agrees in sign with the corresponding entry of M. Although sign rank has found numerous applications in various areas of computer science, we are interested in its applications to communication complexity and boolean circuit complexity. Paturi and Simon [PS86] showed that the logarithm of the sign rank of a (communication) matrix is essentially equivalent to the unbounded-error 2-party communication complexity of the underlying function (Theorem 2.3.9). Forster et al. [FKL+01] showed that proving lower bounds on the sign rank of a function gives lower bounds on the minimum size of any $\mathsf{THR} \circ \mathsf{MAJ}$ circuit computing it. Sign rank is known to be equivalent to dimension complexity, a geometric notion that is of fundamental importance in computational learning theory. Even proving lower bounds on the sign rank of a random function is non-trivial and was first done by Alon et al. [AFR85]. On the other hand, proving strong lower bounds on the sign rank of an explicit function, IP, was a breakthrough achieved by Forster [For02] fifteen years later. Since that work, there have relatively been just a few results proving strong sign rank lower bounds on explicit functions [She11b, RS10, BT16, BCH+16]. While many basic questions about sign rank remain unanswered, new connections between it and other areas of mathematics keep showing up (see for example [AMY16]).

An active research program is to search for functions in AC^0 that are increasingly hard to *approximate* under various natural measures. For example, a recent result of Bun and Thaler [BT17] gave almost optimal bounds for approximate degree, and

Sherstov [She15] gave the best known lower bounds on sign degree for functions in AC^0. Sign rank is arguably one of the hardest notions of approximability to analyze. Razborov and Sherstov [RS10] exhibited a function computable by a *depth-3* read-once formula that has large sign rank. Improving this, Bun and Thaler [BT16] gave the strongest known lower bound of $2^{\tilde{\Omega}(n^{2/5})}$ on the sign rank of a function in AC^0. All these results exploit the considerable computing power of AC^0 to come up with more intricate functions that are harder to approximate. Our work contrasts with these efforts by finding, in some sense, a *simpler* function, still in AC^0, that has large sign rank.

4.1.2 Low-Depth Threshold Circuits

Linear threshold functions (LTF's) form one of the most central classes of Boolean functions that are studied. Every such function corresponds to the halfspace induced by a real weight vector $\mathbf{w} \in \mathbb{R}^{n+1}$ denoted by $\mathsf{THR}_{\mathbf{w}}$ in the following way: For each $x \in \{-1,1\}^n$,

$$\mathsf{THR}_{\mathbf{w}}(x) = \mathsf{sgn}\left(w_0 + \sum_{i=1}^{n} w_i x_i\right).$$

It is well known [Mur71] that for every threshold gate with n inputs, there exists a threshold representation for it that uses only integer weights of magnitude at most $2^{O(n \log n)}$. The power of an LTF depends on the magnitude of the weights allowed. For instance, the Boolean function $\mathrm{GT}(x, y)$ that determines if the n-bit integer x is at least as large as the n-bit integer y is an example of an LTF that has no representation as an LTF with sub-exponentially small weights. Indeed in various areas, several questions and problems have been solved when the LTF's arising in the study are restricted to have small weights, but extending them to unrestricted weights are either open or have been solved after spending much research efforts. Examples of such areas are learning theory [KOS04, She13b], pseudorandom generators [ST17] , analysis of Boolean functions [HKM12] and Boolean circuit complexity [COS17]. Understanding the relative power of large weights vs. small weights in the context of small-depth circuits having LTF's as gates has attracted attention by several works [AM05, GHR92, SB91, HP10, HP15, Raz92a, HMP+93, Hof96, GK98].

The class of all Boolean functions that can be computed by circuits of depth d and polynomial size, comprising gates computing LTF's (of polynomially bounded weights), is denoted by LT_d ($\widehat{LT}_d$). The seminal work of Minsky and Papert [MP69] showed that a simple function, Parity, is not in LT_1. While it is not very hard to

verify that Parity is in $\widehat{LT}_2$, an outstanding problem is to exhibit an explicit function that is not in LT_2. This problem is now a well-identified frontier for research in circuit complexity. LT_2 is one of the smallest known boolean circuit classes against which no strong lower bounds are known.

By contrast, the relatively early work of Hajnal et al. [HMP+93] established the fact that the Inner-Product modulo 2 function (denoted by IP), that is easily seen to be in $\widehat{LT}_3$, is not in $\widehat{LT}_2$. It turns out that there is a natural class sitting between $\widehat{LT}_2$ and LT_2, denoted by $\mathsf{THR} \circ \mathsf{MAJ}$, where the top THR gate has unrestricted weights, but the weights of the bottom MAJ gates are restricted to be only polynomially large.

Goldmann et al. [GHR92] proved several interesting results, which implied the following structure.

$$\widehat{LT}_2 \overset{[\text{GHR92}]}{=} \mathsf{MAJ} \circ \mathsf{THR} \overset{[\text{GHR92}]}{\subsetneq} \mathsf{THR} \circ \mathsf{MAJ} \subseteq LT_2 \overset{[\text{GHR92}]}{\subseteq} \widehat{LT}_3.$$

In a breakthrough work, Forster [For02] showed that IP has sign rank $2^{\Omega(n)}$ for the natural partition of input variables. This yielded an exponential separation between $\mathsf{THR} \circ \mathsf{MAJ}$ and $\widehat{LT}_3$. This meant that at least one of the two containments $\mathsf{THR} \circ \mathsf{MAJ} \subseteq LT_2$ and $LT_2 \subseteq \widehat{LT}_3$ is strict. Alman and Williams [AW17] recently showed interesting upper bounds on the 'probabilistic sign-rank' for functions in LT_2. In contrast, Amano and Maruoka [AM05] and Hansen and Podolskii [HP10] state that *separating* $\mathsf{THR} \circ \mathsf{MAJ}$ from $\mathsf{THR} \circ \mathsf{THR} = LT_2$ would be an important step for shedding more light on the structure of depth-2 boolean circuits. However, as far as we know, there was no clear target function identified for the purpose of separating the two classes. No progress on this question was made until our work. We emphasize here that it is not a priori clear that these classes ought to be different, especially in light of Goldmann et al.'s result that $\mathsf{MAJ} \circ \mathsf{MAJ} = \mathsf{MAJ} \circ \mathsf{THR}$.

We show that indeed $\mathsf{THR} \circ \mathsf{MAJ} \subsetneq \mathsf{THR} \circ \mathsf{THR}$ and elaborate on this in Section 4.1.4.

4.1.3 Communication Complexity Frontiers

Functions whose communication matrix of dimension $2^n \times 2^n$ have sign rank bounded above by a quasi-polynomial in n were shown in [PS86] to correspond exactly to the complexity class UPP (see Theorem 2.3.9). The lower bound on the sign rank by Razborov and Sherstov [RS10] implied that PH (in fact, $\Pi_2\mathsf{P}$) contains functions with

large sign rank, rendering the sign rank technique essentially useless to prove lower bounds against the second level of the polynomial hierarchy.

Indeed, there is a rich landscape of communication complexity classes below the second level as discussed in a recent, almost exhaustive survey by Göös, Pitassi and Watson [GPW18]. UPP is the strongest communication complexity class against which we know how to prove explicit lower bounds. A natural question is until where the sign rank method continues to yield lower bounds. Refer to Chapter 2 for formal descriptions of various communication complexity classes.

Göös et al. [GPW18] conjectured two seemingly incomparable classes, $\mathsf{AM} \cap \mathsf{coAM}, \mathsf{S_2P}$ to contain functions of large sign rank. Each of these classes are contained, plausibly strictly, in $\mathsf{\Pi_2P}$. Bouland et al. [BCH+16] recently resolved the first conjecture, exhibiting a *partial* function in $\mathsf{AM} \cap \mathsf{coAM}$ which has large sign rank.[1] We provide a strong confirmation of the second conjecture by showing that even a sub-class of $\mathsf{S_2P}$ contains a *total* function of large sign rank. We elaborate on this in Section 4.1.4.

4.1.4 Our Work

In order to state our main result regarding unbounded-error communication, we first define decision lists.

Definition 4.1.1 (Decision lists)**.** A decision list of length k, is a sequence $D = (L_1, a_1), (L_2, a_2), \ldots, (L_k, a_k)$, where each $a_i \in \{-1, 1\}$, and L_k is the constant -1 function. The decision list computes a function $f : \{-1,1\}^n \to \{-1,1\}$ as follows. If $L_1(x) = -1$, then $f(x) = a_1$; elseif $L_2(x) = -1$, then $f(x) = a_2$, elseif $\ldots$, elseif $L_k(x) = -1$, then $f(x) = a_k$. That is,

$$f(x) = \bigvee_{i=1}^{k} \left(a_i \bigwedge L_i(x) \bigwedge_{j<i} \neg L_j(x) \right).$$

We now recall the definition of our hard function, F_n, defined in Section 1.5.1. F_n can be described as a decision list of 'Equalities' in the following way. The input, of $n = 2ml$ bits, is split into two disjoint parts, $X \in \{-1,1\}^{ml}$ and $Y \in \{-1,1\}^{ml}$. X and Y are further divided into l disjoint blocks as $X_1, \ldots, X_l, Y_1, \ldots, Y_l$, each of length m. The function F_n outputs -1 iff the largest index $i \in [l]$ for which $X_i = Y_i$ holds is an odd index. We set $m = l^{1/3} + \log l$. Note that F_n is a decision list of

[1] It still remains unknown if there are *total* functions in $\mathsf{AM} \cap \mathsf{coAM}$ that have large sign rank.

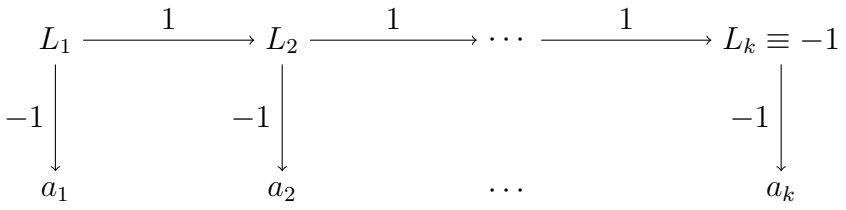

Figure 4.1: The decision list described in Definition 4.1.1

Equalities. This view of F_n is convenient while proving our communication class separation (Theorem 4.1.4).

F_n can also be viewed as an XOR function as follows. Consider a simple adaption of the well known $\mathsf{ODD\text{-}MAX\text{-}BIT}$ function [Bei94], which we define as follows. OMB^0_ℓ acts on an l bit input, and outputs -1 iff the rightmost bit set to 1 occurs at an odd index. OMB^0_ℓ is a linear threshold function, as explained by the following representation.

$$\mathsf{OMB}^0_\ell(x) = -1 \iff \sum_{i=1}^{\ell} (-1)^{i+1} 2^i \, (1 + x_i) \geq 0.5.$$

It is not hard to verify that $F_n = \mathsf{OMB}^0_\ell \circ \mathsf{OR}_{\ell^{1/3}+\log \ell} \circ \mathsf{XOR}_2$. This view of F_n helps while proving our main circuit complexity application (Theorem 4.1.3).

We show a strong lower bound on the sign rank of M_{F_n}, where the rows of M_{F_n} are indexed by the inputs X, the columns by Y, and the (x, y)th entry is $F_n(x, y)$. We overload notation and refer to the sign rank of M_{F_n} as the sign rank of F_n. The following is our main theorem regarding unbounded-error communication (Theorem 2.3.9 shows that $\mathsf{UPP}(F_n)$ is exactly characterized by the logarithm of its sign rank).

Theorem 4.1.2 (Main). The function F_n has sign rank $2^{\Omega(n^{1/4})}$.

A Separation of Threshold Circuit Classes

We first observe that F_n can be computed by linear sized $\mathsf{THR} \circ \mathsf{THR}$ formulas. For each $x \in \{-1, 1\}^n$, let $\mathsf{ETHR}_{\mathbf{w}}(x) = -1 \iff w_0 + w_1 x_1 + \cdots + w_n x_n = 0$. Thus, ETHR gates are also called exact threshold gates. By first observing that every function computed by a formula of the form $\mathsf{THR} \circ \mathsf{OR}$ can also be computed by a formula of the form $\mathsf{THR} \circ \mathsf{AND}$ with a linear blow-up in size, it follows that F_n can be computed by linear size formulas of the form $\mathsf{THR} \circ \mathsf{AND} \circ \mathsf{XOR}_2$. Note that each $\mathsf{AND} \circ \mathsf{XOR}_2$ is computable by an ETHR gate. Hence, F_n is in $\mathsf{THR} \circ \mathsf{ETHR}$, a class

that Hansen and Podolskii [HP10] showed is identical to the class $\mathsf{THR} \circ \mathsf{THR}$. Since sign rank is a lower bound on $\mathsf{THR} \circ \mathsf{MAJ}$ circuit size (see Lemma 4.2.3), our main theorem (Theorem 4.1.2) and the above observation yield the following circuit class separation.

Theorem 4.1.3. The function F_n can be computed by linear sized $\mathsf{THR} \circ \mathsf{THR}$ formulas, but any $\mathsf{THR} \circ \mathsf{MAJ}$ circuit computing it requires size $2^{\Omega(n^{1/4})}$.

Theorem 4.1.3 provides the first explanation for why current lower bound methods fail to get traction with $\mathsf{THR} \circ \mathsf{THR}$. Interestingly, it also suggests some new paths along which progress can be made. This is discussed in Chapter 7.

A Separation of Communication Classes

The application of our main result to give a communication class separation was brought to our notice by Göös [Göö17]. In order to state our communication class separation, let us consider the complexity class P^{MA} that is contained in $\mathsf{S}_2\mathsf{P}$. A function is in P^{MA} if it can be computed by deterministic protocols of polylogarithmic cost, where Alice and Bob have oracle access to any function in MA. The function F_n under the natural input partition (recall that it can be expressed as a decision list of equalities) can be efficiently solved by P^{MA} protocols by an appropriate binary search, and querying an $\mathsf{OR} \circ \mathsf{EQ}$ oracle at each step. A formal description of this protocol is given in Protocol 1.

Since the logarithm of sign rank of f essentially equals $\mathsf{UPP}(f)$ (see Theorem 2.3.9), we prove the following as a consequence of Theorem 4.1.2.

Theorem 4.1.4. The function F_n witnesses the following communication complexity class separation.

$$\mathsf{P}^{\mathsf{MA}} \not\subseteq \mathsf{UPP}.$$

Our result thus strongly confirms a conjecture of Göös et al. by exhibiting the first *total function* in a complexity class contained, plausibly strictly, in $\Pi_2\mathsf{P}$, that has large sign rank. More precisely, our function F_n is in P^{MA}, a class seemingly well below $\mathsf{S}_2\mathsf{P}$. This yields Theorem 4.1.4.

On the other hand, it is known that $\mathsf{P}^{\mathsf{NP}} \subsetneq \mathsf{UPP}$ and $\mathsf{MA} \subsetneq \mathsf{PP} \subsetneq \mathsf{UPP}$. These facts combined with Theorem 4.1.4 shows that P^{MA} is right on the frontier between what we understand and what we do not. Thus, proving lower bounds against P^{MA} protocols emerges as a natural program for advancing the set of currently known techniques, given our work. Future directions are further discussed in Chapter 7.

4.1.5 Our Techniques

We strive to prove a lower bound on the sign rank of a function $F \in \mathsf{THR} \circ \mathsf{THR}$. We are guided by the equivalence of sign rank and unbounded-error communication complexity, due to Paturi and Simon [PS86] (Theorem 2.3.9). This sets our target to showing that there is a function in $\mathsf{THR} \circ \mathsf{THR}$ that has large unbounded-error communication complexity.

Finding a Candidate Function

Why should some function $F \in \mathsf{THR} \circ \mathsf{THR}$ have large unbounded-error communication complexity? A natural protocol one is tempted to use is the following. Sample a sub-circuit of the top gate with a probability proportional to its weight. Then, use the best protocol for the sampled bottom THR gate. Note that for any given input x, with probability $1/2 + \epsilon$, one samples a bottom gate that agrees with the value of $F(x)$. Here, ϵ can be exponentially small in the input size. Thus, if we had a small cost randomized protocol for the bottom THR gate that errs with probability significantly less than ϵ we would have a small cost unbounded-error protocol for F. Fortunately for us (the lower bound prover), there does not seem to exist any such efficient randomized protocol for THR, when $\epsilon = 1/2^{n^{\Omega(1)}}$.

Taking this a step further, one could hope that the bottom gates could be any function that is hard to compute with such tiny error ϵ. The simplest such canonical function is EQ. Therefore, a plausible target is $\mathsf{THR} \circ \mathsf{EQ}$. This still turns out to be in $\mathsf{THR} \circ \mathsf{THR}$ as $\mathsf{EQ} \in \mathsf{ETHR}$. Moreover, EQ has a nice composed structure. It is just $\mathsf{AND} \circ \mathsf{XOR}$, which lets us re-express our target as $F = \mathsf{THR} \circ \mathsf{AND} \circ \mathsf{XOR}$, for some top THR that is 'suitably' hard. At this point, one might be drawn towards the universal threshold function as being a candidate top LTF. However, there is an inherent lack of structure in the universal threshold function, making it difficult to analyze. Thus, we consider a simple LTF, namely OMB. We now view F as an XOR function whose outer function, f turns out to have sufficiently good analytic properties for us to prove that $f \circ \mathsf{XOR}$ has high sign rank.

Analyzing the Candidate Function

We are naturally drawn to the work of Razborov and Sherstov [RS10] for inspiration as they bound the sign rank of a three-level composed function as well. They showed that $\mathsf{AND} \circ \mathsf{OR} \circ \mathsf{AND}_2$ has high sign rank. They exploited the fact that this function embeds a *pattern matrix* inside it, which has nice convenient spectral

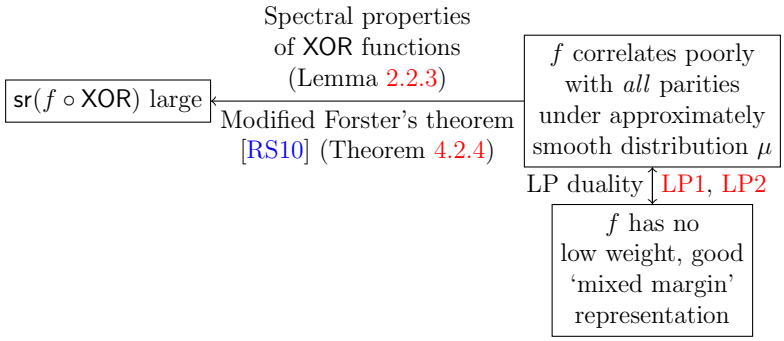

Figure 4.2: Approximation theoretic hardness of f implies large sign rank of $f \circ \mathsf{XOR}$ (Theorem 4.3.1).

properties as observed in [She11a]. These spectral properties dictate them to look for an *approximately smooth orthogonalizing* distribution w.r.t which the outer function $f = \mathsf{AND} \circ \mathsf{OR}$ has zero correlation with small degree parities. This naturally gives rise to a linear program, that seeks to maximize the *smoothness* of the distribution under the constraints of low-degree orthogonality. The main technical challenge that Razborov and Sherstov overcome is the analysis of the dual of this LP using and building appropriate approximation theoretic tools. We follow this general framework of analyzing the dual of a suitable LP. However, as we are forced to work with an XOR function, there are new challenges that crop up. This is understandable, for if we take the same outer function of $\mathsf{AND} \circ \mathsf{OR}$, then the resulting XOR function has small sign rank. Indeed, this remains true even if one were to harden the outer function to $\mathsf{MAJ} \circ \mathsf{OR}$. This is simply because $\mathsf{OR} \circ \mathsf{XOR}$ is non-equality (NEQ). A simple efficient UPP protocol for $\mathsf{MAJ} \circ \mathsf{NEQ}$ exists: pick a random NEQ and then execute a protocol of cost $O(\log n)$ that solves this NEQ with error less than $1/n^2$.

Proof Outline

For the sake of continuity and convenience, we sketch a proof outline along with schematics, as in Section 1.5.1. Figure 4.2 describes a general passage from the problem of showing a lower bound on the sign rank of a function $f \circ \mathsf{XOR}$ to a sufficient problem of proving an approximation theoretic hardness property of f, namely f has no good 'mixed margin' representation by low weight polynomials. Theorem 4.3.1 states the precise connection between the approximation theoretic property of f and the sign rank of $f \circ \mathsf{XOR}$. This passage is made possible by using well known spectral properties of XOR functions and LP duality. This is similar to

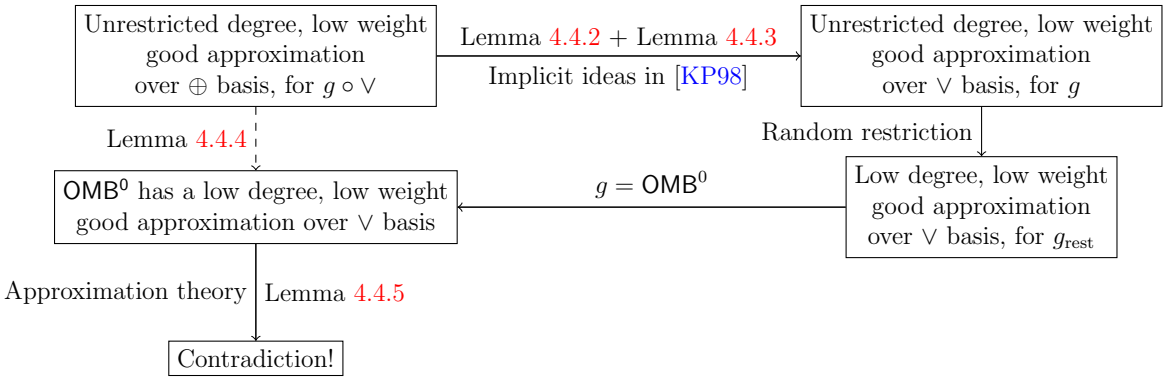

Figure 4.3: Approximation theoretic analysis (Theorem 4.4.1)

earlier works [RS10, She11b, BT16, BCH+16], where the spectral properties of pattern matrices were analyzed. The key difference between our work and theirs is in the nature of the approximation theoretic problem that we end up with. While all these previous works had to rule out good *low degree* representations, our Theorem 4.3.1 stipulates us to rule out good *low weight* representations of otherwise *unrestricted degree*.

Our main technical contribution is Theorem 4.4.1 which shows that the function $\mathsf{OMB}^0 \circ \mathsf{OR}$ is inapproximable by low weight polynomials of *unrestricted degree*, in a sense which we elaborate on below. We prove this by a novel combination of ideas, sketched in Figure 4.3, that differs entirely from analysis in earlier works. One may view this result as a hardness amplification result, albeit specific to the function OMB^0. We start with the function OMB^0 which has no low weight 'worst case margin' representation when the degree of the approximating polynomial is bounded [Bei94]. We show that on composition with large fan-in OR gates, the function $\mathsf{OMB}^0 \circ \mathsf{OR}$ becomes '*mixed margin*'-inapproximable by low weight polynomials, even with *unrestricted degree*. We believe this result to be of independent interest in the area of analysis of Boolean functions and approximation theory.

The first step in our method is to borrow an averaging idea from Krause and Pudlák [KP98] to show the following: a low weight good approximation of $g \circ \mathsf{OR}_m$ by a polynomial p over the parity (Fourier) basis implies that there exists a low weight polynomial q over the OR basis which approximates g as well as p approximates $g \circ \mathsf{OR}_m$, save an additive loss of at most 2^{-m}. This transformation to q is very useful because although it is still unrestricted in degree, it is over the OR basis, that is vulnerable to random restrictions. Indeed, in the next step, we hit q with random restrictions to reduce its degree. At this point, we extract a low weight *and* low

degree polynomial r that still approximates g_{rest}, the restriction of g. We now appeal to interesting properties of the ODD-MAX-BIT function by setting $g = \mathsf{OMB}^0$. First, we observe that OMB^0 on l bits, under random restrictions, retains its hardness as it contains OMB^0 on $l/8$ bits with high probability. Next, we show that OMB^0 does not have low degree good approximations by appealing to classical approximation theoretic tools, suitably modifying the arguments of Buhrman et al. [BVdW07] and Beigel [Bei94]. This provides us with the required contradiction.

4.1.6 Related Work

Long after Forster [For02] showed that an upper bound on the spectral norm of a $\{-1, 1\}$ valued matrix suffices to show sign rank lower bounds, Sherstov [She11b] introduced an innovative method that designed a passage to a suitable approximation problem via LP duality. This basic framework was again used by Razborov and Sherstov [RS10], developing more approximation theoretic tools, to prove the first exponential lower bounds of $2^{\Omega(n^{1/3})}$ on the sign rank of a function in AC^0. This function can be computed by a *depth-3 linear sized circuit*. Later, with a more detailed approximation theoretic analysis, this bound was improved to $2^{\tilde{\Omega}(n^{2/5})}$ by Bun and Thaler [BT16] for a more carefully chosen function, still in depth-3 AC^0. Finally, very recently, Bouland et al. [BCH+16] proved strong sign rank bounds for a *partial function* with interesting applications. All of these works [She11b, RS10, BT16, BCH+16] rely on the passage, invented by Sherstov [She11b] to an approximation theoretic problem involving *low degree* polynomials. This passage is made possible by exploiting the elegant spectral properties of communication matrices of the target functions, following the basic pattern matrix method of Sherstov [She11a].

Unfortunately, it seems difficult to embed a pattern matrix in a function in $\mathsf{THR} \circ \mathsf{THR}$. Consequently, we come up with a different type of function, F_n, that is an *XOR function*. Proving lower bounds on communication complexity of XOR functions, in general, has received a lot of attention recently [MO09, ZS09, LZ10, Zha14, HHL18, KMSY18]. However, there seem to be very few previous works that prove a lower bound on the sign rank of an XOR function. All of these works (Section 4.8 of this thesis, [HQ17, AFK17]) consider the sign rank of functions of the form $f \circ \mathsf{XOR}$ when f is symmetric. In contrast, our target function F_n is not a symmetric XOR function. Moreover, both the works [HQ17] and [AFK17] obtain their result using neat reductions from pattern matrices of symmetric functions, which had been analyzed by Sherstov [She11b]. Such a reduction for a function in $\mathsf{THR} \circ \mathsf{THR}$ is unknown,

and plausibly impossible. This forces us to use a first-principle based argument for bounding the sign rank of an XOR function. Such functions also have nice spectral properties that are however different from those of pattern matrices. More specifically, the approximation theoretic problem that one is led to in this case involves polynomials with *unrestricted* degree but low Fourier weight. A similar flavored but simpler problem was considered in Chapter 3, where we characterized the *discrepancy* of XOR functions. In that chapter, the primal program asked for a distribution μ such that f correlates poorly with all parities w.r.t μ. However, there was no smoothness constraint imposed on μ in, which is what we are constrained to have in this chapter. Analyzing this combination of high degree parity constraints and the smoothness constraints is the main new technical challenge that our work addresses.

It is simple to verify that F_n is computed by a linear size AC^0 circuit. Theorem 4.1.2 therefore yields a new argument to show that AC^0 has large sign rank. While our bounds on the sign rank of F_n are weaker than that of [RS10, BT16], F_n is simpler than the earlier functions in other ways. It is just a decision list of 'Equalities' that is therefore, both in the boolean circuit class $\mathsf{THR} \circ \mathsf{THR}$ and the communication complexity class P^{MA}. It is precisely this property of F_n that allows us to simultaneously answer two open questions.

4.2 Preliminaries

4.2.1 Sign Rank

In a seminal result, Forster [For02] showed that the sign rank of a ± 1 valued matrix is bounded below in terms of the spectral norm of the matrix. This immediately yielded exponential sign rank lower bounds for IP.

Theorem 4.2.1. For any ± 1 valued $m \times n$ matrix M,

$$\mathsf{sr}(M) \geq \frac{\sqrt{mn}}{||M||}.$$

Forster et al. [FKL+01] generalized Forster's result in the following way, which can handle input matrices which are not ± 1 valued, but all of whose entries are large enough in magnitude.

Theorem 4.2.2 (Forster et al. [FKL+01])**.** Let $M_{m\times n}$ be a real matrix with no 0 entries. Then,

$$\mathsf{sr}(M) \geq \frac{\sqrt{mn}}{||M||} \cdot \min_{x,y} |M(x,y)|.$$

Forster et al. also observed that functions with efficient $\mathsf{THR}\circ\mathsf{MAJ}$ representations have small sign rank.

Lemma 4.2.3 (Forster et al. [FKL+01])**.** Let $F : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ be a boolean function computed by a $\mathsf{THR} \circ \mathsf{MAJ}$ circuit of size s. Then,

$$\mathsf{sr}(M_F) \leq sn,$$

where M_F denotes the communication matrix of F.

We also require the following further generalization of Forster's theorem [For02] due to Razborov and Sherstov [RS10].

Theorem 4.2.4 (Razborov and Sherstov [RS10])**.** Let $A = [A_{xy}]_{x\in X, y\in Y}$ be a real valued matrix with $s = |X||Y|$ entries, such that $A \neq 0$. For arbitrary parameters $h, \gamma > 0$, if all but h of the entries of A satisfy $|A_{xy}| \geq \gamma$, then

$$\mathsf{sr}(A) \geq \frac{\gamma s}{||A||\sqrt{s} + \gamma h}.$$

4.2.2 Functions, Polynomials and Approximation

Definition 4.2.5 (OR polynomials)**.** Define a function $p : \{-1,1\}^n \to \mathbb{R}$ of the form $p(x) = \sum_{S\subseteq[n]} a_S \bigvee_{i\in S} x_i$ to be an *OR polynomial*. Define the weight of p (in the OR basis) to be $\sum_{S\subseteq[n]} |a_S|$, and its degree to be $\max_{S\subseteq[n]}\{|S| : a_S \neq 0\}$.

Remark 4.2.6. In the above definition, 'OR monomials' are defined as follows.

$$\bigvee_{i\in S} x_i = \begin{cases} 1 & x_i = 1\ \forall i \in S \\ -1 & \text{otherwise.} \end{cases}$$

Unless mentioned otherwise, all polynomials we consider will be over the *parity* basis.

Hansen and Podolskii [HP10] showed the following.

Theorem 4.2.7 (Hansen and Podolskii [HP10])**.** If a function $f : \{-1,1\}^n \to \{-1,1\}$ can be represented by a $\mathsf{THR}\circ\mathsf{ETHR}$ formula of size s, then it can be represented by a $\mathsf{THR}\circ\mathsf{THR}$ formula of size $2s$.

For the sake of completeness and clarity, we provide the proof below.

Proof. Let h be an exact threshold function with the representation $\sum_{i=1}^{n} w_i x_i = t$. There exists an $\epsilon_h > 0$ such that $\sum_{i=1}^{n} w_i x_i > t \implies \sum_{i=1}^{n} w_i x_i > t + \epsilon_h$. Consider a $\mathsf{THR} \circ \mathsf{ETHR}$ formula of size s which computes f. Say it computes $\mathsf{sgn}(c_0 + \sum_{i=1}^{s} c_i f_i)$, where f_i's have exact threshold representations $\sum_{j=1}^{n} w_{i,j} x_j = t_i$, respectively. Consider the $\mathsf{THR} \circ \mathsf{THR}$ formula of size $2s$, given by $\mathsf{sgn}\left(\sum_{i=1}^{s} c_i (g_{i,1} - g_{i,2} + 1)\right)$, where g_i's are threshold functions with representations as follows.

$$g_{i,1} = 1 \iff \sum_{j=1}^{n} w_{i,j} x_j \geq t_i,$$

$$g_{i,2} = 1 \iff \sum_{j=1}^{n} w_{i,j} x_j \geq t_i + \epsilon_{f_i}.$$

It is easy to verify that this formula computes f. □

Remark 4.2.8. In fact, Hansen and Podolskii [HP10] showed that the circuit class $\mathsf{THR} \circ \mathsf{THR}$ is identical to the circuit class $\mathsf{THR} \circ \mathsf{ETHR}$. However, we do not require the full generality of their result.

We now note that any function computable by a $\mathsf{THR} \circ \mathsf{OR}$ formula can be computed by a $\mathsf{THR} \circ \mathsf{AND}$ formula without a blowup in the size.

Lemma 4.2.9. Suppose $f : \{-1, 1\}^n \to \{-1, 1\}$ can be computed by a $\mathsf{THR} \circ \mathsf{OR}$ formula of size s. Then, f can be computed by a $\mathsf{THR} \circ \mathsf{AND}$ formula of size s.

Proof. Consider a $\mathsf{THR} \circ \mathsf{OR}$ formula of size s, computing f, say

$$f(x) = \mathsf{sgn}\left(\sum_{i=1}^{s} w_i \bigvee_{j \in S_i} x_j\right).$$

Note that

$$\sum_{i=1}^{s} w_i \bigvee_{j \in S_i} x_j = \sum_{i=1}^{s} -w_i \bigwedge_{j \in S_i} x_j^c.$$

Thus, $\mathsf{sgn}\left(\sum_{i=1}^{s} -w_i \bigwedge_{j \in S_i} x_j^c\right)$ is a $\mathsf{THR} \circ \mathsf{AND}$ representation of f, of size s. □

We require the following well-known lemma by Minsky and Papert [MP69].

Lemma 4.2.10 (Minsky and Papert [MP69])**.** Let $p : \{-1, 1\}^n \to \mathbb{R}$ be any symmetric real polynomial of degree d. Then, there exists a univariate polynomial q of

degree at most d, such that for all $x \in \{-1,1\}^n$,

$$p(x) = q(\#1(x))$$

where $\#1(x) = |\{i \in [n] : x_i = 1\}|$.

We require the following approximation-theoretic lemma by Ehlich and Zeller [EZ64] and Rivlin and Cheney [RC66].

Lemma 4.2.11 ([EZ64, RC66])**.** The following holds true for any real valued $\alpha > 0$ and integer $k > 0$. Let $p : \mathbb{R} \to \mathbb{R}$ be a univariate polynomial of degree $d < \sqrt{k/4}$, such that $p(0) \geq \alpha$, and $p(i) \leq 0$ for all $i \in [k]$. Then, there exists $i \in [k]$ such that $p(i) < -2\alpha$.

4.3 Sign Rank to Polynomial Approximation

In this section, we prove how a certain approximation theoretic hardness property of f implies that the sign rank of $f \circ \mathsf{XOR}$ is large, as outlined in Figure 4.2.

Let $f : \{-1,1\}^n \to \{-1,1\}$ be any function, $\delta > 0$ be a parameter, and X be any subset of $\{-1,1\}^n$. We consider the following linear program, which has exactly the same structure as in (LP1) in [She11b] except for one crucial difference described below:

(LP1)

Variables	$\epsilon, \{\mu_x : x \in \{-1,1\}^n\}$		
Minimize	ϵ		
s.t.	$\left\lvert \sum_x \mu(x) f(x) \chi_S(x) \right\rvert$	$\leq \epsilon$	$\forall S \subseteq [n]$
	$\sum_x \mu(x)$	$= 1$	
	$\epsilon \geq 0$		
	$\mu(x) \geq \frac{\delta}{2^n}$		$\forall x \in X$
	$\mu(x) \geq 0$		$\forall x \in \{-1,1\}^n$

The first constraint in (LP1) specifies that correlation of f against *all parities* need to be small w.r.t a distribution μ. Note that in [She11b], this constraint was only imposed for *low degree* parities. This difference between the two linear programs forces us to entirely change the analysis of the dual from the one in [She11b]. As discussed earlier in Section 4.1.5, this analysis is one of our main technical innovations. The second last constraint enforces the fact that μ is 'δ-smooth' over the set X. As we had indicated earlier in Section 4.1.5, these constraints make analyzing the LP challenging.

Standard manipulations (as in Section 3.4, for example) and strong linear programming duality reveal that the optimum of (LP1) equals the optimum of (LP2). Let OPT denote the optima of these programs.

(LP2)

Variables	$\Delta, \{\alpha_S : S \subseteq [n]\}, \{\xi_x : x \in X\}$		
Maximize	$\Delta + \frac{\delta}{2^n} \sum_{x \in X} \xi_x$		
s.t.	$f(x) \sum_{S \subseteq [n]} \alpha_S \chi_S(x)$	$\geq \Delta$	$\forall x \in \{-1,1\}^n$
	$f(x) \sum_{S \subseteq [n]} \alpha_S \chi_S(x)$	$\geq \Delta + \xi_x$	$\forall x \in X$
	$\sum_{S \subseteq [n]} \lvert\alpha_S\rvert$	≤ 1	
	$\Delta \in \mathbb{R}$		
	$\alpha_S \in \mathbb{R}$		$\forall S \subseteq [n]$
	$\xi_x \geq 0$		$\forall x \in X$

The first constraint of (LP2) indicates that the variable Δ represents the worst margin guaranteed to exist at all points. The second constraint says that at each point x over the smooth set X, the dual polynomial has to better the worst margin by at least ξ_x. If OPT is large, then it means that on average, the dual polynomial did significantly better than the worst margin. It is for this reason we call the optimum the 'mixed margin' as mentioned in Section 4.1.5.

We now show that an upper bound on OPT for any function f yields sign rank lower bounds against $f \circ \mathsf{XOR}$. The proof idea is depicted in Figure 4.2.

Theorem 4.3.1. Let $f : \{-1,1\}^n \to \mathbb{R}$ be any function. For any $\delta > 0$ and $X \subseteq \{-1,1\}^n$, suppose the value of the optimum of (LP2) (and hence (LP1)) is at most OPT. Then,

$$\mathsf{sr}(f \circ \mathsf{XOR}) \geq \frac{\delta}{OPT + \delta \cdot \frac{|X^c|}{2^n}}.$$

Proof. By (LP1), there exists a distribution μ on $\{-1,1\}^n$ such that $\mu(x) \geq \frac{\delta}{2^n}$ for all $x \in X$, and $\max_{S \subseteq [n]} |\widehat{f\mu}(S)| \leq \frac{OPT}{2^n}$. By Lemma 2.2.3,

$$||M_{f\mu \circ \mathsf{XOR}}|| = 2^n \cdot \max_{S \subseteq [n]} \left|\widehat{f\mu}(S)\right| \leq OPT.$$

Each $x \in X$ contributes to 2^n entries of $M_{f\mu \circ \mathsf{XOR}}$ whose absolute value is at least δ. Plugging values in Theorem 4.2.4, we obtain

$$\mathsf{sr}(f \circ \mathsf{XOR}) \geq \mathsf{sr}(f\mu \circ \mathsf{XOR}) \geq \frac{\frac{\delta}{2^n} \cdot 2^{2n}}{OPT \cdot 2^n + \frac{\delta}{2^n} \cdot 2^n \cdot |X^c|} = \frac{\delta}{OPT + \delta \cdot \frac{|X^c|}{2^n}},$$

which proves the desired sign rank lower bound.

□

4.4 Hardness of Approximating $\mathsf{OMB}^0_\ell \circ \mathsf{OR}_m$

Below is our main technical result, capturing the essence of Figure 4.3, which says that no dual polynomial exists with a large optimum value for (LP2) when $f = \mathsf{OMB}^0_\ell \circ \bigvee_{\ell^{1/3}+\log \ell} : \{-1,1\}^{\ell^{4/3}+\ell \log \ell} \to \{-1,1\}$, even when the smoothness parameter δ is as high as $1/4$.

Theorem 4.4.1. Let $f = \mathsf{OMB}^0_\ell \circ \bigvee_{\ell^{1/3}+\log \ell} : \{-1,1\}^{\ell^{4/3}+\ell \log \ell} \to \{-1,1\}, \delta = 1/4$ and $X = \{x \in \{-1,1\}^{\ell^{4/3}+\ell \log \ell} : \bigvee(x) = -1^\ell\}$. Then for sufficiently large values of ℓ, the optimal value, OPT, of (LP2) is less than $2^{-\frac{\ell^{1/3}}{81}}$.

Theorem 4.4.1 can be viewed as a hardness amplification theorem as follows. Our base function is OMB^0, which is known to be hard to approximate in the worst case by low degree sign representing polynomials [Bei94, BVdW07]. We show that a lifted version of this function, $\mathsf{OMB}^0_\ell \circ \mathsf{OR}_m$, cannot be approximated well under a significantly weaker notion of approximation where we permit any approximating polynomial to have the following additional power.

- Unrestricted degree but low weight.
- It need not sign represent $\mathsf{OMB}^0_\ell \circ \mathsf{OR}_m$, but a certain linear combination of their worst case and average case margin is small (see (LP2)). In fact, it might agree in sign with $\mathsf{OMB}^0_\ell \circ \mathsf{OR}_m$ on just one input!

In the remnant of this section, we list the various tools that go into proving Theorem 4.4.1. We follow the schematic from Figure 4.3.

We first use an idea from Krause and Pudlák [KP98] which enables us to work with polynomial approximations for g, given a polynomial approximation for $g \circ \bigvee_m$. We use the following notation for the following two lemmas. For any set $I \subseteq [\ell] \times [m]$, define $J \subseteq [\ell]$ to be the projection of I on $[\ell]$; $i \in J \iff \exists j,\ x_{i,j} \in I$. For any

$y \in \{-1,1\}^\ell$, let μ_y denote the uniform distribution over all inputs $x \in \{-1,1\}^{m\ell}$ such that $\bigvee_m(x) = y$. Lemma 4.4.2 and Lemma 4.4.3 represent the first implication in Figure 4.3. The first tool we use is an approximation of monomials (in the parity basis) by OR functions, with a small error.

Lemma 4.4.2. Let ℓ, m be positive integers such that $m > \log \ell$. For any set $I \subseteq [\ell] \times [m], y \in \{-1,1\}^\ell$,

$$\left| \mathbb{E}_{\mu_y} \left[\bigoplus_{(i,j)\in I} x_{i,j} \right] - \frac{1}{2} - \frac{1}{2} \bigvee_{i \in J} y_i \right| \leq 2\ell 2^{-m}.$$

The proof of Lemma 4.4.2 appears in the proof of Lemma 2.3 in [KP98]. However, we reproduce the proof below for clarity and completeness.

Proof of Lemma 4.4.2. First observe that for all $y \in \{-1,1\}^\ell$, and for all x satisfying $\bigvee_m(x) = y$, the monomial corresponding to I equals

$$\bigoplus_{(i,j)\in I} x_{i,j} = \bigoplus_{(i,j)\in I, y_i = -1} x_{i,j}.$$

Let $A = \{i \in [\ell] : y_i = -1\}$. If $A \cap J = \emptyset$, then

$$\mathbb{E}_{\mu_y} \left[\bigoplus_{(i,j)\in I} x_{i,j} \right] = \bigvee_{i \in J} y_i = 1.$$

Else, $\bigvee_{i\in J} y_i = -1$. Also,

$$\mathbb{E}_{\mu_y} \left[\bigoplus_{(i,j)\in I} x_{i,j} \right] = \mathbb{E}_{x \in \{-1,1\}^{(A\cap J)\times[m]} : \bigvee(x) = -1^{|A\cap J|}} \left[\bigoplus_{(i,j)\in I, y_i=-1} x_{i,j} \right]. \tag{4.1}$$

Note that

$$\mathbb{E}_{x\in\{-1,1\}^{(A\cap J)\times[m]}} \left[\bigoplus_{(i,j)\in I, y_i=-1} x_{i,j} \right] = 0. \tag{4.2}$$

Denote $|A \cap J| = t$. Using Equation (4.2) and a simple counting argument, the absolute value of the RHS (and thus the LHS) of Equation (4.1) can be bounded

above as follows (note that we require $1 \leq t \leq \ell$ in the following computations).

$$\begin{aligned}\left|\mathbb{E}_{\mu_y}\left[\bigoplus_{(i,j)\in I} x_{i,j}\right]\right| &\leq \frac{2^{mt} - (2^m - 1)^t}{(2^m - 1)^t} \\ &\leq \frac{2^{mt} - (2^{mt} - t2^{m(t-1)})}{(2^m - 1)^t}\end{aligned}$$

(Sum of remaining terms in binomial expansion of $(2^m - 1)^t$ is positive since $m > \log \ell$)

$$\begin{aligned}&\leq \frac{t \cdot 2^{mt-m}}{2^{mt}/2} && \text{(since } m > \log \ell\text{)} \\ &\leq 2\ell 2^{-m}.\end{aligned}$$

Hence, for all $y \in \{-1,1\}^\ell$, we have

$$\left|\mathbb{E}_{\mu_y}\left[\bigoplus_{(i,j)\in I} x_{i,j}\right] - \frac{1}{2} - \frac{1}{2}\bigvee_{i\in J} y_i\right| \leq 2\ell 2^{-m}. \tag{4.3}$$

□

The next lemma states that g can be approximated well over the OR basis, given a good approximation for $g \circ \bigvee$ over the parity basis.

Lemma 4.4.3. Let ℓ, m be positive integers such that $m > \log \ell$, and $g : \{-1,1\}^\ell \to \{-1,1\}$ be any function. Define $f = g \circ \bigvee_m : \{-1,1\}^{m\ell} \to \{-1,1\}, \Delta \in \mathbb{R}, e_x \geq 0\ \forall x \in X$, where X denotes the set of all inputs x in $\{-1,1\}^{m\ell}$ such that $\bigvee_m(x) = -1^\ell$, and let p be a real polynomial such that

$$\begin{aligned}\forall x \in \{-1,1\}^{m\ell}, &\qquad f(x)p(x) \geq \Delta, \\ \forall x \in X, &\qquad f(x)p(x) \geq \Delta + e_x.\end{aligned}$$

Then, there exists an OR polynomial q, of weight at most $\mathsf{wt}(p)$, such that

$$\begin{aligned}\forall y \in \{-1,1\}^\ell, &\qquad q(y)g(y) \geq \Delta - \mathsf{wt}(p)\left(2\ell \cdot 2^{-m}\right), \\ &\qquad q(-1^\ell)g(-1^\ell) \geq \Delta + \frac{\sum_{x\in X} e_x}{|X|} - \mathsf{wt}(p)\left(2\ell \cdot 2^{-m}\right).\end{aligned}$$

Proof. Note that for any $y \in \{-1,1\}^\ell$,

$$\mathbb{E}_{\mu_y}[f(x)p(x)] = g(y) \cdot \mathbb{E}_{\mu_y}[p(x)] \geq \Delta \tag{4.4}$$

and

$$\mathbb{E}_{\mu_{-1^\ell}}[f(x)p(x)] = g(-1^\ell) \cdot \mathbb{E}_{\mu_{-1^\ell}}[p(x)] \geq \Delta + \frac{\sum_{x \in X} e_x}{|X|}. \tag{4.5}$$

Denote the unique multilinear expansion of p by $p = v_0 + \sum_k v_k p_k$, where $p_k(x) = \oplus_{(i,j) \in I_k} x_{i,j}$. Let J_k denote the projection of I_k on $[\ell]$. Define

$$q = v_0 - \frac{\sum_k v_k}{2} - \sum_k \frac{v_k}{2} \bigvee_{i \in J_k} y_i.$$

Note that

$$\mathsf{wt}(q) = \mathsf{wt}\left(v_0 - \frac{\sum_k v_k}{2} - \sum_k \frac{v_k}{2} \bigvee_{i \in J_k} y_i\right) = \left|v_0 - \frac{\sum_k v_k}{2}\right| + \sum_k \left|\frac{v_k}{2}\right| \leq \mathsf{wt}(p).$$

Thus, using linearity of expectation and Lemma 4.4.2, Equation (4.4) and Equation (4.5) yield that for all $y \in \{-1,1\}^\ell$,

$$q(y) \cdot g(y) \geq \Delta - \mathsf{wt}(p)\left(2\ell \cdot 2^{-m}\right)$$

and

$$q(-1^\ell) \cdot g(-1^\ell) \geq \Delta + \frac{\sum_{x \in X} e_x}{|X|} - \mathsf{wt}(p)\left(2\ell \cdot 2^{-m}\right).$$

□

Next, we use random restrictions which reduces the degree of the approximating OR polynomial, at the cost of a small error. In particular, we consider the case when $g = \mathsf{OMB}^0_\ell$. This represents the dashed implication in Figure 4.3.

Lemma 4.4.4. Let ℓ, m be any positive integers such that $m > \log \ell$. Let $g_\ell = \mathsf{OMB}^0_\ell : \{-1,1\}^\ell \to \{-1,1\}, f = g_\ell \circ \bigvee_m$, and $\Delta, \{e_x \geq 0 : x \in X\}$ (where X is defined as in Lemma 4.4.3), and p be a real polynomial such that

$$\forall x \in \{-1,1\}^{m\ell},\ f(x)p(x) \geq \Delta,$$
$$\forall x \in X,\ p(x) \geq \Delta + e_x.$$

Then, for any integer $d > 0$, there exists an OR polynomial $r : \{-1,1\}^{\ell/8} \to \mathbb{R}$, of degree d and weight at most $\mathsf{wt}(p)$, such that

$$\begin{aligned} &\text{For all } y \in \{-1,1\}^{\ell/8}, \qquad r(y)g_{\ell/8}(y) \geq \Delta - \mathsf{wt}(p)\left(2\ell \cdot 2^{-m} + 2^{-(d-1)}\right) \\ &\qquad\qquad \text{and} \qquad\qquad r(-1^{\ell/8}) \geq \Delta + \frac{\sum_{x\in X} e_x}{|X|} - \mathsf{wt}(p)\left(2\ell \cdot 2^{-m} + 2^{-(d-1)}\right). \end{aligned}$$

Proof. Lemma 4.4.3 guarantees the existence of an OR polynomial q, of weight at most $\mathsf{wt}(p)$, such that

$$\begin{aligned} &\forall y \in \{-1,1\}^{\ell}, \qquad q(y)g_{\ell}(y) \geq \Delta - \mathsf{wt}(p)\left(2\ell \cdot 2^{-m}\right) \\ &\qquad\qquad \text{and} \qquad\qquad q(-1^{\ell})g(-1^{\ell}) \geq \Delta + \frac{\sum_{x\in X} e_x}{|X|} - \mathsf{wt}(p)\left(2\ell \cdot 2^{-m}\right). \end{aligned} \tag{4.6}$$

Now, set each of the ℓ variables to -1 with probability $1/2$, and leave it unset with probability $1/2$. Call this random restriction R. Any OR monomial of degree at least d gets fixed to -1 with probability $1 - 2^{-d}$. Thus, by linearity of expectation, the expected weight of surviving monomials of degree at least d in q is at most $\mathsf{wt}(p) \cdot 2^{-d}$. Let $M|_R$ denote the value of a monomial M after the restriction R. By Markov's inequality,

$$\Pr_R\left[\sum_{M:\mathsf{deg}(M|_R)\geq d} \mathsf{wt}(M|_R) > \mathsf{wt}(p) \cdot 2^{-d+1}\right] < 1/2.$$

Consider $\ell/2$ pairs of variables, $\{(x_i, x_{i+1}) : i \in [\ell/2]\}$ (assume w.l.o.g that ℓ is even). For any pair, the probability that both of its variables remain unset is $1/4$. This probability is independent over pairs. Thus, by a Chernoff bound, the probability that at most $\ell/16$ pairs remain unset is at most $2^{-\frac{\ell}{64}}$.

By a union bound, there exists a setting of variables such that at least $\ell/16$ pairs of variables are left free, and the weight of degree $\geq d$ monomials in q is at most $\mathsf{wt}(p) \cdot 2^{-d+1}$. Set the remaining $7\ell/8$ variables to the value -1. After the restriction, drop the monomials of degree $\geq d$ from q to obtain r, which is now an OR polynomial of degree less than d and weight at most $\mathsf{wt}(p)$. Note that the function g_ℓ hit with this restriction just becomes $g_{\ell/8}$.

Thus, Equation (4.6) yields the following.

$$\begin{aligned} &\text{For all } y \in \{-1,1\}^{\ell/8}, \qquad r(y)g_{\ell/8}(y) \geq \Delta - \mathsf{wt}(p)\left(2\ell \cdot 2^{-m} + 2^{-(d-1)}\right) \\ &\qquad\qquad \text{and} \qquad\qquad r(-1^{\ell/8}) \geq \Delta + \frac{\sum_{x\in X} e_x}{|X|} - \mathsf{wt}(p)\left(2\ell \cdot 2^{-m} + 2^{-(d-1)}\right). \end{aligned}$$

□

The following lemma states that approximating OMB^0 well by a low weight polynomial p is not possible unless the degree of p is large. This captures the last implication in Figure 4.3.

Lemma 4.4.5. Suppose $p : \{-1,1\}^n \to \mathbb{R}$ is a polynomial of degree $d < \sqrt{n/4}$ and $a > 0, b \in \mathbb{R}$ be reals such that $p(-1^n) \geq a$ and $\mathsf{OMB}^0_n(x)p(x) \geq b$ for all $x \in \{-1,1\}^n$. Define

$$p_{max} = \max_{i \in \{0,\ldots,\lfloor n/10d^2 \rfloor\}} \{2^i a + \left(3 \cdot 2^i - 3\right) b\}.$$

Then, there exists $x \in \{-1,1\}^n$ such that $|p(x)| \geq p_{max}$.

A simple consequence of the above lemma is that the weight of a polynomial p (in either the OR basis, or the parity basis) satisfying the assumptions of Lemma 4.4.5 is at least p_{max}. This property of p suffices for our need.

The proof of Lemma 4.4.5 follows an iterative argument, making repeated use of Lemma 4.2.11, inspired by the arguments of Beigel [Bei94] and Buhrman et al. [BVdW07].

Remark 4.4.6. We remark here that this strengthens the result of Beigel [Bei94], who proved that any good approximation by a low degree *sign representing* polynomial for OMB^0 must have large weight. Our approximating polynomial is not constrained to be sign representing (b might be negative in Lemma 4.4.5). In fact, it might disagree in sign with OMB^0 on all points but -1^n.

We first require the following intermediate claim.

Claim 4.4.7. If a and b are reals such that $a > 0, b \in \mathbb{R}$ and $2^i a + (3 \cdot 2^i - 2)\, b < 0$ for some integer $i \geq 0$, then $2^j a + (3 \cdot 2^j - 3)\, b < 0$ for all integers $j > i$.

Proof. Note that since $a > 0$ and $2^i a + (3 \cdot 2^i - 2)\, b < 0$, b must be negative. For any $j > i$, write $2^j a + (3 \cdot 2^j - 3)\, b = 2^{j-i}\left(2^i a + (3 \cdot 2^i - 2)\, b\right) + (2^{j-i+1} - 3)b < 0$. □

Proof of Lemma 4.4.5. Divide the n variables into $\lfloor n/10d^2 \rfloor$ contiguous blocks of size $10d^2$ each.

Induction hypothesis: For each $i \in \{0,\ldots,\lfloor n/10d^2 \rfloor\}$, there exists an input $x^i \in \{-1,1\}^n$ such that

- $x^i_j = -1$ for all indices j to the right of the ith block (thus, $x^0 = (-1)^n$).

- The values of x^i_j for indices j to the left of the ith block are set as dictated by the previous step. That is, $x^i_j = x^{i-1}_j$ for all indices j to the left of the ith block.
- $|p(x^i)| \geq 2^i a + (3 \cdot 2^i - 3)\, b$.
- The value of $p(x^i)$ is negative if i is odd, and positive if i is even.

Clearly, proving this hypothesis proves Lemma 4.4.5. We now prove the induction hypothesis.

- **Base case:** Say $i = 0$. By assumption, $p(-1^n) \geq a$.
- **Inductive step:** Say the hypothesis is true for all $0 \leq j \leq i-1$ for some $i \geq 1$. In the ith block, set the variables corresponding to the even indices to -1 if i is odd, and set the odd indexed variables to -1 if i is even. Set the variables outside the ith block as dictated by the previous step. Assume that i is odd (the argument for even i follows in a similar fashion, with suitable sign changes). Denote the free variables by $y_1, \ldots, y_{5d^2}$. Define a polynomial $p_i : \{-1,1\}^{5d^2} \to \mathbb{R}$ by $p_i(y) = \mathbb{E}_{\sigma \in S_{5d^2}} \tilde{p}(\sigma(y))$, where $\tilde{p}(y)$ denotes the value of p on input $y_1, \ldots, y_{5d^2}$, and the remaining variables are set as described earlier. The expectation is over the uniform distribution. Note that p_i is a symmetric polynomial of degree at most d, and satisfies

$$p_i(-1^{5d^2}) \geq 2^{i-1}a + \left(3 \cdot 2^{i-1} - 3\right) b, \qquad p_i(y) \leq -b \;\forall y \neq -1^{5d^2}.$$

By Lemma 4.2.10, there exists a univariate polynomial p'_i such that for all $j \in \{0\} \cup [5d^2]$,

$$p'_i(j) = p_i(y) \;\forall y \text{ such that } \#1(y) = j.$$

Thus,

$$p'_i(0) \geq 2^{i-1}a + \left(3 \cdot 2^{i-1} - 3\right) b, \qquad p'_1(j) \leq -b \;\forall j \in [5d^2].$$

Define $p''_i = p'_i + b$. Thus,

$$p''_i(0) \geq 2^{i-1}a + \left(3 \cdot 2^{i-1} - 2\right) b \qquad p''_i(j) \leq 0 \;\forall j \in [5d^2].$$

If $2^{i-1}a + (3 \cdot 2^{i-1} - 2)\, b < 0$, then by Claim 4.4.7, the inductive hypothesis is true for all integers $j \geq i$. Thus, assume $2^{i-1}a + (3 \cdot 2^{i-1} - 2)\, b \geq 0$.

By Lemma 4.2.11, there exists a $j \in [5d^2]$ such that $p''_i(j) \leq -2^i a - (3 \cdot 2^i - 4)\, b$, and hence $p'_i(j) \leq -2^i a - (3 \cdot 2^i - 3)\, b$. This implies the existence of an x^i in

$\{-1,1\}^n$ (with all variables to the right of the ith block still set to -1, and variables to the left of the ith block as dictated by the previous step) such that $p(x^i) < -2^i a - (3 \cdot 2^i - 3)\, b$.

□

We are now ready to prove Theorem 4.4.1.

Proof of Theorem 4.4.1. Let p be a polynomial of weight 1, for which (LP2) attains its optimum. Denote the values taken by the variables at the optimum by $\Delta_{\mathsf{OPT}}, \{\xi_{x,\mathsf{OPT}} : x \in X\}$. Towards a contradiction, assume $\mathsf{OPT} \geq 2^{-\frac{\ell^{1/3}}{81}}$.

Lemma 4.4.4 (set $m = \ell^{1/3} + \log \ell$) shows the existence of an OR polynomial r, on $\ell/8$ variables, of degree $\ell^{1/3}$ and weight 1, such that

$$\text{For all } y \in \{-1,1\}^{\ell/8}, \qquad r(y)\mathsf{OMB}^0_{\ell/8}(y) \geq \Delta_{\mathsf{OPT}} - 2 \cdot 2^{-\ell^{1/3}} - 2 \cdot 2^{-\ell^{1/3}}$$

$$\text{and} \qquad r(-1^{\ell/8}) \geq \Delta + \frac{\sum_{x \in X} \xi_{x,\mathsf{OPT}}}{|X|} - 2 \cdot 2^{-\ell^{1/3}} - 2 \cdot 2^{-\ell^{1/3}}.$$

Note that

$$\mathsf{OPT} \geq 2^{-\frac{\ell^{1/3}}{81}} \implies \Delta_{\mathsf{OPT}} \geq 2^{-\frac{\ell^{1/3}}{81}} - \delta \frac{\sum_{x \in X} \xi_{x,\mathsf{OPT}}}{2^n}. \tag{4.7}$$

r satisfies the assumptions of Lemma 4.4.5 with $d = \mathsf{deg}(r) = \ell^{1/3} < \sqrt{\ell/32}$ (since any OR polynomial of degree d can be represented by a polynomial of degree at most d), $a = \Delta_{\mathsf{OPT}} + \frac{\sum_{x \in X} \xi_{x,\mathsf{OPT}}}{|X|} - 4 \cdot 2^{-\ell^{1/3}}$, and $b = \Delta_{\mathsf{OPT}} - 4 \cdot 2^{-\ell^{1/3}}$. a is non-negative because of the following.

$$a = \Delta_{\mathsf{OPT}} + \frac{\sum_{x \in X} \xi_{x,\mathsf{OPT}}}{|X|} - 4 \cdot 2^{-\ell^{1/3}} \geq 2^{-\frac{\ell^{1/3}}{81}} - 4 \cdot 2^{-\ell^{1/3}} \geq 0.$$

Set $k = \ell^{1/3}/80$ for the remaining of this proof. By Lemma 4.4.5, there exists an $x \in \{-1,1\}^{\ell/8}$ such that

$$\begin{aligned}
|r(x)| &\geq 2^k a + \left(3 \cdot 2^k - 3\right) b \geq \Delta_{\mathsf{OPT}}(4 \cdot 2^k - 3) + 2^k \frac{\sum_{x \in X} \xi_{x,\mathsf{OPT}}}{|X|} - 4 \cdot 2^{-80k}(4 \cdot 2^k - 3) \\
&\geq \left(4 \cdot 2^k - 3\right) \left(2^{-\frac{\ell^{1/3}}{81}} - \delta \frac{\sum_{x \in X} \xi_{x,\mathsf{OPT}}}{2^n}\right) + 2^k \frac{\sum_{x \in X} \xi_{x,\mathsf{OPT}}}{|X|} - 4 \cdot 2^{-80k}(4 \cdot 2^k - 3) \\
&\qquad \text{(Using Equation 4.7)} \\
&\geq \left(4 \cdot 2^k - 3\right) \left(2^{-80k/81} - 4 \cdot 2^{-80k}\right) > 1. \\
&\qquad \text{(Since } \delta = 1/4, \text{ and assuming } k \geq 1)
\end{aligned}$$

This yields a contradiction, since r was a polynomial of weight at most 1 (in the OR basis). □

4.5 Class Separations

We are now ready to prove our main theorem and its applications to complexity class separations.

Theorem 4.5.1 (Restatement of Theorem 4.1.2). Let $f = \mathsf{OMB}^0_\ell \circ \bigvee_{\ell^{1/3}+\log \ell} : \{-1,1\}^{\ell^{4/3}+\ell \log \ell} \to \{-1,1\}$. Then, for sufficiently large values of ℓ,

$$\mathsf{sr}(f \circ \mathsf{XOR}) \geq 2^{\frac{\ell^{1/3}}{81}-3}.$$

Proof. Let $n = \ell^{4/3} + \ell \log \ell$. Theorem 4.4.1 says that the optimum of (LP2) (and hence (LP1), by duality) is at most $2^{-\frac{\ell^{1/3}}{81}}$, when $f = \mathsf{OMB}^0_\ell \circ \bigvee_{\ell^{1/3}+\log \ell}, \delta = 1/4$, and $X = \{x \in \{-1,1\}^{\ell^{4/3}+\ell \log \ell} : \bigvee(x) = -1^\ell\}$. We now estimate the size of X^c. The probability (over the uniform distribution on the inputs) of a particular OR gate firing a 1 is $\frac{1}{2^{\ell^{1/3}+\log \ell}}$. By a union bound, the probability of any OR gate firing a 1 is at most $\frac{1}{2^{\ell^{1/3}}}$, hence $|X^c| \leq 2^n \cdot \frac{1}{2^{\ell^{1/3}}}$. Plugging these values in Theorem 4.3.1, we obtain

$$\mathsf{sr}(f \circ \mathsf{XOR}) \geq \frac{1/4}{2^{-\frac{\ell^{1/3}}{81}} + 2^{-\ell^{1/3}-2}} \geq 2^{\frac{\ell^{1/3}}{81}-3}.$$

□

4.5.1 A Separation of Depth-2 Threshold Circuit Classes

We are now ready to prove Theorem 4.1.3, which uses F_n to separate the circuit classes $\mathsf{THR} \circ \mathsf{MAJ}$ and $\mathsf{THR} \circ \mathsf{THR}$, resolving an open question posed in [AM05, HP10].

Proof of Theorem 4.1.3. First, we show that F_n is computable by linear sized $\mathsf{THR} \circ \mathsf{THR}$ formulas. Let $n = 2\ell^{4/3} + 2\ell \log \ell$ denote the number of input bits to $F_n = \mathsf{OMB}^0_\ell \circ \bigvee_{\ell^{1/3}+\log \ell} \circ \mathsf{XOR}_2$. By Lemma 4.2.9, F_n can be computed by a $\mathsf{THR} \circ \mathsf{AND} \circ \mathsf{XOR}_2$ formula of size $2\ell^{4/3} + 2\ell \log \ell$. Hence $F_n \in \mathsf{THR} \circ \mathsf{ETHR} = \mathsf{THR} \circ \mathsf{THR}$, by Theorem 4.2.7.

Next, we show a lower bound on the size of any $\mathsf{THR} \circ \mathsf{MAJ}$ circuit computing F_n. Suppose $\mathsf{OMB}^0_l \circ \bigvee_{l^{1/3}+\log l} \circ \mathsf{XOR}_2$ could be represented by a $\mathsf{THR} \circ \mathsf{MAJ}$ circuit of size

s. By Lemma 4.2.3 and Theorem 4.5.1,

$$s\left(2\ell^{4/3} + 2\ell\log\ell\right) \geq \mathsf{sr}(f) \geq 2^{\frac{\ell^{1/3}}{81}-3}.$$

Thus, $s = 2^{\Omega\left(n^{1/4}\right)}$.

□

4.5.2 Communication Complexity Class Separations

In this section, we show explicit separations between certain communication complexity classes, resolving an open question posed in [GPW18]. This application of our main result was brought to our attention by Göös [Göö17]. Precise definitions of communication complexity classes of interest may be found in Chapter 2.

Theorem 4.5.2. Let $f = \mathsf{OMB}^0_\ell \circ \bigvee_{\ell^{1/3}+\log\ell} : \{-1,1\}^{\ell^{4/3}+\ell\log\ell} \to \{-1,1\}$, and let $n = \ell^{4/3} + \ell\log\ell$ denote the number of input variables. Then, for sufficiently large values of n,

$$\mathsf{UPP}(f \circ \mathsf{XOR}) = \Omega\left(n^{1/4}\right).$$

Proof. It follows from Theorem 4.5.1 and Theorem 2.3.9. □

Note that $F_n = \mathsf{OMB}_\ell \circ \mathsf{EQ}_{\ell^{1/3}+\log\ell}$, where OMB_ℓ outputs -1 iff the rightmost bit of the input set to -1 occurs at an odd index.

It is not hard to see that there is an MA protocol for $\bigvee_\ell \circ \mathsf{EQ}_{\ell^{1/3}+\log\ell}$ of cost polylogarithmic in ℓ. Using this, and a binary search, we exhibit a P^{MA} upper bound for F_n under the natural partition of the inputs.

Protocol 1 P^{MA} protocol for $\mathsf{OMB}(\mathsf{EQ}_1, \ldots, \mathsf{EQ}_\ell)$

if $\bigvee_{i=1}^{\ell}(\mathsf{EQ}_i) = 1$ **then** Output 1.
end if
$start = 1$
$end = \ell$
$mid = \lceil \frac{start+end}{2} \rceil$
while $start \neq end$ **do**
  if $\bigvee_{i=mid}^{end}(\mathsf{EQ}_i) = -1$ **then** $start \leftarrow mid$
  else if $\bigvee_{i=mid}^{end}(\mathsf{EQ}_i) = 1$ **then** $end \leftarrow mid - 1$
  end if
end while
Output -1 iff $start$ is odd.

Hence, we obtain $F_n \in \mathsf{P}^{\mathsf{MA}}$. Along with Theorem 4.5.2, this yields the following result.

Theorem 4.5.3.

$$\mathsf{P}^{\mathsf{MA}} \not\subseteq \mathsf{UPP}.$$

It is known that $\mathsf{P}^{\mathsf{MA}} \subseteq \mathsf{S}_2\mathsf{P}$, and $\mathsf{P}^{\mathsf{MA}} \subseteq \mathsf{BPP}^{\mathsf{NP}}$ (see e.g. [GPW18] for references for such containments, and an excellent overview on the landscape of two-party communication complexity classes).

Thus, Theorem 4.5.3 yields

$$\mathsf{S}_2\mathsf{P} \not\subseteq \mathsf{UPP} \qquad \text{and} \qquad \mathsf{BPP}^{\mathsf{NP}} \not\subseteq \mathsf{UPP}.$$

The first non-inclusion resolves an open question posed in [GPW18]. To the best of our knowledge, ours is the first explicit total function to witness the second non-inclusion. We remark here that Bouland et al. [BCH+16] used a partial function to witness the same separation.

A natural question to ask is whether the class P^{MA} becomes weaker if we restrict its power. On the one hand it is known that $\mathsf{MA} \subsetneq \mathsf{PP} \subsetneq \mathsf{UPP}$ and on the other hand, $\mathsf{P}^{\mathsf{NP}} \subsetneq \mathsf{UPP}$. Thus, Theorem 4.5.3 places the communication complexity class P^{MA} right at the frontier of our current knowledge.

4.6 An Upper Bound

In this section, we observe that the function F_n has sign rank $2^{O(n^{1/4})}$, showing that our lower bound in Theorem 4.1.2 is essentially tight for F_n.

Theorem 4.6.1. The function F_n has sign rank $2^{O(n^{1/4})}$.

Proof. As noted in the previous section, F_n is expressible as a circuit of the form $\mathsf{THR}_\ell \circ \mathsf{EQ}_{\ell^{1/3}+\log \ell}$. Claim 2.3.10 shows that $\mathsf{UPP}(F_n) = O(n^{1/4})$ (since each Equality is trivially solvable by deterministic protocols of cost $\ell^{1/3} + \log \ell$). By Theorem 2.3.9, F_n has sign rank $2^{O(n^{1/4})}$. □

4.7 Signed Monomial Complexity Lower Bounds

In this section, we show how an upper bound on the optimum of LP1 (and LP2) w.r.t a function f yields signed monomial complexity lower bounds for f. This is already implied by Theorem 4.3.1, as a sign rank lower bound on $f \circ \mathsf{XOR}$ directly implies a signed monomial complexity lower bound on f (a simple consequence of Claim 2.3.10). The use of Theorem 4.3.1, whose proof makes use of the deep result of Forster [For02], seems an overkill to just prove a lower bound on signed monomial complexity. In this section, we give a much more direct proof of this fact, entirely avoiding the use of Forster's theorem. In the process, we generalize a classical result of Bruck [Bru90] that gives a sufficient condition for showing lower bounds on signed monomial complexity. The interested reader may note that our generalization of Bruck's theorem is analogous to Razborov and Sherstov's [RS10] generalization of Forster's theorem. Further, our generalized result, Theorem 4.7.2, along with Theorem 4.4.1, will directly imply that there are functions in poly-size $\mathsf{THR} \circ \mathsf{OR}$ circuits that cannot be computed in sub-exponential size by $\mathsf{THR} \circ \mathsf{XOR}$ circuits. Such a result was first proved by Krause and Pudlák [KP98], using a different technique. Interestingly, Krause and Pudlák expressed the belief that such a separation cannot be done based on a spectral technique like that of Bruck's Theorem [Bru90]. Our argument here shows that this belief was false.

We recall Bruck's Theorem below.

Theorem 4.7.1 ([Bru90])**.** Let $f : \{-1,1\}^n \to \{-1,1\}$ be any function. If $\max_{S \subseteq [n]} \left|\hat{f}(S)\right| \leq \epsilon$, then

$$\mathsf{mon}_{\pm}(f) \geq \frac{1}{\epsilon}.$$

The following is our generalization of Theorem 4.7.1.

Theorem 4.7.2. Let $f : \{-1,1\}^n \to \{-1,1\}$ be any function, and X any subset of $\{-1,1\}^n$. Suppose there exists a distribution μ on $\{-1,1\}^n$ such that $\max_{S\subseteq[n]} \left|\widehat{f\mu}(S)\right| \leq \epsilon$ and $\min_{x\in X} \mu(x) \geq \delta$. Then,

$$\mathsf{mon}_{\pm}(f) \geq \frac{\delta}{\epsilon + \delta \cdot \frac{|X^c|}{2^n}}.$$

Proof. Let $p : \{-1,1\}^n \to \mathbb{R}$ be any polynomial which sign represents f. By Fact 2.2.2,

$$\mathbb{E}_x[f(x)\mu(x)p(x)] = \sum_{S\subseteq[n]} \widehat{f\mu}(S)\widehat{p}(S) \leq \max_{S\subseteq[n]} \left|\widehat{f\mu}(S)\right| \cdot \max_{S\subseteq[n]} |\widehat{p}(S)| \cdot \mathsf{mon}(p) \tag{4.8}$$

$$\leq \epsilon \cdot \max_{S\subseteq[n]} |\widehat{p}(S)| \cdot \mathsf{mon}(p). \tag{4.9}$$

Note that

$$\begin{aligned}
\mathbb{E}_x[f(x)\mu(x)p(x)] &= \frac{1}{2^n}\sum_{x\in X} f(x)\mu(x)p(x) + \frac{1}{2^n}\sum_{x\in X^c} f(x)\mu(x)p(x) \\
&\geq \frac{\min_{x\in X}\mu(x)}{2^n}\left[\sum_{x\in\{-1,1\}^n} |p(x)| - |X^c| \cdot \max_{x\in X^c}|p(x)|\right] \\
&\qquad\qquad \text{Since } p \text{ sign represents } f \\
&\geq \delta \cdot \max_{S\subseteq[n]} |\widehat{p}(S)| - \frac{\delta}{2^n} \cdot |X^c| \cdot \max_{x\in\{-1,1\}^n} |p(x)|. \quad \text{Using Lemma 2.2.1}
\end{aligned}$$

Combining the above and Equation 4.8, we obtain

$$\begin{aligned}
&\epsilon \cdot \max_{S\subseteq[n]} |\widehat{p}(S)| \cdot \mathsf{mon}(p) \geq \delta \cdot \max_{S\subseteq[n]} |\widehat{p}(S)| - \frac{\delta}{2^n} \cdot |X^c| \cdot \max_{x\in\{-1,1\}^n} |p(x)| \\
\implies &\epsilon \cdot \mathsf{mon}(p) \geq \delta - \frac{\delta}{2^n} \cdot |X^c| \cdot \frac{\max_{x\in\{-1,1\}^n}|p(x)|}{\max_{S\subseteq[n]}|\widehat{p}(S)|} \geq \delta - \frac{\delta}{2^n} \cdot |X^c| \cdot \mathsf{mon}(p) \\
\implies &\mathsf{mon}(p) \geq \frac{\delta}{\epsilon + \delta \cdot \frac{|X^c|}{2^n}}.
\end{aligned}$$

□

The following theorem provides a signed monomial complexity lower bound against a function in $\mathsf{THR} \circ \mathsf{OR}$.

Theorem 4.7.3. Let $f = \mathsf{OMB}^0_\ell \circ \bigvee_{\ell^{1/3}+\log \ell} : \{-1,1\}^{\ell^{4/3}+\ell \log \ell} \to \{-1,1\}$. Then,

$$\mathsf{mon}_{\pm}(f) \geq 2^{\frac{\ell^{1/3}}{81}-3}.$$

Proof. The proof immediately follows from Theorem 4.7.2 and Theorem 4.4.1. □

This gives us a function f on n input variables, computable by linear sized $\mathsf{THR} \circ \mathsf{AND}$ circuits, such that for large enough n,

$$\mathsf{mon}_{\pm}(f) \geq 2^{\Omega\left(n^{1/4}\right)}.$$

4.8 Lower Bounds Against $\mathsf{MOD}_m \circ \mathsf{XOR}$

In this section, we analyze the unbounded-error communication complexity of XOR functions where the outer function is symmetric and its spectrum is periodic.

4.8.1 Introduction

We prove a UPP lower bound against functions of the form $\mathsf{MOD}^A_m \circ \mathsf{XOR}$ when MOD^A_m does not represent a constant function, the parity function, or the complement of parity. We remark here that, although very recent independent results of Hatami and Qian [HQ17] and Ada, Fawzi and Kulkarni [AFK17] subsume our results on UPP complexity of symmetric XOR functions, our methods vary vastly from theirs. We prove our lower bounds from first principles, and do not make a reduction to Sherstov's result [She11b] on symmetric AND functions. Interestingly, our UPP lower bounds are not obtained via linear programming duality, as opposed to our UPP lower bounds sketched earlier in this chapter, or even our PP and BPP lower bounds from earlier chapters.

The starting point of our work is Theorem 4.2.2 which relates the sign rank of a function $f : \{0,1\}^n \times \{0,1\}^n \to \mathbb{R}$ in terms of the minimum value taken by f and the spectral norm of the communication matrix of f.

Informally, the unbounded-error complexity of f is large if the minimum value taken by it is not too small, and the spectral norm is small. We then note in Lemma 2.2.3 that the spectral norm of $f \circ \mathsf{XOR}$ is just a scaling of the maximum Fourier coefficient of f. It turns out that $\mathsf{MOD}^{\{0\}}_3$ has a large principal Fourier coef-

ficient even though the other coefficients are inverse exponentially small. Thus, one cannot use Theorem 4.2.2 directly. Next, we prove in Theorem 4.8.6 that if the L_1 mass of a subset of the Fourier coefficients of f is sufficiently bounded away from 1, and the remaining coefficients are sufficiently small, we can still obtain a strong unbounded-error lower bound for $f \circ \mathsf{XOR}$. We then analyze the Fourier coefficients of MOD functions, to show that they satisfy the above properties, and this helps us prove lower bounds for $\mathsf{MOD}_m^A \circ \mathsf{XOR}$ for *odd* integers m with values upto $O(n^{1/2-\epsilon})$ as long as MOD_m^A does not represent a constant or parity function. This still does not prove hardness for all MOD functions with period at most $O(n^{1/2-\epsilon})$ since it can be proved, for example, $\left|\widehat{\mathsf{MOD}_4^{\{0\}}}(\emptyset)\right| + \left|\widehat{\mathsf{MOD}_4^{\{0\}}}([n])\right| = 1$, thus not allowing us to use Theorem 4.8.6. To handle this case, we make two crucial observations. One is that setting a few variables (which we can view as shifting the accepting set by a small amount) does not change the unbounded-error communication complexity of $\mathsf{MOD}_m^A \circ \mathsf{XOR}$ by much. The second is the fact that the unbounded-error complexity of $f \oplus g$ is at most the unbounded-error complexity of f plus that of g (Lemma 2.3.12). Armed with these facts, we use a shifting and XORing trick that enables us to reduce the modulus of the target MOD_m^A function to either 4 or a prime without using too large or too many shifts, or too many XORs. We then use induction on m to finish the proof of our main theorem regarding unbounded-error communication in this section (Theorem 4.8.2).

Let us first recall the definition of MOD functions.

Definition 4.8.1 (MOD functions and simple accepting sets)**.** A function $f : \{0,1\}^n \to \{-1,1\}$ is called a MOD function if there exists a positive integer $m < n$ and an 'accepting' set $A \subseteq [m]$ such that

$$f(x) = \begin{cases} -1 & \sum_{i=1}^{n} x_i \equiv k \bmod m \text{ for some } k \in A \\ 1 & \text{otherwise.} \end{cases}$$

We write $f = \mathsf{MOD}_m^A$. We call an accepting set A *simple* if MOD_m^A either represents the constant 0 function, constant 1 function, or the parity function or its negation. We also call the corresponding predicate *simple* in this case.

Our main theorem in this section is as follows.

Theorem 4.8.2. For any integer $m \geq 3$, express $m = j2^k$ uniquely, where j is either odd or 4, and k is a positive integer. Then for any non-simple A,

$$\mathsf{UPP}(\mathsf{MOD}_m^A \circ \mathsf{XOR}) \geq \Omega\left(\frac{n-km}{jm}\right) - \frac{2j\log j}{m}.$$

4.8.2 Fourier Analysis of Some Modular Functions

We first closely analyze the Fourier coefficients of functions of the type MOD_m^A, when m is odd, using exponential sums.

Claim 4.8.3. For odd m, and any $A \subseteq \{0, 1, \ldots, m-1\}$ which is not the full or empty set,

$$\left|\widehat{\mathsf{MOD}_m^A}(S)\right| \leq \begin{cases} 1 - \frac{2}{m} + 2m\left(\cos\left(\frac{\pi}{2m}\right)\right)^n & S = \emptyset \\ 2m\left(\cos\left(\frac{\pi}{2m}\right)\right)^n & S \neq \emptyset. \end{cases}$$

Zhang [Zha92] showed that for a fixed prime p, $\left|\widehat{\mathsf{MOD}_p}^{\{0\}}(\emptyset)\right| < 1 - \frac{1}{p}$, and $\left|\widehat{\mathsf{MOD}_p}^{\{0\}}(S)\right| = O\left(\frac{1}{2^{\Omega(n)}}\right)$ when $S \neq \emptyset$. We show that a similar bound holds for odd integers m for values up to $m = O(n^{1/2-\epsilon})$ using a different technique. In particular, we show that for $m = O(n^{1/2-\epsilon})$, the principal coefficient is roughly $1 - \frac{1}{m}$, and all other coefficients are exponentially small $\left(\frac{1}{2^{n^{\Omega(1)}}}\right)$, for any non simple accepting set A.

We use the characterization of the MOD_m^A function in terms of exponential sums to analyze its Fourier coefficients. Note that exponential sums have been used in similar contexts in previous papers as well. For example, the reader may refer to [Bou05, CGPT06, ACFN15]. The notation we use is that from [ACFN15].

Definition 4.8.4. Let $\omega = e^{2\pi i/m}$ be a primitive m-th root of unity. Then, for $x = \{0,1\}^n$, define

$$\mathsf{EXP}_m^{a,b}(x_1, \ldots, x_n) = \omega^{a\left(\left(\sum_{j=1}^n x_j\right) - b\right)}.$$

Let us now prove Claim 4.8.3.

Proof. First, we use exponential sums to represent a MOD_m^A function for odd m.

It is easy to check that for any integer k, and any input $x = (x_1, \ldots, x_n)$,

$$\frac{1}{m}\sum_{a=0}^{m-1} \mathsf{EXP}_m^{a,k}(x) = \begin{cases} 1 & |x| \equiv k \pmod m \\ 0 & \text{otherwise.} \end{cases}$$

Thus, for a general accepting set $A \subseteq [m]$,

$$\sum_{k\in A}\left(\frac{1}{m}\sum_{a=0}^{m-1}\mathsf{EXP}_m^{a,k}(x)\right) = \begin{cases} 1 & |x| \equiv k \pmod m \text{ for some } k \in A \\ 0 & \text{otherwise.} \end{cases}$$

Just by a simple linear transformation from $\{0,1\}$ to $\{-1,1\}$, we can express the MOD_m^A function in terms of exponential sums as follows.

$$\mathsf{MOD}_m^A(x) = 1 - \frac{2}{m}\sum_{k\in A}\left(\sum_{a=0}^{m-1}\mathsf{EXP}_m^{a,k}(x)\right) = \begin{cases} -1 & |x| \equiv k \pmod m \text{ for some } k \in A \\ 1 & \text{otherwise.} \end{cases} \tag{4.10}$$

Let us now look at the Fourier coefficients of MOD_m^A for odd m, and A not $\emptyset$ or $[m]$. Let us consider 2 cases, the first where S is non-empty, and the second where S is empty.

1. $S \neq \emptyset$.
   By Equation (2.1),

$$\begin{aligned}\widehat{\mathsf{MOD}_m^A}(S) &= \mathbb{E}_{x\in\{0,1\}^n}\left[\mathsf{MOD}_m^A(x)\chi_S(x)\right] \\ &= \mathbb{E}_{x\in\{0,1\}^n}\left[\chi_S(x)\right] - \frac{2}{m}\sum_{k\in A}\sum_{a=0}^{m-1}\mathbb{E}_{x\in\{0,1\}^n}\left[\mathsf{EXP}_m^{a,k}(x)\chi_S(x)\right]. \end{aligned} \tag{4.11}$$

   where the second equality follows from Equation (4.10) and linearity of expectation. Recall from Definition 4.8.4 that $\mathsf{EXP}_m^{a,b}(x) = \omega^{a\left(\left(\sum_{j=1}^n x_j\right)-b\right)}$. Note that when $a = 0$, $\mathbb{E}_{x\in\{0,1\}^n}\left[\mathsf{EXP}_m^{0,b}(x)\chi_S(x)\right] = \mathbb{E}_{x\in\{0,1\}^n}\left[\chi_S(x)\right] = 0$ since $S \neq \emptyset$. For $a \in \{1,\ldots,m-1\}$,

$$\begin{aligned}\mathsf{EXP}_m^{a,k}(x)\chi_S(x) &= \omega^{a\left(\left(\sum_{j=1}^n x_j\right)-k\right)}(-1)^{\sum_{i\in S}x_i} \\ &= \omega^{a\sum_{j=1}^n x_j}\cdot\omega^{-ak}\cdot(-1)^{\sum_{i\in S}x_i} \\ &= \omega^{-ak}\cdot(-\omega)^{a\sum_{i\in S}x_i}\cdot\omega^{a\sum_{j\notin S}x_j}.\end{aligned}$$

   Thus, in Equation (4.11), the first term is 0 since $S \neq \emptyset$. The summands with $a = 0$ contribute 0 to the expectation. Every other summand in the second term is of the form $\mathbb{E}_{x\in\{0,1\}^n}\left[\mathsf{EXP}_m^{a,k}(x)\chi_S(x)\right]$. Since the expectation is over the uniform distribution which is uniform and independent over the input bits,

the absolute value of such a term can be bounded as follows.

$$\begin{aligned}
\left|\mathbb{E}_{x\in\{0,1\}^n}\left[\mathsf{EXP}_m^{a,k}(x)\chi_S(x)\right]\right| &\leq \left|\mathbb{E}_{x\in\{0,1\}^n}\left[\omega^{-ak}\cdot(-\omega)^{a\sum_{i\in S}x_i}\cdot\omega^{a\sum_{j\notin S}x_j}\right]\right| \\
&\leq \left|\prod_{i\in S}\mathbb{E}_{x_i}(-\omega)^{ax_i}\right|\cdot\left|\prod_{j\notin S}\mathbb{E}_{x_j}\omega^{ax_j}\right| \\
&\leq \left|\left(\frac{1-\omega^a}{2}\right)\right|^{|S|}\left|\left(\frac{1+\omega^a}{2}\right)\right|^{n-|S|} \\
&\leq \max_{a\in\{1,\ldots,m-1\}}\left\{\left|\frac{1-\omega^a}{2}\right|^n,\left|\frac{1+\omega^a}{2}\right|^n\right\}.
\end{aligned}$$

Since $a\in\{1,\ldots,m-1\}$ and m is odd, it is fairly straightforward to check that the value of $\max_a\left\{\left|\frac{1-\omega^a}{2}\right|,\left|\frac{1+\omega^a}{2}\right|\right\}$ is maximized at $a=\frac{m\pm1}{2}$, and the value attained at the maximum is $\frac{1}{2}\sqrt{(1+\cos(\pi/m))^2+\sin^2(\pi/m)} = \frac{1}{2}\sqrt{2+2\cos(\pi/m)} = \cos(\pi/2m)$. Thus, the above, along with Equation (4.11) gives us

$$\left|\widehat{\mathsf{MOD}_m^A}(S)\right| \leq \left|\mathbb{E}_{x\in\{0,1\}^n}\left[\chi_S(x)\right]\right| + \left|\frac{2}{m}\sum_{k\in A}\sum_{a=0}^{m-1}\mathbb{E}_{x\in\{0,1\}^n}\left[\mathsf{EXP}_m^{a,k}(x)\chi_S(x)\right]\right| \tag{4.12}$$

$$\leq \frac{2(m-1)^2}{m}\cdot\left(\cos\left(\frac{\pi}{2m}\right)\right)^n \leq 2m\left(\cos\left(\frac{\pi}{2m}\right)\right)^n. \tag{4.13}$$

2. $S=\emptyset$.

One can follow a similar argument as above to analyze the absolute value of the principal Fourier coefficient. Note that in this case, the first term on the right hand side of Equation (4.11) is not 0, but 1. Next, note that for $a\in\{1,\ldots,m-1\}$, the same bound as in the previous case holds. That is,

$$\begin{aligned}
\left|\mathbb{E}_{x\in\{0,1^n\}}\left[\mathsf{EXP}_m^{a,k}(x)\chi_S(x)\right]\right| &\leq \prod_{i\in S}\mathbb{E}_{x_i}(-\omega)^{ax_i}\cdot\prod_{j\notin S}\mathbb{E}_{x_j}\omega^{ax_j} \\
&\leq \left|\left(\frac{1-\omega^a}{2}\right)\right|^{|S|}\cdot\left|\left(\frac{1+\omega^a}{2}\right)\right|^{n-|S|} \\
&\leq \left(\cos\left(\frac{\pi}{2m}\right)\right)^n.
\end{aligned}$$

by the same argument as in the case of $S\neq\emptyset$. However, when $S=\emptyset$ and $a=0$, we have $\mathbb{E}_{x\in\{0,1\}^n}\left[\mathsf{EXP}_m^{a,b}(x)\chi_\emptyset(x)\right]=1$ (unlike the case when $S\neq\emptyset$, where this expectation was 0).

Plugging these values into Equation (4.11) and using the above observations, we get

$$\begin{aligned}
\left|\widehat{\mathsf{MOD}_m^A}(\emptyset)\right| &\leq \left|\mathbb{E}_{x\in\{0,1\}^n}\left[\chi_\emptyset(x)\right] - \frac{2}{m}\sum_{k\in A}\mathbb{E}_{x\in\{0,1\}^n}\left[\mathsf{EXP}_m^{0,k}(x)\chi_\emptyset(x)\right]\right| \\
&\quad + \left|\frac{2}{m}\sum_{k\in A}\sum_{a=1}^{m-1}\mathbb{E}_{x\in\{0,1\}^n}\left[\mathsf{EXP}_m^{a,k}(x)\chi_\emptyset(x)\right]\right| && (4.14) \\
&\leq \left|1 - 2\frac{|A|}{m}\right| + 2m\left(\cos\left(\frac{\pi}{2m}\right)\right)^n && (4.15) \\
&\leq 1 - \frac{2}{m} + 2m\left(\cos\left(\frac{\pi}{2m}\right)\right)^n. && (\text{since } A \neq \emptyset, [m])
\end{aligned}$$

□

4.8.3 A Lower Bound for $\mathsf{MOD}_m^A \circ \mathsf{XOR}$

In this section, we show unbounded-error lower bounds for functions of the type $\mathsf{MOD}_m^A \circ \mathsf{XOR}$ for values of m up to $O(n^{1/2-\epsilon})$, when A is non-simple. Note that if A is a simple set, then either $\mathsf{MOD}_m^A \circ \mathsf{XOR}$ is a constant or MOD_m^A represents parity (or its negation), in which case $\mathsf{MOD}_m^A \circ \mathsf{XOR}$ just represents the parity function (or its negation), so its communication complexity (even deterministic) is very small. We prove a new sign rank lower bound criterion for XOR functions. Theorem 2.3.9 tells us that the log of the sign rank of a communication matrix is essentially equivalent to the unbounded-error communication complexity of the function.

Let $f : \{0,1\}^n \to \mathbb{R}$, and let A denote the communication matrix of $f \circ \mathsf{XOR}$. In order to show a lower bound on the sign rank of $f \circ \mathsf{XOR}$, it suffices to show an upper bound on the spectral norm of the communication matrix of $f \circ \mathsf{XOR}$.

Combining Theorem 4.2.2 and Theorem 2.2.3, we get

Corollary 4.8.5. Let $f : \{0,1\}^n \to \mathbb{R}$ be any real valued function and let A denote the communication matrix of $f \circ \mathsf{XOR}$. Then,

$$\mathsf{sr}(A) \geq \frac{1}{\max\limits_{S\subseteq[n]}\left|\widehat{f}(S)\right|} \cdot \min_x |f(x)|.$$

Thus, $\mathsf{sr}(f\circ\mathsf{XOR}) = 2^{\Omega(n)}$ for any $\{-1,1\}$ valued function with exponentially small L_∞ Fourier norm.

Note that we cannot use the outer function to be MOD_p (for a constant p) in Corollary 4.8.5, since its principal Fourier coefficient is a constant (though sufficiently bounded away from 1, which we crucially require). The following theorem allows us to ignore a subset of large Fourier coefficients, as long as their mass is not too large, which gives us a stronger condition for unbounded-error hardness of XOR functions.

Theorem 4.8.6. For any function $f : \{0,1\}^n \to \{-1,1\}$, and any collection of sets $\mathcal{S} \subseteq supp(\widehat{f})$, if $\sum_{S\in\mathcal{S}} \left|\widehat{f}(S)\right| \leq 1-\delta$, and $\max_{S\notin\mathcal{S}} \left|\widehat{f}(S)\right| \leq c$. Then, $\mathsf{sr}(f\circ\mathsf{XOR}) \geq \frac{\delta}{c}$.

Proof. Define $f' : \{0,1\}^n \to \mathbb{R}$ by $f'(x) = f(x) - \sum\limits_{S\in\mathcal{S}} \widehat{f}(S)\chi_S(x)$. Notice

$$\min_{x\in\{0,1\}^n} |f'(x)| \geq 1 - \sum_{S\in\mathcal{S}} \left|\widehat{f}(S)\right| \geq \delta.$$

Also note that $\forall S \in \mathcal{S}, \widehat{f'}(S) = 0$, and $\forall S \notin \mathcal{S}, \widehat{f'}(S) = \widehat{f}(S)$. Thus, $\max\limits_{S\subseteq[n]} \left|\widehat{f'}(S)\right| \leq c$. It is easy to see that f' sign agrees with f. Thus, the sign rank of these functions agree by definition. Using Corollary 4.8.5, we have

$$\mathsf{sr}(f\circ\mathsf{XOR}) = \mathsf{sr}(f'\circ\mathsf{XOR}) \geq \frac{1}{\max\limits_{S\notin\mathcal{S}} \left|\widehat{f'}(S)\right|} \cdot \min_x |f'(x)| \geq \frac{\delta}{c}. \tag{4.16}$$

□

Let us first recall the Complete Quadratic function, whose Fourier coefficients were analyzed by Bruck [Bru90]. Define $\mathsf{CQ} : \{0,1\}^n \to \{-1,1\}$ by

$$\mathsf{CQ}(x) = \mathsf{MOD}_4^{\{0,1\}}(x).$$

Lemma 4.8.7 ([Bru90])**.** For even n, $\left|\widehat{\mathsf{CQ}}(S)\right| = 2^{-n/2}$ for all $S \subseteq [n]$. For odd n, $\left|\widehat{\mathsf{CQ}}(S)\right| \in \{0, 2^{-(n-1)/2}\}$ for all $S \subseteq [n]$.

For notational convenience, we use the notation $U(f)$ to represent $\mathsf{UPP}(f\circ\mathsf{XOR})$ in this section. We also use the notation $U(\mathsf{MOD}_m)$ to denote the minimum value of $U(\mathsf{MOD}_m^A)$ over all non-simple accepting sets A.

Theorem 4.8.8. For m odd, and and $A \subseteq \{0,1,\ldots,m-1\}$ which is not the empty set or full set,

$$U(\mathsf{MOD}_m^A) = \Omega(n/m^2) - 2\log(m).$$

Proof. In Theorem 4.8.6, use $\mathcal{S} = \emptyset$. The values obtained using Claim 4.8.3 are $\delta = \frac{2}{m} - 2m\left(\cos\left(\frac{\pi}{2m}\right)\right)^n$, and $c = 2m\left(\cos\left(\frac{\pi}{2m}\right)\right)^n$. Hence,

$$\begin{aligned}\mathsf{sr}(\mathsf{MOD}_m^A \circ \mathsf{XOR}) &\geq \left(\frac{2}{m} - 2m\left(\cos\left(\frac{\pi}{2m}\right)\right)^n\right) \cdot \frac{1}{2m\left(\cos\left(\frac{\pi}{2m}\right)\right)^n} \\ &\geq \frac{1}{m^2\left(\cos\left(\frac{\pi}{2m}\right)\right)^n} - 1.\end{aligned}$$

Using a standard series expansion for $\cos\theta$, and the fact that $1 - x \leq e^{-x}$ for all $x \in \mathbb{R}$, we get

$$\mathsf{sr}(\mathsf{MOD}_m^A \circ \mathsf{XOR}) \geq \frac{2^{\Omega(n/m^2)}}{m^2} - O(1).$$

Thus, using the equivalence between sign rank and unbounded-error communication complexity from Theorem 2.3.9,

$$U(\mathsf{MOD}_m^A) = \Omega(n/m^2) - 2\log(m).$$

□

This already shows us that the unbounded-error complexity of functions of the type MOD_m^A are large when m is odd, and A is not the full set or empty set, for m up to $O(n^{1/2-\epsilon})$. Note that one cannot use Theorem 4.8.6 to prove a sign rank lower bound for $\mathsf{MOD}_4^{\{0\}}$, since $\left|\widehat{\mathsf{MOD}_4^{\{0\}}}(\emptyset)\right| + \left|\widehat{\mathsf{MOD}_4^{\{0\}}}([n])\right| = 1$, which can be easily checked. In Claim 4.8.11, we also show hardness for the case when $m = 4$ and A is not a simple accepting set.

In the analysis of our main claim (Theorem 4.8.12), we will be concerned with the size of the input string. For notational convenience, we add a subscript to MOD_m^A which denotes the input size. That is,

$$\mathsf{MOD}_{m,n}^A : \{0,1\}^n \to \{-1,1\},$$

and we define it exactly the same as in Definition 4.8.1.

We denote the sumset $A + \{p\} = \{a + p \mid a \in A\}$ (the sums are modulo m, where m is the period of the MOD function we are interested in) by $A + p$ for convenience.

Lemma 4.8.9. Suppose $\mathsf{MOD}_{p,n}^{A'} = \mathsf{MOD}_{m,n}^A \oplus \mathsf{MOD}_{m,n}^{A+i}$ for some $p < m$, and any integer i. Then,

$$U(\mathsf{MOD}_{m,n}^A) \geq \frac{U(\mathsf{MOD}_{p,n-m}^{A'})}{2}.$$

We recall the following simple yet powerful lemma.

Lemma 4.8.10 (Restatement of Lemma 2.3.12)**.** For any functions $f, g : \{0,1\}^n \to \{-1,1\}$,

$$U(f \oplus g) \leq U(f) + U(g).$$

Proof of Lemma 4.8.9. Since $\mathsf{MOD}^{A'}_{p,n} = \mathsf{MOD}^{A}_{m,n} \oplus \mathsf{MOD}^{A+i}_{m,n}$, applying Lemma 4.8.10 gives us

$$U(\mathsf{MOD}^{A'}_{p,n-m}) \leq U(\mathsf{MOD}^{A}_{m,n-m}) + U(\mathsf{MOD}^{A+i}_{m,n-m}).$$

The first term on the right is at most $U(\mathsf{MOD}^{A}_{m,n})$ since we can just pad m number of 0's each to Alice's and Bob's inputs and obtain a protocol (of the same cost) for $\mathsf{MOD}^{A}_{m,n-m}$ given a protocol for $\mathsf{MOD}^{A}_{m,n}$ The second term is also at most $U(\mathsf{MOD}^{A}_{m,n})$ for a similar reason. Pad $m-i$ number of 1's and i number of 0's each to Alice's and Bob's inputs. It is easy to see that $\mathsf{MOD}^{A+i}_{m,n-m}(x,y) = -1$ if and only if $\mathsf{MOD}^{A}_{m,n}(x',y') = -1$, where x' and y' are x and y padded with $m-i$ 1's and i 0's respectively. The lemma now follows. □

Let us analyze the unbounded-error communication complexity of $\mathsf{MOD}^{A}_{4} \circ \mathsf{XOR}$ for various accepting sets A. Note that if $A = \{0,2\}$ or $\{1,3\}$, then $\mathsf{MOD}^{A}_{4} \circ \mathsf{XOR}$ is just parity or its negation respectively. Its communication complexity is a constant in these cases. Let us look at the other cases.

Claim 4.8.11. Suppose A is not a simple accepting set. Then, $U(\mathsf{MOD}^{A}_{4}) = \Omega(n)$.

Proof.

1. $A = \{0,1\}$. Then, $\mathsf{MOD}^{A}_{4} = \mathsf{CQ}$, and by Lemma 4.8.7,

$$U(\mathsf{CQ}) \geq n/2.$$

2. $|A| = 2$, and MOD^{A}_{4} does not represent parity (or its negation). Then, this is clearly a translate of CQ, and

$$U(\mathsf{MOD}^{A}_{4,n}) \geq U(\mathsf{MOD}^{\{0,1\}}_{4,n-4}) \geq (n-4)/2.$$

3. A is non simple and does not fall in the previous 2 cases. Without loss of generality, may assume $|A| = 1$ because if it was 3, the complexity of MOD^{A}_{m} is the same as $\mathsf{MOD}^{A^c}_{m}$, and $|A^c| = 1$. In this case, we can use Lemma 4.8.9 to get

$$U(\mathsf{MOD}^{A}_{4} \oplus \mathsf{MOD}^{A+1}_{4}) \geq U(\mathsf{MOD}^{A'}_{m}).$$

for some non simple A' of size 2. From the previous case, we conclude,

$$U(\mathsf{MOD}^A_{4,n}) \geq U(\mathsf{MOD}^{A'}_{4,n-4}) \geq \frac{((n-4)/2)-4}{2} = (n-12)/4.$$

□

Recall our main theorem of this section (Theorem 4.8.2), which says that any function of the type $\mathsf{MOD}^A_m \circ \mathsf{XOR}$ for any non-simple A is hard for UPP protocols, for values of m up to $O(n^{1/2-\epsilon})$.

Theorem. For any integer $m \geq 3$, express $m = j2^k$ uniquely, where j is either odd or 4, and k is a positive integer. Then for any non-simple A,

$$U(\mathsf{MOD}^A_{m,n}) \geq \Omega\left(\frac{n-km}{jm}\right) - \frac{2j\log j}{m}.$$

Note that since k is at most $\log(n)$, and j is at most m, this gives us an $n^{\Omega(1)}$ lower bound on the unbounded-error communication complexity of $\mathsf{MOD}^A_m \circ \mathsf{XOR}$ for any non-simple accepting set A, for m as large as $O(n^{1/2-\epsilon})$.

We require the following claim to prove Theorem 4.8.2.

Claim 4.8.12. For any integer $m \geq 3$, and for all representations $m = j2^k$ for some $j \geq 3$ and a positive integer k, and any non-simple $A \subseteq [m]$, we have

$$U(\mathsf{MOD}^A_{m,n}) \geq \frac{U(\mathsf{MOD}_{j,n-km})}{2^k}.$$

Let us first see how Claim 4.8.12 implies Theorem 4.8.2. Recall that Theorem 4.8.8 gave us

$$U(\mathsf{MOD}_{j,n}) = \Omega(n/j^2) - 2\log(j).$$

This, along with Claim 4.8.11 and Claim 4.8.12, implies that if $m = j2^k$ where j is either 4 or odd,

$$U(\mathsf{MOD}^A_{m,n}) \geq \frac{U(\mathsf{MOD}_{j,n-km})}{2^k} \geq \frac{\Omega\left(\frac{(n-km)}{j^2}\right) - 2\log(j)}{m/j} \geq \Omega\left(\frac{n-km}{jm}\right) - \frac{2j\log j}{m}$$

Let us now prove Claim 4.8.12.

Proof. We prove this by induction on m.

1. The base cases are when m is odd. In this case, the hypothesis is trivially true since $m = j2^k$ can only imply $j = m, k = 0$.

2. Suppose $m = 2p$, where p is odd. Let $a = xy$ denote the characteristic vector of the accepting set A, where x corresponds to the first p elements, and y the last p elements. We interchangeably use the notation MOD_m^A and MOD_m^a when a is the characteristic vector of the set A. Our assumption is that a is not the all 0, or all 1, or the parity (negation of parity) vector. Let $x \oplus y$ denote the bitwise XOR of x and y.

   (a) Suppose $x \oplus y$ is neither the all 0 or all 1 vector. Since $x \oplus y$ does not represent a simple accepting set A, in this case, $\mathsf{MOD}_m^A \oplus \mathsf{MOD}_m^{A+p} = \mathsf{MOD}_p^{x \oplus y}$. By Lemma 4.8.9,

$$U(\mathsf{MOD}_{m,n}^A) \geq \frac{U(\mathsf{MOD}_{p,n-m}^{x \oplus y})}{2}.$$

   (b) If $x \oplus y$ is the all 0 vector, then $x = y$, and neither of them are all 0 or all 1. This means $\mathsf{MOD}_m^a = \mathsf{MOD}_p^x$.

   (c) If $x \oplus y$ is the all 1 vector, this means $y = x^c$. Consider $A' = A + 1$. One may verify that $A \oplus A'$ has characteristic vector $a'' = bb$.

      i. If b is not the all 0 or all 1 string, $\mathsf{MOD}_p^b = \mathsf{MOD}_m^A \oplus \mathsf{MOD}_m^{A+1}$. Use Lemma 4.8.9 and conclude

$$U(\mathsf{MOD}_{m,n}^A) \geq \frac{U(\mathsf{MOD}_{p,n-m}^b)}{2}.$$

      ii. It is easy to check that b can never be the all 0 vector.

      iii. Close inspection reveals that if b is the all 1 vector, then the original vector a must represent parity or its negation, which was not the case by assumption.

3. Suppose $m = 2k$, where k is even. Again, let $a = xy$, where a is the characteristic vector of accepting set A.

   (a) If $x \oplus y$ is neither the all 0 string, all 1 string, nor does it represent parity (or its negation), then $\mathsf{MOD}_m^A \oplus \mathsf{MOD}_m^{A+k} = \mathsf{MOD}_k^{x \oplus y}$. By Lemma 4.8.9,

$$U(\mathsf{MOD}_{m,n}^A) \geq \frac{U(\mathsf{MOD}_{k,n-m}^{x \oplus y})}{2}.$$

   By the induction hypothesis, the claim is true for $\mathsf{MOD}_{k,n-m}^{x \oplus y}$. It is easy to see that this implies the claim for $\mathsf{MOD}_{m,n}^A$.

(b) If $x \oplus y$ is the all 0 vector, then $x = y$, and neither of them are all 0 or all 1. This means MOD_m^a is the same as MOD_k^x.

(c) If $x \oplus y$ is the all 1 vector, this means $y = x^c$. Consider $A' = A + 1$. One may verify that $A \oplus A'$ has a characteristic vector of the form $a'' = bb$.

i. If b is neither the all 0 or all 1 string, nor does it represent parity (or its negation), then $\mathsf{MOD}_k^b = \mathsf{MOD}_m^A \oplus \mathsf{MOD}_m^{A+1}$. Use Lemma 4.8.9 and conclude

$$U(\mathsf{MOD}_{m,n}^A) \geq \frac{U(\mathsf{MOD}_{k,n-m}^b)}{2}.$$

By the induction hypothesis, the theorem is true for MOD_k^b since b does not represent the all 0, all 1, or parity (or its negation) string. The theorem now follows easily for $\mathsf{MOD}_{m,n}^A$.

ii. It is easy to check that b can never be the all 0 or all 1 vector.

iii. One may check that b can be the parity (or its negation) vector only if $k \equiv 2 (\text{mod } 4)$, and A must have represented CQ (or a translate of it by at most 2) which we know to be hard. In this case, we obtain

$$U(\mathsf{MOD}_m^a) = \Omega(n).$$

(d) If $x \oplus y$ represents the parity (or negation of parity) vector, then consider $A' = A + 2$. It is simple to verify that the characteristic vector of $A \oplus A'$ is of the form zz.

i. If z is not the all 0 or all 1 string, or does not represent parity (or its negation), then we have $\mathsf{MOD}_m^A \oplus \mathsf{MOD}_m^{A+2} = \mathsf{MOD}_k^z$. Use Lemma 4.8.9 to say

$$U(\mathsf{MOD}_{m,n}^A) \geq \frac{U(\mathsf{MOD}_{k,n-m}^z)}{2}.$$

The claim now follows because of the induction hypothesis.

ii. One may verify (by considering cases when k has residue either 0 or 2 modulo 4) that z cannot be the all 0 or all 1 string.

iii. If z represents parity or its negation, then it can be checked that the only case when this occurs is when A represented a non simple accepting set, say X, modulo 4. Thus,

$$U(\mathsf{MOD}_m^a) = U(\mathsf{MOD}_4^X) = \Omega(n).$$

□

4.8.4 Circuits

In this section, we observe the consequences of Theorem 4.8.2 to circuit lower bounds.

Theorem 4.8.13. Any $\mathsf{THR} \circ C$ circuit computing $\mathsf{MOD}_m^A \circ \mathsf{XOR}$ must have size

$$s \geq 2^{\Omega\left(\frac{n-km}{jm}\right)-\frac{2j\log j}{m}-c},$$

where c is the deterministic communication complexity of C, and $m = j2^k$ is the unique representation of $m \geq 3$, where j is either odd or 4, and k is a positive integer.

Proof. It follows directly from Theorem 2.3.9, Claim 2.3.10 and Theorem 4.8.2. □

Thus, we obtain that for m up to $O(n^{1/2-\epsilon})$, and any non-simple A, $\mathsf{MOD}_m^A \circ \mathsf{XOR}$ is not in subexponential sized $\mathsf{THR} \circ \mathsf{MAJ}$. A similar argument shows that $\mathsf{MOD}_m^A \circ \mathsf{XOR}$ is not even in subexponential size $\mathsf{THR} \circ \mathsf{SYM}$, where SYM denotes the class of all symmetric functions. This is because all symmetric functions have deterministic communication complexity bounded above by $O(\log(n))$.

This generalizes one particular result of Krause and Pudlák [KP97], and of Zhang [Zha92] which state that $\mathsf{MOD}_m^{\{0\}} \notin \mathsf{THR} \circ \mathsf{PAR}$, where PAR denotes the class of all parity gates. This is because we have shown that $\mathsf{MOD}_m^A \circ \mathsf{XOR} \notin \mathsf{THR} \circ \mathsf{SYM}$, which implies $\mathsf{MOD}_m^A \circ \mathsf{XOR} \notin \mathsf{THR} \circ \mathsf{PAR}$. This implies $\mathsf{MOD}_m^A \notin \mathsf{THR} \circ \mathsf{PAR}$.

4.9 References

The results presented in this chapter are based on joint work with Arkadev Chattopadhyay ([CM17c] and Section 6 in [CM17a]).

Chapter 5

Multi-Party Communication

5.1 Introduction

We first recall the "number-on-forehead" (NOF) model of multi-party communication, introduced by Chandra, Furst and Lipton [CFL83]. In this model, there are k players each with an input metaphorically held on their foreheads. Every forehead is visible to a player except her own. Several communication complexity class separations are known in the two-player setting. The interested reader may refer to [GPW18] for an excellent overview of such known separations. Recall that Babai et al. [BFS86] argue that protocols with polylogarithmic (of input length) communication cost is a natural notion of efficient protocols, just as polynomial time is a notion of efficient computation on Turing machines. This correspondence also naturally extends easily to the NOF model and gives rise to complexity classes such as $\mathsf{P}_k, \mathsf{BPP}_k, \mathsf{NP}_k, \mathsf{PP}_k$, etc.

We consider the classes PP_k and UPP_k. The definitions of these classes are analogous to those in the two-party setting as in earlier chapters. We formally define PP_k and UPP_k in Section 5.2. The inclusion $\mathsf{PP}_k \subseteq \mathsf{UPP}_k$ is straightforward. While a strict separation between the classes was known for $k = 2$ ([BVdW07, She08], and reproved by us in Chapter 3 using different techqniues), the corresponding separation question for $k \geq 3$ players remained unaddressed in the literature.

5.1.1 Our Results

We consider a simple and natural extension of the function defined by Goldmann, Håstad and Razborov [GHR92], which we define as follows.

Definition 5.1.1. Let

$$P(x, y_1, \ldots, y_k) := \sum_{i=0}^{n-1} \sum_{j=0}^{n4^k-1} 2^i y_{1j} \ldots y_{kj}(x_{i,2j} + x_{i,2j+1})$$

where $x \in \{\pm 1\}^{2n^2 4^k}$, $y_i \in \{\pm 1\}^{n4^k}$ for each i.
We set $\mathsf{GHR}_k^N(x, y_1, \ldots, y_k) := \mathsf{sgn}(2P(x, y_1, \ldots, y_k) + 1)$, where $N = 2n^2 4^k$.

Our main theorem regarding multi-party communication uses GHR_k^N to separate PP_k from UPP_k for $k \leq \delta \log N$, for any constant $\delta < 1/4$. Note that there is a natural way to assign the input variables to GHR_k^N to $k+1$ players as follows: x is Player 1's input, and y_i is Player $(i+1)$'s input (for $i = 1, \ldots, k$). We recall our main theorem regarding multi-party communication below.

Theorem 5.1.2. Let Π be any $(k+1)$-party probabilistic public-coin protocol computing the GHR_k^N function with advantage $\epsilon > 0$. Then,

$$\mathrm{Cost}(\Pi) + \log(1/\epsilon) \geq \Omega\left(\frac{\sqrt{N}}{4^k} - \log N - k\right).$$

Theorem 5.1.2 gives a PP_{k+1} lower bound for GHR_k^N. On the other hand, note that GHR_k^N is a composition of a linear threshold function with N parities of arity $k+1$. A well-known simple fact (essentially Claim 2.3.10) says that every such function has a UPP_{k+1} protocol of cost $O(\log N)$. This immediately yields the following separation result.

Corollary 5.1.3. For all $1 \leq k \leq \delta \log N$, the GHR_k^N function is not in PP_{k+1} but is in the class UPP_{k+1}, for any constant $0 < \delta < \frac{1}{4}$.

An additional motivation for our work comes from the study of constant-depth Threshold circuits. We work with the GHR function [GHR92] which is easily seen to be the composition of a linear threshold function and Parity.

Goldmann et al. [GHR92] showed that although $\mathsf{THR} \in \mathsf{MAJ} \circ \mathsf{MAJ}$, a simple function computable by linear-size circuits of the form $\mathsf{THR} \circ \mathsf{PAR}_2$ requires exponential size to be computed by $\mathsf{MAJ} \circ \mathsf{SYM}$ circuits, where SYM denotes gates computing arbitrary symmetric functions. We strengthen their result to depth-three circuits as follows.

Theorem 5.1.4. For each $k \geq 2$, the function GHR_k^N can be computed by linear-size $\mathsf{THR} \circ \mathsf{PAR}_{k+1}$ circuits, but requires size $2^{\Omega\left(\frac{\sqrt{N}}{k4^k} - \frac{\log N}{k}\right)}$ to be computed by depth-three circuits of the form $\mathsf{MAJ} \circ \mathsf{SYM} \circ \mathsf{ANY}_k$.

Let us remark that Theorem 5.1.4 continues to yield non-trivial bounds as long as $k < \delta \log N$ for any constant $0 < \delta < 1/4$. It is also worth noting that a result of [GHR92] ($\mathsf{MAJ} \circ \mathsf{MAJ} = \mathsf{MAJ} \circ \mathsf{THR}$) immediately yields, from the above theorem, the following interesting result.

Corollary 5.1.5. The function GHR_k^N can be computed by linear-size $\mathsf{THR} \circ \mathsf{PAR}_{k+1}$ circuits but requires size $2^{\Omega\left(\frac{\sqrt{N}}{k4^k} - \frac{\log N}{k}\right)}$ to be computed by depth-three circuits of the form $\mathsf{MAJ} \circ \mathsf{THR} \circ \mathsf{ANY}_k$.

We obtain another corollary, demonstrating the existence of a linear threshold function not computable by a low-weighted signed sum of low $\mathbb{F}_2$-degree polynomials (see Corollary 5.4.4).

5.1.2 Related Work

An anonymous reviewer, and subsequently Sherstov [She16a], pointed out that a more off-the-shelf $\Omega(N^{1/7})$ separation between PP_k and UPP_k is implicit in prior work by combining known results of Sherstov [She11a] and Beigel [Bei94]. The best PP_k lower bound that one would get using functions obtained in this way is $\Omega(N^{2/11})$, using a more recent result of Thaler [Tha16], which is weaker than the $\Omega(N^{1/2})$ bound obtained in our Theorem 5.1.2. After our result was published in a technical report, Sherstov [She16c] showed that by carefully piecing together approximation-theoretic ideas from his earlier work [She13a] and the result in [She16b], one can obtain an $\Omega(N/4^k)$ lower bound for a non-explicit function. This can be made to reproduce our lower bound, for an explicit function that is similar to ours. We note that while our result separates PP_k from UPP_k for up to $k \leq (1/4 - \epsilon) \log N$ players, Sherstov's separation [She16c] extends to $k \leq (1/2 - \epsilon) \log N$ players. On the other hand, our method is elementary and self-contained. Using first principles, we prove a strong PP_k lower bound for a function which remained unanalyzed until this result.

The route of combining earlier work of Sherstov [She16b] uses unique-disjointness as the inner function. With such an inner function, the previous techniques work with any outer function, like OMB, that has large approximation error for any polynomial of degree sufficiently smaller than N [Bei94]. This is in contrast to our use of XOR as the inner function. It is not very difficult to see that $\mathsf{OMB} \circ \mathsf{XOR}$ has very efficient PP_k protocols for all $k \geq 2$. Thus, our argument has to exploit some feature of the outer function that is not possessed by functions like OMB. We find this an independently interesting aspect of the technique used in this work. Indeed, there

has been considerable recent interest in studying the communication complexity of XOR functions, this thesis hopefully proving to be an example.

In summary, progress on separating communication complexity classes in the NOF model has been slow. This work is the first one to explicitly address the question of separating PP_k and UPP_k for $k > 2$.

5.1.3 Our Proof Technique and Organization

Recall (cf. Theorem 2.3.8) that PP complexity and discrepancy are equivalent notions in the 2-party setting. It is not hard to show that a similar equivalence holds in the multiparty setting as well. Thus, proving a PP lower bound against GHR_k^N is equivalent to proving a discrepancy upper bound. To estimate the discrepancy of GHR_k^N, we extend ideas from [GHR92] who estimated this in the setting of two players. The basic intuition can be seen after observing that for a given setting of $y_1, \ldots, y_k$ the function GHR_k^N essentially depends on the sign of a plus-minus combination of A_j's for $0 \leq j \leq n4^k - 1$, where

$$A_j := \frac{1}{2} \sum_{i=0}^{n-1} 2^i \big(x_{i,2j} + x_{i,2j+1}\big).$$

The relevant sign of each A_j depends on the parity of the bits $y_{1,j}, \ldots, y_{k,j}$. Further, the set of bits in x that each A_j depends on is disjoint from the set of bits that $A_{j'}$ depends on, whenever $j \neq j'$. We sample x such that each A_j is an i.i.d. binomial distribution centered at 0 with range $[-2^n + 1, 2^n - 1]$. Let this distribution be μ_X. We sample each y_i uniformly at random. We want to ensure that GHR_k^N, under this distribution, behaves in a way that leaves the players with little clue about the outcome unless the relevant sign to be associated with each A_j is determined. The distribution defined above is a product distribution. Sherstov [She08] showed that GHR has large discrepancy under product distributions. Thus, as done in [GHR92], one is forced to sample in a slightly more involved way. First sample y's uniformly at random. Then sample x according to μ_X, conditioned on the fact that $P = \sum_{j=0}^{n4^k-1} A_j y_{1,j} \cdots y_{k,j}$ is very close to its mean compared to its standard deviation (which is as high as $2^{\Omega(n)}$). Note that the mean of each A_j is 0, which gives us plenty of room to exploit. This turns out to be the hard distribution but to establish this requires technical work. This is mainly because analyzing the discrepancy under non-product distributions is difficult. As a first step to overcome this difficulty, we follow the ideas of Goldmann et al. [GHR92], and show that it is sufficient to show

an upper bound on the discrepancy of a function related to the GHR function under a particular product distribution. Analyzing the discrepancy of this related function on the obtained product distribution is still non-trivial, and this is the main technical contribution of our work.

Organization: Section 5.2 develops the basic notions and lemmas. Section 5.3 establishes our main technical result, Theorem 5.3.1, which gives an upper bound on the k-wise discrepancy of the GHR function. Using this, we prove Theorem 5.1.2 and Corollary 5.1.3. Section 5.4 derives the circuit consequences of Theorem 5.1.4 and Corollary 5.1.5. We discuss future directions and open problems in Chapter 7.

5.2 Preliminaries

5.2.1 The NOF Model

In the k-party model of Chandra et al. [CFL83], k players with unlimited computational power wish to compute a function $f : X_1 \times \cdots \times X_k \to \{-1, 1\}$ on some input $x = (x_1, \ldots, x_k)$. For the purpose of this work, we consider inputs of the form $X_i = \{-1, 1\}^{n_i}$. On input x, player i is given $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k)$, which is why it is figuratively said that x_i is on the i'th player's forehead. Players communicate by writing on a blackboard, so every player sees every message. We denote by $D_k(f)$ the deterministic k-party communication complexity of f, namely the number of bits exchanged in the best deterministic protocol for f on the worst-case input.

The PP_k and UPP_k costs of functions are defined analogous to the definitions in the two-party setting. However, we recall the definitions here for completeness.

A probabilistic protocol Π with access to public (private) randomness computes f with advantage ϵ if the probability that Π and f agree is at least $1/2 + \epsilon$ for all inputs. The cost of Π is the maximum number of bits it communicates over it's internal random choices in the worst case. Let us define $R_\epsilon^{pub}(f)$ ($R_\epsilon^{priv}(f)$) to be the cost of the best such protocol. Note that for convenience, we deviate from the notation defined in [KN97]. Define

$$\mathsf{PP}_k(f) := \min_\epsilon \left[R_\epsilon^{pub}(f) + \log\left(\frac{1}{\epsilon}\right) \right], \qquad \mathsf{UPP}_k(f) := \min_\epsilon \left[R_\epsilon^{priv}(f) \right]. \quad (5.1)$$

Note that privateness of the random coins is essential in the definition of UPP_k. It is a simple exercise to show that every function can be computed by unbounded-error protocols using 2 bits if allowed public coins. Define $\mathsf{PP}_k = \{f : \mathsf{PP}_k(f) =$

polylog$(N)\}$ and $\mathsf{UPP}_k = \{f : \mathsf{UPP}_k(f) = \text{polylog}(N)\}$, where N is the maximum length of an input to a player. Each element in either of these classes refers to a family of functions, f, one for each input length.

5.2.2 Cylinder Intersections, Discrepancy and the Cube Norm

Let $f : X_1 \times \cdots \times X_k \to \{-1, 1\}$. A subset $S_i \subseteq X_1 \times \cdots \times X_k$ is a cylinder in the i'th direction if membership in S_i does not depend on the i'th coordinate. A subset S is called a cylinder intersection if it can be represented as the intersection of k cylinders, $S = \cap_{i=1}^{k} S_i$, where S_i is a cylinder in the i'th direction.

Definition 5.2.1. Let μ be a distribution on $X_1 \times \cdots \times X_k$. The discrepancy of f according to μ, $\mathsf{disc}_\mu^k(f)$ is

$$\max_S \Big| \Pr_\mu[f(x_1, \ldots, x_k) = 1 \wedge (x_1, \ldots, x_k) \in S] - \Pr_\mu[f(x_1, \ldots, x_k) = -1 \wedge (x_1, \ldots, x_k) \in S] \Big| \quad (5.2)$$

where the maximum is taken over all cylinder intersections S.

The k in disc_μ^k denotes the dimension of the underlying cylinder intersections. We will drop this superscript when it is clear from the context what k is. Let $\mathsf{disc}(f) = \min_\mu \mathsf{disc}_\mu^k(f)$.

The discrepancy method, due to Babai, Nisan and Szegedy [BNS92], is a powerful tool that gives lower bounds on randomized communication complexity in terms of discrepancy. The following lemma can be found, for example, in [KN97].

Lemma 5.2.2. $R_\epsilon^{pub}(f) \geq \log(2\epsilon/\mathsf{disc}(f))$.

We now recall a useful technique that helps prove upper bounds on the discrepancy of a function under a product distribution. It is a standard lemma (see, for example, [Cha09] and [Raz00]).

Lemma 5.2.3. Let $f : X \times Y_1 \times \cdots \times Y_k \to \mathbb{R}$, $\mu = \mu_X \times \mu_1 \times \cdots \times \mu_k$ be any product distribution, and let $\phi : X \times Y_1 \times \cdots \times Y_k \to \{0, 1\}$ be the characteristic function of

a cylinder intersection. Then,

$$|\mathbb{E}_{\mu}[f(x, y_1, \ldots, y_k)\phi(x, y_1, \ldots, y_k)]| \leq \left(\mathbb{E}_{y_1^0, y_1^1, \ldots, y_k^0, y_k^1}\left[\left|\mathbb{E}_{x \sim \mu_X} \Pi_{a \in \{0,1\}^k} f(x, y_1^{a_1}, \ldots, y_k^{a_k})\right|\right]\right)^{1/2^k}$$

where $y_i^0 \sim \mu_i,\ y_i^1 \sim \mu_i$ are sampled independently for each $i \in [k]$.

Remark 5.2.4. When f is $\{-1, 1\}$ valued, the left hand side represents the discrepancy of f over the cylinder intersection ϕ with respect to the distribution μ. However, for our purposes, we are required to use the inequality when f is $\{-1, 1, 0\}$ valued.

5.2.3 The Binomial Distribution

Definition 5.2.5. Let $B(N)$ denote the distribution obtained as the sum of $2N$ independent Bernoulli variables, each of which take values $1/2, -1/2$ with probability $1/2$ each.

A few important things to observe are that $B(N)$ takes only integral values, it is centered and symmetric around 0, so $B(N)$ is identically distributed to $-B(N)$. Its range is $[-N, N]$.
Let us denote $\Pr[B(N) = 0]$ by p_0. It is a well-known fact that $p_0 = \frac{\binom{2N}{N}}{4^N} = \Theta\left(\frac{1}{N^{1/2}}\right)$. The following lemma tells us that the probability of a binomial distribution taking any value close to its mean is significantly high.

Lemma 5.2.6. Let W be a binomial random variable distributed according to $B(N)$. Let p_0 denote $\Pr[W = 0]$. Then for all $j \in [-N, N]$,

$$p_0 - O\left(\frac{j^2}{N^{3/2}}\right) \leq \Pr[W = j] \leq p_0.$$

Proof. Note that for $|j| \geq N/2$ (in fact, for all $|j| = \omega(N^{3/4})$), the bound to be proved is trivial. Thus we assume $|j| < N/2$.

$$\begin{aligned}
\Pr[W = j-1] - \Pr[W = j] &= \left[\binom{2N}{N+j-1} - \binom{2N}{N+j}\right] \cdot \frac{1}{2^{2N}} \\
&= \left[\frac{(2N)!}{(N+j-1)!(N-j+1)!} - \frac{(2N)!}{(N+j)!(N-j)!}\right] \cdot \frac{1}{2^{2N}} \\
&= \frac{(2N)!}{(N+j-1)!(N-j)!} \cdot \frac{2j-1}{(N-j+1)(N+j)} \cdot \frac{1}{2^{2N}} \\
&= \binom{2N}{N+j} \cdot \frac{2j-1}{N-j+1} \cdot \frac{1}{2^{2N}} \\
&\leq \binom{2N}{N} \cdot \frac{1}{2^{2N}} \cdot \frac{2j}{N-j}.
\end{aligned}$$

since the middle binomial coefficient is the maximum. Thus, we have $\forall i, |i| \leq j$,

$$\Pr[W = i-1] - \Pr[W = i] \leq \binom{2N}{N} \frac{2j}{N-j} \cdot \frac{1}{2^{2N}}.$$

Since $\frac{\binom{2N}{N}}{4^N} = \Theta\left(\frac{1}{N^{1/2}}\right)$,

$$\begin{aligned}
\Pr[W = 0] - \Pr[W = j] &\leq \sum_{i=1}^{j} |\Pr[W = i-1] - \Pr[W = i]| \leq \frac{2j^2}{N-j} \cdot O\left(\frac{1}{N^{1/2}}\right) \\
&\leq \frac{2 \cdot 2j^2}{N} \cdot O\left(\frac{1}{N^{1/2}}\right) && \text{Since } |j| \leq N/2 \\
&\leq O\left(\frac{j^2}{N^{3/2}}\right). && \square
\end{aligned}$$

5.3 A Discrepancy Upper Bound for the Multi-Party GHR Function

In this section, we prove essentially a $2^{-\sqrt{N}/4^k}$ upper bound on the discrepancy of the GHR_k^N function where the first player gets N input bits. This gives us a $2^{-n^{\Omega(1)}}$ upper bound on the discrepancy if $k \leq \epsilon \log(N)$ for any constant $0 < \epsilon < 1/4$.

Goldmann et al. [GHR92] showed that when $k = 2$, if there is a low cost one-way protocol for GHR_2^N, then it must have low advantage. Sherstov [She08] noted that the same proof technique can be adapted to prove an upper bound on the discrepancy on GHR_2^N. We generalize this for higher k. In particular, we show

Theorem 5.3.1. For any $k \geq 1$,

$$\mathsf{disc}(\mathsf{GHR}_k^N) = O\left(\frac{(8e)^k N^{1/4}}{2^{\sqrt{N}/4^k} \cdot 2^{k/2}}\right),$$

where GHR_k^N is defined as in Definition 5.1.1, and N is the maximum number of bits a player gets (in this case the first player).

Proof of Theorem 5.1.2. It follows directly from Theorem 5.3.1 and Lemma 5.2.2. □

Proof of Corollary 5.1.3. From Theorem 5.1.2, it follows that for all $1 \leq k \leq \delta \cdot \log n$, the function GHR_k^N is not in PP_{k+1} for any constant $0 < \delta < 1/4$. The upper bound follows the same proof as that of Claim 2.3.10. □

Recall that $N = 2n^2 4^k$. The proof technique of Theorem 5.3.1 is inspired from that of Goldmann et al. [GHR92].

Proof of Theorem 5.3.1. Let $A_j = \frac{1}{2}\sum_{i=0}^{n-1} 2^i (x_{i,2j} + x_{i,2j+1})$. It is easy to see that A_j can take any integer value in $[-2^n + 1, 2^n - 1]$. Let μ_X be a distribution on the x's that make the A_j's independent and binomially distributed according to $B(2^n - 1)$ as defined in Definition 5.2.5. Such a distribution exists because each A_j depends on a disjoint set of variables. Let $\mathcal{U}$ be the uniform distribution on $\{-1, 1\}^{n4^k}$. We choose a tuple $(x, y_1, \ldots, y_k)$ by first picking $y_i \sim \mathcal{U}$ independently for each i, and then picking $x \sim \mu_X$ under the condition that $|P(x, y_1, \ldots, y_k)| = 2^k$. Let us define μ to be the distribution obtained by this sampling procedure.

We will now show an upper bound on the discrepancy of GHR_k^N under the distribution μ. Let S denote the characteristic function (0-1 valued) of a cylinder intersection. By Definition 5.2.1, the discrepancy of GHR_k^N according to μ is

$$\mathsf{disc}_\mu(\mathsf{GHR}_k^N) = \max_S \left|\mathbb{E}_\mu \left[\mathsf{GHR}_k^N(x, y_1, \ldots, y_k) S(x, y_1, \ldots, y_k)\right]\right|. \tag{5.3}$$

The following lemma will enable us to switch to working with a product distribution on the inputs, for which we have convenient techniques for proving discrepancy upper bounds via Lemma 5.2.3.

Lemma 5.3.2. For $\mu_X, \mathcal{U}$ as defined above,

$$\Pr_{\mu_X \times \mathcal{U}^k}[|P(x, y_1, \ldots, y_k)| = 2^k] \geq \Omega\left(\frac{1}{\sqrt{n} 2^{(n+2k)/2}}\right).$$

Proof. We will show that for any fixed $y_1, \ldots, y_k$, if we sample x according to μ_X, then

$P(x, y_1, \ldots, y_k)/2 = \sum_{j=0}^{n4^k-1} A_j y_{1j} \cdots y_{kj}$ is distributed according to $B(n4^k(2^n-1))$. Note that $A_j y_{1j} \cdots y_{kj}$ is always distributed according to $B(2^n-1)$, no matter what the values of $y_1, \ldots, y_k$ are. Next, observe that the sum of binomial distributions is a binomial distribution. This shows that $\sum_{j=0}^{n4^k-1} A_j y_{1j} \cdots y_{kj}$ is distributed according to $B(n4^k(2^n-1))$.

Hence, by plugging in $N = n4^k(2^n-1)$ and $j = 2^k$ in Lemma 5.2.6,

$$\begin{aligned}\Pr_{\mu_X \times \mathcal{U}^k}[|P(x, y_1, \ldots, y_k)| = 2^k] &\geq \Omega\left(\frac{1}{(n4^k(2^n-1))^{1/2}}\right) - O\left(\frac{4^k}{(n4^k(2^n-1))^{3/2}}\right) \\ &= \Omega\left(\frac{1}{\sqrt{n}2^{(n+2k)/2}}\right).\end{aligned}$$

We can discard the second term since it equals $O\left(\frac{1}{(4^k)^{1/2} \cdot (n(2^n-1)^{3/2})}\right)$, and is dominated by the first term. □

Let us now recall the law of total expectation.

Fact 5.3.3 (Law of total expectation)**.** For any probability space $(\Omega, \mathcal{F}, \nu)$, any event $E \in \mathcal{F}$, and any random variable Z, the following equality holds.

$$\mathbb{E}_\nu[Z] = \mathbb{E}_\nu[Z \mid E] \cdot \Pr_\nu[E] + \mathbb{E}_\nu[Z \mid \bar{E}] \cdot (1 - \Pr_\nu[E]).$$

Define a function q by

$$q(x, y_1, \ldots, y_k) = \begin{cases} P(x, y_1, \ldots, y_k)/2^k & \text{if } |P(x, y_1, \ldots y_k)| = 2^k \\ 0 & \text{otherwise.} \end{cases}$$

This means that if $(x, y_1, \ldots, y_k)$ is chosen according to the distribution $\mu_X \times \mathcal{U}^k$, then $q(x, y_1, \ldots, y_k) = \mathsf{GHR}_k^N(x, y_1, \ldots, y_k)$ on the support of μ, and 0 otherwise. For any cylinder intersection S, let Z denote the random variable $q(x, y_1, \ldots, y_k) \cdot S(x, y_1, \ldots, y_k)$, let E denote the event $|P(x, y_1, \ldots y_k)| = 2^k$. Using Fact 5.3.3 and the fact that $\mathbb{E}_{\mu_X \times \mathcal{U}^k}[Z \mid \bar{E}] = 0$, we obtain

$$\mathbb{E}_\mu[\mathsf{GHR}(x, y_1, \ldots, y_k) S(x, y_1, \ldots, y_k)] = \frac{\mathbb{E}_{\mu_X \times \mathcal{U}^k}[q(x, y_1, \ldots, y_k) \cdot S(x, y_1, \ldots, y_k)]}{\Pr_{\mu_X \times \mathcal{U}^k}[|P(x, y_1, \ldots y_k)| = 2^k]}. \tag{5.4}$$

Using Equation (5.3), Lemma 5.3.2 and Equation (5.4), we obtain the following.

$$\mathsf{disc}_{\mu}(\mathrm{GHR}_k^N) \leq \max_S \left|\mathbb{E}_{\mu_X \times \mathcal{U}^k}[q(x, y_1, \ldots, y_k) S(x, y_1, \ldots, y_k)]\right| \cdot O\left(\sqrt{n} 2^{\frac{n+2k}{2}}\right) \quad (5.5)$$

where S denotes a cylinder intersection. It therefore suffices to show that for all cylinder intersections S,

$$\left|\mathbb{E}_{\mu_X \times \mathcal{U}^k}[q(x, y_1, \ldots, y_k) S(x, y_1, \ldots, y_k)]\right| \leq O\left(2^{-\frac{n+2k}{2} - \epsilon n}\right) \quad (5.6)$$

for some constant $\epsilon > 0$ to give us a discrepancy upper bound of $2^{-n^{\Omega(1)}}$. For notational convenience, we may switch between the notations $\mathbb{E}_x$ and $\mathbb{E}_{x \sim \mu_X}$ from now on. Now that we have a product distribution, we can use Lemma 5.2.3,

$$\begin{aligned} &\left|\mathbb{E}_{\mu_X \times \mathcal{U}^k}[q(x, y_1, \ldots, y_k) S(x, y_1, \ldots, y_k)]\right| \\ &\qquad \leq \left(\mathbb{E}_{y_1^0, y_1^1, \ldots, y_k^0, y_k^1} \left| \mathbb{E}_x \left[\prod_{a_1, \ldots, a_k \in \{0,1\}} q(x, y_1^{a_1}, \ldots, y_k^{a_k}) \right] \right| \right)^{1/2^k}. \end{aligned} \quad (5.7)$$

We will now show an upper bound on the RHS of the above equation by splitting the outer expectation into two terms, the first of which has low probability. We will require certain properties of Hadamard matrices to give an upper bound on the second term. Let $\beta \in \{0,1\}^k$. Define 2^k subsets of indices as $I_\beta = \{j \in [n4^k] : \forall i \in [k],\ (y_i^0)_j = (-1)^{\beta_i} \cdot (y_i^1)_j\}$. Note that $\{I_\beta : \beta \in \{0,1\}^k\}$ forms a partition of the indices. Since our distributions on y_i^0, y_i^1's are uniform and independent, each I_β is empty with equal probability. An easy counting argument tells us that the probability of I_β being empty is $\left(\frac{2^k - 1}{2^k}\right)^{n4^k}$. By a union bound, the probability that any one of them is empty is at most $2^k \cdot \left(\frac{2^k - 1}{2^k}\right)^{n4^k}$. We have the following.

$$\begin{aligned} &\left(\mathbb{E}_{y_1^0, y_1^1, \ldots, y_k^0, y_k^1} \left| \mathbb{E}_x \left[\prod_{a_1, \ldots, a_k \in \{0,1\}} q(x, y_1^{a_1}, \ldots, y_k^{a_k}) \right] \right| \right)^{1/2^k} \\ &\qquad \leq \left(2^k \left(1 - \frac{1}{2^k}\right)^{n4^k} + Z\right)^{1/2^k} \end{aligned}$$

where $Z = \mathbb{E}_{y_1^0, y_1^1, \ldots, y_k^0, y_k^1 : \forall \beta, I_\beta \neq \emptyset} \left| \mathbb{E}_x \prod_{a_1, \ldots, a_k \in \{0,1\}} q(x, y_1^{a_1}, \ldots, y_k^{a_k}) \right|$

Claim 5.3.4. For all $y_1^0, \ldots, y_k^0, y_1^1, \ldots, y_k^1$ such that I_β is non-empty for each $\beta \in \{0,1\}^k$, we have

$$\left| \mathbb{E}_x \left[\prod_{a_1,\ldots,a_k \in \{0,1\}} q(x, y_1^{a_1}, \ldots, y_k^{a_k}) \right] \right| \leq O\left(2^{k \log(e) 2^k} \cdot 2^{2^k} \frac{1}{(2^{n/2})^{2^k - 1}} \cdot \frac{2^{(k+1)2^{k+1}}}{2^{3n/2}} \right)$$

Let us assume the claim to be true for now. We have from Equation (5.5) that

$$\begin{aligned}
\mathsf{disc}_\mu(\mathsf{GHR}_k^N) &\leq \left| \mathbb{E}_{\mu_X \times \mathcal{U}^k}[q(x, y_1, \ldots, y_k) S(x, y_1, \ldots, y_k)] \right| \; O\left(\sqrt{n} 2^{\frac{n+2k}{2}} \right) \\
&\leq \left(2^k \left(1 - \frac{1}{2^k} \right)^{n4^k} + O\left(2^{k \log(e) 2^k} \cdot 2^{2^k} \frac{1}{(2^{n/2})^{2^k - 1}} \cdot \frac{2^{(k+1)2^{k+1}}}{2^{3n/2}} \right) \right)^{1/2^k} \\
&\cdot O\left(\sqrt{n} 2^{\frac{n+2k}{2}} \right) \\
&\leq \left[2^{k/2^k} \left(1 - \frac{1}{2^k} \right)^{n2^k} + O\left(\frac{(4e)^k}{(2^{\frac{n}{2}})^{1 - \frac{1}{2^k}} \cdot 2^{\frac{3n}{2} \cdot \frac{1}{2^k}}} \right) \right] O\left(\sqrt{n} 2^{\frac{n+2k}{2}} \right) \\
&\leq O\left(\left(e^{-1/2^k} \right)^{n2^k} \cdot 2^{n/2 + k + k/2^k} \cdot \sqrt{n} + \frac{(8e)^k \sqrt{n}}{2^{(\frac{3n}{2} - \frac{n}{2}) \cdot \frac{1}{2^k}}} \right) \\
&\qquad \text{Using the fact that } \left(1 - \tfrac{1}{\gamma} \right) < e^{-1/\gamma} \\
&= O\left(e^{-n} \cdot 2^{n/2 + k + k/2^k} \cdot \sqrt{n} + \frac{(8e)^k \sqrt{n}}{2^{n/2^k}} \right) = O\left(\frac{(8e)^k \sqrt{n}}{2^{n/2^k}} \right) \\
&\qquad \text{Assuming } k < n/3 \\
&= O\left(\frac{(8e)^k N^{1/4}}{2^{\sqrt{N}/4^k} \cdot 2^{k/2}} \right) \qquad \text{Recall that } N = 2n^2 4^k
\end{aligned}$$

which proves Theorem 5.3.1. □

Now it only remains to prove Claim 5.3.4.

5.3.1 Proof of Claim 5.3.4

Recall that we need to show the following. For all $y_1^0, \ldots, y_k^0, y_1^1, \ldots, y_k^1$ such that I_β is non-empty for each β, we want

$$\left| \mathbb{E}_x \left[\prod_{a_1,\ldots,a_k \in \{0,1\}} q(x, y_1^{a_1}, \ldots, y_k^{a_k}) \right] \right| \leq O\left(2^{k \log(e) 2^k} \cdot 2^{2^k} \frac{1}{(2^{n/2})^{2^k - 1}} \cdot \frac{2^{(k+1)2^{k+1}}}{2^{3n/2}} \right).$$

Fix any such $y_1^0, \ldots, y_k^0, y_1^1, \ldots, y_k^1$. Note that the LHS of the above equation is

$$\left| \Pr_x \left[\prod_{a_1,\ldots,a_k \in \{0,1\}} q(x, y_1^{a_1}, \ldots, y_k^{a_k}) = 1 \right] - \Pr_x \left[\prod_{a_1,\ldots,a_k \in \{0,1\}} q(x, y_1^{a_1}, \ldots, y_k^{a_k}) = -1 \right] \right|.$$

For convenience, for all $a \in \{0,1\}^k$ let us denote $P(x, y_1^{a_1}, \ldots, y_k^{a_k})$ by P_a and let S_a denote $P_a/2$. By the definition of q, we have

$$\begin{aligned} \left| \mathbb{E}_x \left[\prod_{a \in \{0,1\}^k} q(x, y_1^{a_1}, \ldots, y_k^{a_k}) \right] \right| &= \left| \Pr_x \left[\prod_{a \in \{0,1\}^k} \frac{P_a}{2^k} = 1 \right] - \Pr_x \left[\prod_{a \in \{0,1\}^k} \frac{P_a}{2^k} = -1 \right] \right| \\ &= \left| \Pr_x \left[\prod_{a \in \{0,1\}^k} S_a = 2^{(k-1)2^k} \right] - \Pr_x \left[\prod_{a \in \{0,1\}^k} S_a = -2^{(k-1)2^k} \right] \right|. \quad (5.8) \end{aligned}$$

Let $W_\beta = \sum_{j \in I_\beta} A_j (y_1^0)_j \ldots (y_k^0)_j$. It will be useful to note here that W_β only takes integral values. We will use this fact crucially later. Let $\mathbf{p_k}$ denote the $2^k \times 1$ column vector whose elements are indexed by $a = (a_1, \ldots, a_k) \in \{0,1\}^k$, and the a'th element of $\mathbf{p_k}$ is $P(x, y_1^{a_1}, \ldots, y_k^{a_k})$.

Similarly define column vectors $\mathbf{s_k}$ ($\mathbf{w_k}$, respectively) whose a'th entries are S_a (W_a, respectively) for all $a \in \{0,1\}^k$. Although $\mathbf{p_k}, \mathbf{s_k}$ and $\mathbf{w_k}$ depend on $x, y_1^0, \ldots, y_k^0, y_1^1, \ldots, y_k^1$, we do not make this dependence explicit in the following discussion in order to avoid clutter.

Claim 5.3.5. The following holds true for all k, and all $x, y_1^0, \ldots, y_k^0, y_1^1, \ldots, y_k^1$.

$$\mathbf{p_k} = 2\mathbf{s_k} = 2\mathbf{H_k} \cdot \mathbf{w_k} \quad (5.9)$$

where $\mathbf{H_k}$ is a $2^k \times 2^k$ Hadamard matrix defined[1] as $\mathbf{H_k} = \begin{bmatrix} \mathbf{H_{k-1}} & \mathbf{H_{k-1}} \\ \mathbf{H_{k-1}} & -\mathbf{H_{k-1}} \end{bmatrix}$ and $\mathbf{H_0} = \begin{bmatrix} 1 \end{bmatrix}$.

Let us first state a well-known property of $\mathbf{H_k}$.

Fact 5.3.6. Let $\mathbf{H_k}$ be as defined above. Then, $(\mathbf{H_k})_{ij} = (-1)^{\langle i,j \rangle}$ for all $i, j \in \{0,1\}^k$.

In other words, $\mathbf{H_k}$ is the communication matrix of IP. Let us now prove Claim 5.3.5.

[1] This is the Sylvester construction of Hadamard matrices.

Proof of Claim 5.3.5. Let $a \in \{0,1\}^k$, $P_a = 2\sum_{j=1}^{n4^k} A_j(y_1^{a_1})_j \cdots (y_k^{a_k})_j$ and $W_\beta = \sum_{j \in I_\beta} A_j(y_1^0)_j \cdots (y_k^0)_j$. Say $j \in I_\beta$ where $\beta \in \{0,1\}^k$. Note that $(y_i^{a_i})_j = -1 \cdot (y_i^0)_j$ iff $a_i = 1, \beta_i = 1$. Hence, we have $(y_1^{a_1})_j \cdots (y_k^{a_k})_j = (-1)^{\left(\sum_i a_i \cdot \beta_i\right)}(y_1^0)_j \cdots (y_k^0)_j = (-1)^{\langle a,\beta\rangle}(y_1^0)_j \cdots (y_k^0)_j$.

$$\begin{aligned} P_a &= 2\sum_{j=1}^{n4^k} A_j(y_1^{a_1})_j \cdots (y_k^{a_k})_j = 2\left(\sum_{\beta\in\{0,1\}^k}\sum_{j\in I_\beta}(-1)^{\langle a,\beta\rangle} A_j(y_1^0)_j \cdots (y_k^0)_j\right) \\ &= 2\left(\sum_{\beta\in\{0,1\}^k}(-1)^{\langle a,\beta\rangle} W_\beta\right) \\ &= 2(\mathbf{H_k})_a \cdot \mathbf{w_k} \end{aligned}$$

where $(\mathbf{H_k})_a$ denotes the a'th row of $\mathbf{H_k}$. Thus, $\mathbf{p_k} = 2\mathbf{s_k} = 2\mathbf{H_k} \cdot \mathbf{w_k}$. □

On Integral Solutions to Hadamard Constraints

In the remainder of this section, we shall refer to an integral assignment to $\mathbf{w_k}$ as a *valid integral assignment* if it satisfies Equation (5.9) for some setting of $x, y_1^0, \ldots, y_k^0, y_1^1, \ldots, y_k^1$. The conditions on $\mathbf{s_k}$ will be explicitly stated in each usage.

First, we prove that the number of valid integral assignments to $\mathbf{w_k}$ satisfying $\prod_{a\in\{0,1\}^k} S_a = 2^{(k-1)2^k}$ is equal to the number of valid integral assignments to $\mathbf{w_k}$ satisfying $\prod_{a\in\{0,1\}^k} S_a = -2^{(k-1)2^k}$. Moreover, we show that the total number of such valid integral assignments is small, and the values of $|W_a|$ are not too large in any such valid assignment. Recall from Equation (5.8) that for all $y_1^0, \ldots, y_k^0, y_1^1, \ldots, y_k^1$ such that I_β is non-empty for each β, we have

$$\left|\mathbb{E}_x\left[\prod_{a\in\{0,1\}^k} q(x, y_1^{a_1}, \ldots, y_k^{a_k})\right]\right| = \left|\Pr_x\left[\prod_{a\in\{0,1\}^k} S_a = 2^{(k-1)2^k}\right] - \Pr_x\left[\prod_{a\in\{0,1\}^k} S_a = -2^{(k-1)2^k}\right]\right|.$$

Thus, we can pair valid "positive" and "negative" assignments. Higher-order terms in the difference $\left|\Pr_x\left[\prod_{a\in\{0,1\}^k} S_a = 2^{(k-1)2^k}\right] - \Pr_x\left[\prod_{a\in\{0,1\}^k} S_a = -2^{(k-1)2^k}\right]\right|$ cancel out. We require the following well-known property of Hadamard matrices.

Fact 5.3.7. Let $\mathbf{H}$ be an $N \times N$ Hadamard matrix. Then, $\mathbf{H}$ is invertible, and $\mathbf{H}^{-1} = \frac{1}{N}\mathbf{H}$.

Claim 5.3.8. The number of valid integral assignments to $\mathbf{w_k}$ that satisfy $\prod_{a\in\{0,1\}^k} S_a = +2^{(k-1)2^k}$ equals the number of valid integral assignments that satisfy $\prod_{a\in\{0,1\}^k} S_a = -2^{(k-1)2^k}$.

Proof. The constraints we have are $\mathbf{H_k}\cdot\mathbf{w_k} = \mathbf{s_k}$. Since W_a is integral for all a, and $\mathbf{H_k}$ is a ±1 matrix, this implies that the S_a's are integral as well. Thus, using Fact 5.3.7 we get $\frac{1}{2^k}\mathbf{H_k}\cdot\mathbf{s_k} = \mathbf{w_k}$, or $\mathbf{H_k}\cdot\frac{\mathbf{s_k}}{2^k} = \mathbf{w_k}$. Let us consider two cases, one where $\forall a \in \{0,1\}^k, \left|\frac{S_a}{2^k}\right| = 1/2$, and another where there exists an a such that $\left|\frac{S_a}{2^k}\right| \neq 1/2$.

- Let us assume $\forall a, \left|\frac{S_a}{2^k}\right| = 1/2$. We show something slightly stronger, namely that every setting of each $\frac{S_a}{2^k}$ to $\pm1/2$ gives us a valid assignment to the W_a's. Since $\mathbf{H_k}$ is a ±1 matrix of even dimension, the parity of the number of appearances of $+1/2$ equals the parity of number of appearances of $-1/2$ in the sum $(\mathbf{H_k})_R\cdot\frac{\mathbf{s_k}}{2^k}$, where $(\mathbf{H_k})_R$ is the R'th row of $\mathbf{H_k}$. This holds for every row R. Thus, W_R is always an integer. This means the number of valid positive assignments equals the number of valid negative assignments in this case.

- The absolute value of S_a must equal a power of 2 for each a since the product of them is a power of 2. If there exists an S_a whose value is not $\pm2^{k-1}$, then there must exist an S_b (consider the last such one) which is a multiple of 2^k since $\prod_{a\in\{0,1\}^k} S_a = \pm2^{(k-1)2^k}$. Since $S_b/2^k$ is an integer, and we had a valid integral assignment to $\mathbf{w_k}$, flipping the sign of S_b can change the value of any W_c to $W_c \pm 2\cdot S_b/2^k$, which remains an integer. This is a bijection between valid positive and negative assignments.

□

Lemma 5.3.9. The number of valid integral assignments to $\mathbf{w_k}$ satisfying $\prod_{a\in\{0,1\}^k} S_a = \pm2^{(k-1)2^k}$ is at most $2^{k\log(e)2^k}$.

We use the following standard fact about binomial coefficients.

Fact 5.3.10. For all $n\in\mathbb{N}$ and for all $k\in[n]$, $\binom{n}{k} \leq \left(\frac{n\cdot e}{k}\right)^k$.

Proof of Lemma 5.3.9. Suppose $\prod_{a\in\{0,1\}^k} S_a = \pm2^{(k-1)2^k}$. This means we have to distribute $(k-1)2^k$ powers of 2 among $2^k S_a$'s (which are all integers). The total number of ways to do this equals the number of non-negative integer solutions to $m_1+\cdots+m_{2^k} = (k-1)2^k$, which equals $\binom{k2^k-1}{(k-1)2^k}$. Note that $\binom{k2^k-1}{(k-1)2^k} = \binom{k2^k}{(k-1)2^k} \leq \left(\frac{k2^k\cdot e}{(k-1)2^k}\right)^{(k-1)2^k}$, where the last inequality follows by Fact 5.3.10. Now we use the fact

that $1 + x \leq e^x$ and conclude that $\left(\frac{k2^k \cdot e}{(k-1)2^k}\right)^{(k-1)2^k}$ is bounded above by e^{k2^k}, which equals $2^{k\log(e)2^k}$. Each of these can give at most one integral assignment to the W_a's because the system of constraints is linearly independent. □

We now state an upper bound on the value of $|W_a|$ in every integral assignment.

Lemma 5.3.11. For all $a \in \{0,1\}^k, |W_a| \leq 2^{(k+1)2^k}$ for any valid integral assignment to $\mathbf{w_k}$ satisfying $\prod_{a\in\{0,1\}^k} S_a = \pm 2^{(k-1)2^k}$.

Proof. First note that for each a, $|W_a| \leq \sum_{a\in\{0,1\}^k} \frac{|S_a|}{2^k}$ since $\mathbf{H_k} \cdot \mathbf{s_k} = \mathbf{w_k}$. We show that $\sum_{a\in\{0,1\}^k} |S_a|$ is at most 2^{k2^k}. Suppose not. By a simple averaging argument, there must be a b such that $|S_b| > \frac{2^{k2^k}}{2^k}$, which is $2^{k(2^k-1)}$, which is at least $2^{(k-1)2^k}$ if $k \geq 1$. But this is not possible since $\prod_{a\in\{0,1\}^k} S_a = \pm 2^{(k-1)2^k}$ and the S_a's are integers. □

Using Properties of the Binomial Distribution

Recall from Equation (5.8) that for all $y_1^0, \ldots, y_k^0, y_1^1, \ldots, y_k^1$ for which I_β is non-empty for each β, we want to show an upper bound on the quantity $\left|\Pr_x\left[\prod_{a\in\{0,1\}^k} S_a = 2^{(k-1)2^k}\right] - \Pr_x\left[\prod_{a\in\{0,1\}^k} S_a = -2^{(k-1)2^k}\right]\right|$. Recall that we defined $W_\beta = \sum_{j\in I_\beta} A_j (y_1^0)_j \ldots (y_k^0)_j$. For any $\beta \in \{0,1\}^k$, note that W_β is always distributed according to $B(c_\beta(2^n - 1))$, where $c_\beta = |I_\beta| \geq 1$. We can prove this in a manner similar to that in the proof of Lemma 5.3.2. In Claim 5.3.8, we showed that the number of valid integral assignments to $\mathbf{w_k}$ such that $\prod_{a\in\{0,1\}^k} S_a = 2^{(k-1)2^k}$ equals the number of integral assignments such that $\prod_{a\in\{0,1\}^k} S_a = -2^{(k-1)2^k}$. Note that if the assignment to $\mathbf{w_k}$ is not integral, then it has probability 0, since for each a, W_a takes only integral values. Let us call an assignment to $\mathbf{w_k}$ *positive* if the corresponding value of $\prod_{a\in\{0,1\}^k} S_a = 2^{(k-1)2^k}$, and *negative* if the value of $\prod_{a\in\{0,1\}^k} S_a = -2^{(k-1)2^k}$. Arbitrarily form a matching, denoted by $\mathcal{M}$, between the positive and negative assignments. We will bound the difference of probabilities of each match.

$$\begin{aligned}
&\left|\Pr_x\left[\prod_{a\in\{0,1\}^k} S_a = 2^{(k-1)2^k}\right] - \Pr_x\left[\prod_{a\in\{0,1\}^k} S_a = -2^{(k-1)2^k}\right]\right| \\
&\leq \sum_{(w,w')\in\mathcal{M}} \left|\Pr_x[\mathbf{w_k} = w] - \Pr_x[\mathbf{w_k} = w']\right|.
\end{aligned}$$

where $w = (w_a)_{a\in\{0,1\}^k}$ is a valid positive assignment and $w' = (w'_a)_{a\in\{0,1\}^k}$ is the valid negative assignment that is the unique match of w according to $\mathcal{M}$. The term

$\Pr_x[\mathbf{w_k} = w]$ is equal to $\Pr_x[\bigwedge_{a\in\{0,1\}^k} W_a = w_a]$. In Lemma 5.3.11 we showed that for each β, the absolute value of W_β in any integral assignment can be at most $2^{(k+1)2^k}$. Each W_β is distributed according to $B(c_\beta(2^n-1)), c_\beta > 0$, since $\forall \beta \in \{0,1\}^k, |I_\beta| > 0$. For a particular positive assignment w, negative assignment w' and any $y_1^0, \ldots, y_k^0, y_1^1, \ldots, y_k^1$ such that I_β is non-empty for each β,

$$\left|\Pr_x[\mathbf{w_k} = w] - \Pr_x[\mathbf{w_k} = w']\right| = \left|\Pr\left[\bigwedge_{a\in\{0,1\}^k} W_a = w_a\right] - \Pr\left[\bigwedge_{a\in\{0,1\}^k} W_a = w'_a\right]\right|.$$

By plugging in $N = c_\beta(2^n-1)$ and $j = 2^{(k+1)2^k}$ in Lemma 5.2.6, we obtain $p_0 \geq \Pr_x[W_\beta = w_\beta] \geq p_0 - O\left(\frac{2^{(k+1)2^{k+1}}}{2^{3n/2}}\right)$, where $p_0 = \Pr[W_\beta = 0] = O\left(\frac{1}{2^{n/2}}\right)$. For convenience in calculations, let us say $\Pr_x[W_\beta = w_\beta] \in \left(p_0 \pm O\left(\frac{2^{(k+1)2^{k+1}}}{2^{3n/2}}\right)\right)$. Recall that the W_β's are independent of each other since they depend on disjoint variables. Thus,

$$\begin{aligned}
&\left|\Pr[\textstyle\bigwedge_{a\in\{0,1\}^k} W_a = w_a] - \Pr[\bigwedge_{a\in\{0,1\}^k} W_a = w'_a]\right| \\
&\leq \left|\left(p_0 \pm O\left(\frac{2^{(k+1)2^{k+1}}}{2^{3n/2}}\right)\right)^{2^k} - \left(p_0 \pm O\left(\frac{2^{(k+1)2^{k+1}}}{2^{3n/2}}\right)\right)^{2^k}\right| \leq \frac{2\cdot 2^{2^k}}{(2^{n/2})^{2^k-1}} \cdot \frac{2^{(k+1)2^{k+1}}}{2^{3n/2}}.
\end{aligned}$$

The last inequality holds because the highest-order term after binomially expanding both terms is $(p_0)^{2^k}$, which cancel each other. Note that the sum of the binomial coefficients is 2^{2^k}, and each term after the first is at most $\frac{1}{(2^{n/2})^{2^k-1}} \cdot \frac{2^{(k+1)2^{k+1}}}{2^{3n/2}}$. Thus, the sum of the remaining terms can be bounded above by $2^{2^k}\frac{1}{(2^{n/2})^{2^k-1}} \cdot \frac{2^{(k+1)2^{k+1}}}{2^{3n/2}}$. By Lemma 5.3.9, the number of assignments is at most $2^{k\log(e)2^k}$. Thus,

$$\begin{aligned}
&\left|\Pr_x\left[\textstyle\prod_{a\in\{0,1\}^k} S_a = 2^{(k-1)2^k}\right] - \Pr_x\left[\prod_{a\in\{0,1\}^k} S_a = -2^{(k-1)2^k}\right]\right| \\
&\leq \sum_{(w,w')\in\mathcal{M}} \left|\Pr_x[\mathbf{w_k} = w] - \Pr_x[\mathbf{w_k} = w']\right| \leq 2^{k\log(e)2^k} \cdot 2^{2^k} \frac{1}{(2^{n/2})^{2^k-1}} \cdot \frac{2^{(k+1)2^{k+1}}}{2^{3n/2}}
\end{aligned}$$

which proves Claim 5.3.4. Using Equation (5.5), this proves Theorem 5.3.1.

5.4 Circuit Lower Bounds

In this section, we will show how we obtain lower bounds on the size of depth-3 circuits of the type $\mathsf{MAJ} \circ \mathsf{THR} \circ \mathsf{ANY}_k$ computing the GHR_k^N function. Recall that

GHR_k^N can be computed by linear-size $\mathsf{THR} \circ \mathsf{PAR}_{k+1}$ circuits. First let us state the results that were known prior to this work.

Lemma 5.4.1 (Folklore)**.** Any function f computable by size s circuits of the type $\mathsf{SYM} \circ \mathsf{ANY}_k$ has a deterministic simultaneous $(k+1)$-player protocol of cost $O(k \log(s))$ for any partitioning of the input bits.

Proof. Since each of the bottom layer gates has fan-in at most k, there must exist a player who sees all the inputs to it. The protocol decides beforehand which gate 'belongs' to which player. All players simultaneously broadcast their contribution to the top SYM gate using at most $\log(s)$ bits each. □

A consequence of this is an upper bound for randomized protocols for depth-3 circuits, which may be found in [Cha07] for example, and is stated below without proof.

Lemma 5.4.2 (Folklore)**.** Given any function f computable by size s circuits of the type $\mathsf{MAJ} \circ \mathsf{SYM} \circ \mathsf{ANY}_k$, and any partition of the input bits, there exists a public coin $(k+1)$-player randomized protocol computing f with advantage $\Omega(1/s)$ and cost $O(k \log(s))$. In other words, $\mathsf{PP}_{k+1}(f) = O(k \log s)$.

A similar upper bound is as follows.

Lemma 5.4.3 (Folklore)**.** Given any function f computable by size s circuits of the type $\mathsf{MAJ} \circ \mathsf{XOR} \circ \mathsf{ANY}_k \circ \mathsf{XOR}$ and for any partition of the input bits, $\mathsf{PP}_{k+1}(f) = O(k \log s)$.

Proof. The proof follows along the same lines as that of Lemma 5.4.2 along with the observation that each $\mathsf{XOR} \circ \mathsf{ANY}_k \circ \mathsf{XOR}$ sub-circuit has $\mathbb{F}_2$-degree at most k. □

Let us now prove Theorem 5.1.4.

Proof. Suppose GHR_k^N could be computed by $\mathsf{MAJ} \circ \mathsf{SYM} \circ \mathsf{ANY}_k$ circuits of size s. Using the protocol mentioned in Lemma 5.4.2, the cost of the protocol is $O(k \log(s))$ and advantage $\Omega(1/s)$. Using Theorem 5.1.2, $O(k \log(s) + \log(s)) \geq \Omega\left(\frac{\sqrt{N}}{4^k} - \log(N) - k\right)$, which gives $\log(s) \geq \Omega\left(\frac{\sqrt{N}}{k4^k} - \frac{\log(N)}{k} - 1\right)$. Thus, $s \geq 2^{\Omega\left(\frac{\sqrt{N}}{k4^k} - \frac{\log(N)}{k} - 1\right)} \geq 2^{\Omega\left(\frac{\sqrt{N}}{k4^k} - \frac{\log(N)}{k}\right)}$. □

By definition, polynomial-size $\mathsf{MAJ} \circ \mathsf{MAJ}$ circuits can be simulated by polynomial-size $\mathsf{MAJ} \circ \mathsf{SYM}$ circuits. Also, Goldmann et al. [GHR92] (Theorem 26) showed that $\mathsf{MAJ} \circ \mathsf{THR}$ circuits can be simulated by $\mathsf{MAJ} \circ \mathsf{MAJ}$ circuits with a polynomial blowup.

More precisely, a $\mathsf{MAJ} \circ \mathsf{THR}$ circuit of size s can be simulated by a $\mathsf{MAJ} \circ \mathsf{MAJ}$ circuit of size $s^{\alpha} \cdot N^{\beta}$ for some constants α, β where N is the input size. Hence, Corollary 5.1.5 follows by a similar proof as that of Lemma 5.4.2.

The communication lower bound of Goldmann et al. [GHR92] also implies that $\mathsf{THR} \notin \mathsf{MAJ} \circ \mathsf{XOR}$. This result of theirs may be interpreted as following: there exists a linear threshold function that cannot be represented by a low-weight signed sum of polynomials of $\mathbb{F}_2$-degree 1. As a corollary of our main result in this chapter, we generalize their result and show that there exists a linear threshold function that cannot be represented as a low-weight signed sum of polynomials of $\mathbb{F}_2$-degree $O(\log N)$. Formally, we obtain the following.

Corollary 5.4.4. There exists a constant c and a linear threshold function $f : \{-1,1\}^N \to \{-1,1\}$ such that f cannot be computed by polynomial sized $\mathsf{MAJ} \circ \mathsf{XOR} \circ \mathsf{ANY}_k$ circuits for any $k < c \log N$.

Proof. By definition, GHR_k^N can be expressed as $f \circ \mathsf{XOR}$, where $f \in \mathsf{THR}$. Thus, if f had a polynomial sized representation of the form $\mathsf{MAJ} \circ \mathsf{XOR} \circ \mathsf{ANY}_k$, then $\mathsf{GHR}_k^N \in \mathsf{MAJ} \circ \mathsf{XOR} \circ \mathsf{ANY}_k \circ \mathsf{XOR}$. By Lemma 5.4.3, this implies $\mathsf{PP}_{k+1}(\mathsf{GHR}_k^N)$ would be polylogarithmic in N, which is a contradiction by Corollary 5.1.3. □

5.5 References

The results presented in this chapter are based on joint work with Arkadev Chattopadhyay [CM16].

Chapter 6

Linear Decision Lists

In this chapter, we take a short digression and study the power of linear decision lists, which are decision lists where the queries are linear threshold functions.

As mentioned in Section 4.5.2, a natural program arising from our work is to show lower bounds against the class P^{MA}. Recall that F_n could be expressed as a decision list of exact thresholds (in particular, a decision list of Equalities), which are easy to compute in P^{MA}. Thus, a plausibly easier first step is to show lower bounds against decision lists of exact threshold functions.

A similar model to consider is the class of decision lists of *linear* threshold functions, which we denote by linear decision lists. A simple observation is that this class is a sub-class of $\mathsf{THR} \circ \mathsf{THR}$. Turán and Vatan [TV97] showed that decision lists of linear threshold functions must have large monochromatic rectangles, and thus require exponential size to compute IP. An open question they posed was how the power of linear decision lists compares with $\mathsf{MAJ} \circ \mathsf{MAJ}$. Buhrman et al. [BVdW07] and Sherstov [She11a] independently exhibited a function, efficiently computable by linear decision lists, but not by $\mathsf{MAJ} \circ \mathsf{MAJ}$.

Towards the goal of proving lower bounds against decision lists of exact threshold functions, we exploit the aforementioned lower bound technique for linear decision lists due to Turán and Vatan to show that a simple function in $\mathsf{MAJ} \circ \mathsf{MAJ}$ requires exponential sized linear decision lists to compute it. This completely answers the open question posed by Turán and Vatan.

6.1 Introduction

Decision lists are a widely studied model of computation, first introduced by Rivest [Riv87]. Recall that a decision list L of size ℓ computing a Boolean function $f \in B_n$ is

a sequence of $\ell - 1$ instructions of the form **if** $f_i(x) = a_i$ **then output** b_i **and stop**, followed by the instruction **output** $\neg b_{\ell-1}$ **and stop**. Here B_n denotes the set of all Boolean functions in n variables, each $f_i \in B_n$ is called a *query function*, and a_i and b_i are Boolean constants. If the functions f_i all belong to a function class $S \subseteq B_n$, then L is said to be an S-decision list.

Krause [Kra06] claimed that there are functions with small representation as AND-decision lists, but requiring exponential size $\mathsf{THR} \circ \mathsf{XOR}$ circuits. On the other hand, Impagliazzo and Williams [IW10] showed that a certain condition is sufficient to prove lower bounds against decision lists of rectangles.

Lower bounds against linear decision lists (and even against bounded-rank linear decision trees, a natural generalization) for IP were proved by Gröger, Turán and Vatan, in [GT91, TV97]. Subsequently, in [UT11, UT15], Uchizawa and Takimoto observed that the class of linear decision lists and linear decision trees when the weights of the linear threshold queries are bounded by a polynomial in the input length, cannot compute functions outside UPP, by noting that functions in this class can be efficiently compute in $\mathsf{THR} \circ \mathsf{MAJ}$.

We observe that the lower bound argument in [TV97] shows that functions efficiently computable by linear decision lists (with no restrictions on the weights of the queried linear threshold functions) must have large monochromatic rectangles. We then use this fact to establish a lower bound for a seemingly simple function, in $\mathsf{MAJ} \circ \mathsf{MAJ}$, thus completely resolving the open question posed by Turán and Vatan. Our main theorem regarding linear decision lists is as follows.

Theorem 6.1.1. There exists a function that can be computed by polynomial sized $\mathsf{MAJ} \circ \mathsf{MAJ}$ circuits, but any linear decision list computing it requires exponential size.

We prove this by showing that $\mathsf{MAJ} \circ \mathsf{XOR}_2$ cannot be computed efficiently by LDL's.

Theorem 6.1.2. Any linear decision list computing $\mathsf{MAJ}_n \circ \mathsf{XOR}_2$ must have size $2^{\Omega(n)}$.

It is not hard to see that $\mathsf{MAJ} \circ \mathsf{XOR}$ can be simulated by linear sized $\mathsf{MAJ} \circ \mathsf{MAJ}$ circuits with only a *linear* blow-up in size. This immediately yields Theorem 6.1.1.

Impagliazzo and Williams [IW10] demonstrated a function, implicitly computable by polynomial sized $\mathsf{MAJ} \circ \mathsf{MAJ}$ circuits, which cannot be computed by polynomial sized rectangle-decision lists. We observe that Turán and Vatan's lower bound technique against linear decision lists also applies to this function. Thus, their function

also separates linear decision lists from $\mathsf{MAJ} \circ \mathsf{MAJ}$. We elaborate on this in Section 6.4.

We now formally define some notions of interest in this chapter.

Definition 6.1.3 (Monochromatic squares)**.** Let $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be any function. A monochromatic b-square of size s is a tuple (X, Y), where $X, Y \subseteq \{0,1\}^n$ such that $|X| = |Y| = s$, and $F(x, y) = b$ for every $(x, y) \in X \times Y$. We say that (X, Y) is a monochromatic square if it is a monochromatic 0-square or 1-square.

Definition 6.1.4 (Hamming distance)**.** The (Hamming) distance between any two strings $x, y \in \{0,1\}^n$, denoted $d(x, y)$, is defined as $d(x, y) \triangleq |\{i : x_i \neq y_i\}|$. The Hamming distance between any two sets $A, B \subseteq \{0,1\}^n$, denoted $d(A, B)$, is the minimum pairwise distance; $d(A, B) = \min_{x \in A, y \in B} d(x, y)$.

Definition 6.1.5 (Hamming balls)**.** Let $c \in \{0,1\}^n$ and $k \in \{0, 1, \ldots, n\}$. A set $A \subseteq \{0,1\}^n$ is called a Hamming ball with centre c and radius k if

$$\{s \in \{0,1\}^n \mid d(s, c) \leq k - 1\} \subset A \subseteq \{s \in \{0,1\}^n \mid d(s, c) \leq k\}$$

For a set $A \subseteq \{0,1\}^n$, the boundary of A is the set $\{s \in \{0,1\}^n | d(s, A) = 1\}$. In [Har66], Harper proved a isoperimetry result: among all sets of a given size, Hamming balls have the smallest boundary set size. A simplified proof was given by Frankl and Füredi [FF81], who also stated the theorem in the equivalent form we mention below. See also the presentation in [Bol86]).

Theorem 6.1.6 (Harper's Theorem)**.** Let $A, B \subseteq \{0,1\}^n$ be non-empty sets. Then, we can find a Hamming ball A_0 with centre 0^n and a Hamming ball B_0 with centre 1^n such that $|A_0| = |A|$, $|B_0| = B$, and $d(A_0, B_0) \geq d(A, B)$.

Definition 6.1.7 (Binary Entropy)**.** The binary entropy function $\mathbb{H} : [0,1] \to [0,1]$ is defined as follows: $\mathbb{H}(p) = -p \log p - (1-p) \log(1-p)$.

Fact 6.1.8. $\mathbb{H}(1/4) < 0.82$.

6.2 Linear Decision Lists Contain Large Monochromatic Squares

In this section, we observe that the argument of Turán and Vatan from [TV97] in fact shows that all functions computable by small sized linear decision lists must contain

large monochromatic rectangles. For completeness, we first reproduce the following lemma and proof.

Lemma 6.2.1 (Lemma 2 in [TV97])**.** Let f be a linear threshold function over the variables $x_1, \ldots, x_n, y_1, \ldots, y_n$. Let $X, Y \subseteq \{0,1\}^n$, $|X| = |Y| = m$, and $v \in [m]$. Then, exactly one of the following is true.

1. There is a monochromatic 1-square (X', Y') of size v within $X \times Y$. (That is, $X' \subseteq X$ and $Y' \subseteq Y$.)

2. There is a monochromatic 0-square (X', Y') of size $m - v + 1$ within $X \times Y$.

Proof. Let M be the submatrix of M_f restricted to $X \times Y$. Let the threshold function f be given by $\mathsf{sgn}(a + \langle \alpha \cdot x \rangle + \langle \beta \cdot y \rangle)$. Reorder the rows and columns of A in decreasing order of $a + \langle \alpha \cdot x \rangle$ and $\langle \beta \cdot y \rangle$ to get the matrix B. Consider the $[v, v]$'th entry of B. If this is positive, then the top-left submatrix of B gives a 1-square of size v. Otherwise the bottom-right submatrix of B gives a 0-square of size $m - v + 1$.

The set sizes ensure that both such squares cannot simultaneously exist, since 0-squares and 1-squares must be disjoint. □

Lemma 6.2.2. Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be any function with no monochromatic square of size greater than t. Then, any LDL computing f must have size at least $2^n/t$.

Proof. (The argument below is presented for the Inner Product function in the proof of Theorem 1 in [TV97].)

Let $(L_1, a_1), (L_2, a_2), \ldots, (L_k, a_k)$ be an LDL of size k computing f. We construct, for each $i \in [k-1]$, a square $S_i = (X_i, Y_i)$ of size $2^n - it$ which is a 0-square for all L_j with $j \leq i$. We proceed by induction on i. Let $v = t + 1$.

For the base case $i = 1$, let $S_0 = (X_0, Y_0)$ be the entire $2^n \times 2^n$ matrix. Suppose S_0 has a square of size v that is a 1-square of L_1. Then everywhere in this square, f will be a_1. But f has no monochromatic square as large as $v = t + 1$. So S_0 has no square of size v that is a 1-square of L_1. By Lemma 6.2.1, S_0 must then contain a 0-square of L_1 of size $2^n - v + 1 = 2^n - t$. This establishes the base case.

For the inductive step, we already have a square S_{i-1} of size $2^n - (i-1)t$ which is a 0-square for $L_1, L_2, \ldots L_{i-1}$. Within this square, suppose L_i has a 1-square of size v. Then $f = a_i$ in this square, giving a monochromatic square of f of size $t + 1$. Since such squares do not exist, we can use Lemma 6.2.1, to conclude that S_{i-1} must contain a 0-square of L_i of size $2^n - (i-1)t - v + 1 = 2^n - it$. Since this square, say S_i, is contained in S_{i-1}, it is a 0-square for all L_j with $j \leq i$.

Thus, we have a square S_{k-1} of size $2^n - (k-1)t$ on which $L_1, L_2, \ldots, L_{k-1}$ are 0, and $L_k = 1$ because L_k is the constant function 1. Everywhere on this square, f evaluates to a_k. So S_{k-1} is a monochromatic square for f. Hence it cannot have size more than t. Thus $2^n - (k-1)t \leq t$, yielding $k \geq \frac{2^n}{t}$. □

6.3 MAJ∘XOR has no Large Monochromatic Squares

In this section, we show an upper bound and a matching tight lower bound on the size of a largest monochromatic square in the communication matrix of $\mathsf{MAJ} \circ \mathsf{XOR}$.

Lemma 6.3.1. Let $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be the function $\mathsf{MAJ}_n \circ \mathsf{XOR}$. Then, for any $b \in \{0,1\}$, M_F has a monochromatic b-square of size at least $\sum_{i=0}^{\lfloor n/4 \rfloor} \binom{n}{i}$.

Proof. Define the sets X, Y, Z as follows:

$$\begin{aligned} X = Y &= \{x \in \{0,1\}^n : |x| \leq \lfloor n/4 \rfloor\}. \\ Z &= \{x \in \{0,1\}^n : |x| \geq n - \lfloor n/4 \rfloor\}. \end{aligned}$$

Note that $F(x,y) = 0$ for all $x \in X, y \in Y$, and $F(x,z) = 1$ for all $x \in X$, $z \in Z$. Thus (X,Y) and (X,Z) are a monochromatic 0-square and 1-square respectively, each of size $\sum_{i=0}^{\lfloor n/4 \rfloor} \binom{n}{i}$. □

We now show that this bound is tight.

Theorem 6.3.2. Let $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be the function $\mathsf{MAJ}_n \circ \mathsf{XOR}$. For odd n, M_F has no monochromatic squares of size greater than $\sum_{i=0}^{\lfloor n/4 \rfloor} \binom{n}{i}$.

Proof. Suppose, to the contrary that there are sets $A, B \subseteq \{0,1\}^n$ be such that $|A| = |B| > \sum_{i<n/4} \binom{n}{i}$ and $A \times B$ is a monochromatic 1-square in M_F. By the definition of F, this implies $d(A,B) > n/2$. By Theorem 6.1.6, there exist Hamming balls A_0 around 0^n, and B_0 around 1^n such that $|A_0| = |A|, |B_0| = |B|$ and $d(A_0, B_0) \geq d(A,B)$. The size lower bound enforces that the radius of A_0 and B_0 must be greater than $\lfloor n/4 \rfloor$, and since they are centred on 0^n and 1^n, it follows that $d(A_0, B_0) < n/2$. But then $d(A,B)$ is also at most $n/2$. Hence, there exist $x \in A, y \in B$ such that $d(x,y) < n/2$, which means $F(x,y) = \mathsf{MAJ}_n \circ \mathsf{XOR}(x,y) = 0$, which contradicts our assumption. Therefore, any monochromatic 1-square in M_F has size at most $\sum_{i=0}^{\lfloor n/4 \rfloor} \binom{n}{i}$.

A similar argument, using the additional assumption that n is odd (not required in the case of 1-squares), shows the same upper bound on the size of monochromatic 0-squares. □

Now we can put things together to prove our main theorem in this chapter.

Proof of Theorem 6.1.2. Let s_n be the minimum size of an LDL computing $\mathsf{MAJ}_n \circ \mathsf{XOR}$. By Lemma 6.2.2 and Theorem 6.3.2, for all odd n,

$$\begin{aligned} s_n &\geq \frac{2^n}{\sum_{i=0}^{\lfloor n/4 \rfloor} \binom{n}{i}} \\ &\geq \frac{2^n}{2^{n \cdot H(1/4)}} && \text{using Stirling's approximation} \\ &\geq 2^{0.18n}. && \text{using Fact 6.1.8} \end{aligned}$$

□

6.4 LDL's and the Threshold Circuit Hierarchy

In this section, we see how the class of functions computable by polynomial sized LDLs fits into the low depth threshold circuit hierarchy. The reader is referred to Razborov's survey [Raz92a] for a detailed exposition on the low depth threshold circuits hierarchy.

6.5 Definitions

Definition 6.5.1 (LDL)**.** Define LDL to be the class of all functions computable by polynomial sized linear decision lists.

Definition 6.5.2 ($\widehat{\mathsf{LDL}}$)**.** Define $\widehat{\mathsf{LDL}}$ to be the class of all functions computable by polynomial sized linear decision lists where, furthermore, weights of the linear threshold queries are integers with values bounded by a polynomial in the number of variables.

Definition 6.5.3 (PL_1)**.** The class PL_1 consists of all functions $f : \{0,1\}^n \to \{0,1\}$ for which $\sum_{S \subseteq [n]} |\widehat{f}(S)| \leq \text{poly}(n)$.

Definition 6.5.4 ($\widehat{\mathsf{PT}_1}$)**.** The class $\widehat{\mathsf{PT}_1}$ consists of all functions $f : \{0,1\}^n \to \{0,1\}$ which can be represented by polynomial sized $\mathsf{MAJ} \circ \mathsf{PARITY}$ circuits.

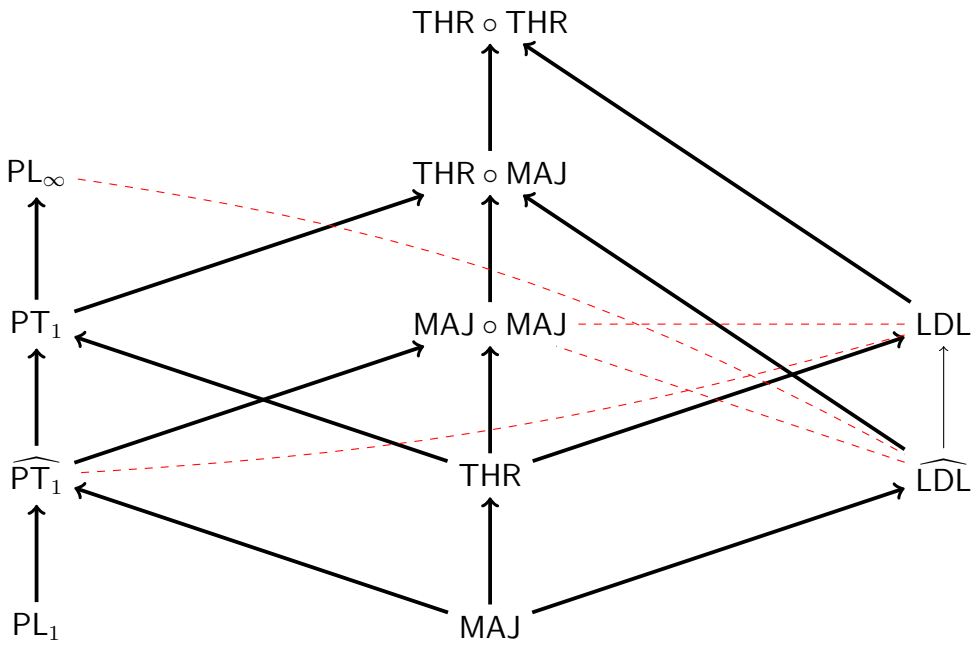

Figure 6.1: Low depth threshold circuit hierarchy

Definition 6.5.5 (PT_1)**.** The class PT_1 consists of all functions $f : \{0,1\}^n \to \{0,1\}$ which can be represented by polynomial sized $\mathsf{THR} \circ \mathsf{PARITY}$ circuits.

Definition 6.5.6 (PL_∞)**.** The class PL_∞ consists of all functions $f : \{0,1\}^n \to \{0,1\}$ for which $\max_{S \subseteq [n]} |\widehat{f}(S)| \geq \frac{1}{\mathrm{poly}(n)}$.

Figure 6.1 depicts the currently known status of low depth circuit class containments, and shows where linear decision lists fit in this hierarchy. Functions witnessing various class separations, other than those shown in this thesis, can be found in Razborov's survey [Raz92a].

A thick solid arrow from $\mathcal{C}_1$ to $\mathcal{C}_2$ denotes $\mathcal{C}_1 \subsetneq \mathcal{C}_2$, a thin solid arrow from $\mathcal{C}_1$ to $\mathcal{C}_2$ denotes $\mathcal{C}_1 \subseteq \mathcal{C}_2$, and a dashed line between $\mathcal{C}_1$ and $\mathcal{C}_2$ denotes incomparability. In the figure, we only show the newly established incomparabilities.

The leftmost column has the classes defined based on spectral representation, and the middle column has the classes based on depth-2 circuits. Concerning these classes, the picture was already completely clear: All containments shown among classes in these columns are known to be strict, and wherever no arrow connects two classes, they are known to be incomparable. Essentially this part of the figure appears in [GHR92]; a subsequent refinement is the insertion of the class $\mathsf{THR} \circ \mathsf{MAJ}$, separated

from $\mathsf{MAJ} \circ \mathsf{MAJ}$ in [GHR92], from PT_1 in [IW10] and most recently from $\mathsf{THR} \circ \mathsf{THR}$, proved earlier in this chapter.

The two classes $\widehat{\mathsf{LDL}}$ and LDL form the new column on the right. In the following section we explain their position with respect to the other two columns. However here the picture is not yet completely clear, and there are still several open questions.

6.6 New Results

By definition, $\mathsf{MAJ} \subseteq \widehat{\mathsf{LDL}}$ and $\mathsf{THR} \subseteq \mathsf{LDL}$ via lists of size 2. The parity function is known to not be in THR, and it has a simple LDL with 0-1 weights in the query functions. Thus both these containments are proper, and $\widehat{\mathsf{LDL}}$ is not contained in THR. We now observe that, implicit from prior work, $\widehat{\mathsf{LDL}}$ is not even contained in $\mathsf{MAJ} \circ \mathsf{MAJ}$.

Theorem 6.6.1.

$$\widehat{\mathsf{LDL}} \not\subseteq \mathsf{MAJ} \circ \mathsf{MAJ}.$$

Proof. Buhrman et al. [BVdW07], and independently Sherstov [She11a], showed that the PP cost of $\mathsf{OMB}_n \circ \mathsf{AND}_2$ is $\Omega(n^{1/3})$. By a result of Hajnal et al. [HMP+93], this implies that any $\mathsf{MAJ} \circ \mathsf{MAJ}$ circuit computing $\mathsf{OMB}_n \circ \mathsf{AND}_2$ requires $2^{\Omega(n^{1/3})}$ size.

Note that OMB can be computed by a linear sized decision list by querying the variables in decreasing order of their indices. Thus $\mathsf{OMB} \circ \mathsf{AND}$ can be computed by a linear sized decision list of AND's, and hence by a linear decision list with 0-1 weights. □

On the other hand, it is easily seen that $\mathsf{MAJ}_n \circ \mathsf{XOR}$ is in $\mathsf{MAJ} \circ \mathsf{MAJ}$, and even in $\widehat{\mathsf{PT}_1}$ (see for instance [Bru90]). Combining this with Theorem 6.1.2, we obtain:

Theorem 6.6.2.

$$\widehat{\mathsf{PT}_1} \not\subseteq \mathsf{LDL}.$$

Putting together these separations with the known containment $\widehat{\mathsf{PT}_1} \subseteq \mathsf{MAJ} \circ \mathsf{MAJ}$, we obtain a slew of incomparability results.

Corollary 6.6.3. For any class $A \in \{\widehat{\mathsf{LDL}}, \mathsf{LDL}\}$ and $B \in \{\widehat{\mathsf{PT}_1}, \mathsf{MAJ} \circ \mathsf{MAJ}\}$, the classes A and B are incomparable.

In particular, the classes LDL and $\mathsf{MAJ} \circ \mathsf{MAJ}$ are incomparable (as claimed in Theorem 6.1.1). This completely answers the open question posed by Turán and Vatan [TV97].

Impagliazzo and Williams [IW10] showed that the function $\mathsf{OR} \circ \mathsf{EQ}$ (also called Block-Equality) does not contain large monochromatic rectangles (in fact they showed that it does not contain large monochromatic rectangles under any product distribution). We now observe that $\mathsf{OR} \circ \mathsf{EQ} \in \mathsf{MAJ} \circ \mathsf{MAJ}$. Thus $\mathsf{OR} \circ \mathsf{EQ}$ also witnesses $\mathsf{MAJ} \circ \mathsf{MAJ} \not\subseteq \mathsf{LDL}$.

Theorem 6.6.4.

$$\mathsf{OR} \circ \mathsf{EQ} \in \mathsf{MAJ} \circ \mathsf{MAJ}.$$

Proof. First observe that $\mathsf{OR} \circ \mathsf{EQ}$ can be computed by a $\mathsf{MAJ} \circ \mathsf{EQ}$ circuit by suitably padding constants to the input. Next, note that EQ is an *exact threshold function*, that is there exist reals $a_1, \ldots, a_n, b_1, \ldots, b_n, c$ such that $\mathsf{EQ}(x, y) = 1$ iff $\sum_{i=1}^{n} a_i x_i + b_i y_i = c$. Hansen and Podolskii [HP10] showed that such functions can be efficiently simulated by $\mathsf{MAJ} \circ \mathsf{THR}$ circuits. However, we do not need the full strength of their result, and the proof that $\mathsf{MAJ} \circ \mathsf{EQ} \in \mathsf{MAJ} \circ \mathsf{THR}$ is essentially the same as the proof of Theorem 4.2.7.

Finally, Goldmann, Håstad and Razborov [GHR92] showed that $\mathsf{MAJ} \circ \mathsf{THR} = \mathsf{MAJ} \circ \mathsf{MAJ}$. Thus, $\mathsf{OR} \circ \mathsf{EQ} \in \mathsf{MAJ} \circ \mathsf{MAJ}$. □

Theorem 6.6.5.

$$\widehat{\mathsf{LDL}} \not\subseteq \mathsf{PL}_\infty.$$

Proof. It is easy to see that any symmetric function can be computed by linear sized linear decision lists where query functions are majority: the linear threshold queries can be used to determine the Hamming weight of the input, and the decision list outputs the appropriate answer at each decision.

Bruck [Bru90] showed that the *Complete Quadratic* function, which is a symmetric function, is not in PL_∞. This function yields the required separation. □

Combining Theorems 6.6.2, 6.6.5 yields more incomparability results.

Corollary 6.6.6. For any class $A \in \{\widehat{\mathsf{LDL}}, \mathsf{LDL}\}$ and $B \in \{\mathsf{PT}_1, \mathsf{PL}_\infty\}$, the classes A and B are incomparable.

Finally, as noted in [TV97], LDL is contained in $\mathsf{THR} \circ \mathsf{THR}$. The same argument shows that $\widehat{\mathsf{LDL}}$ is contained in $\mathsf{THR} \circ \mathsf{MAJ}$. Corollary 6.6.3 implies that these containments are strict.

6.7 Conclusions

We exhibited a function ($\mathsf{MAJ} \circ \mathsf{XOR}$) which is efficiently computable in $\mathsf{MAJ} \circ \mathsf{MAJ}$, but which cannot be computed by polynomial sized linear decision lists, resolving an open question of Turán and Vatan [TV97]. Figure 6.1 depicts where the class LDL, and its small-weight version $\widehat{\mathsf{LDL}}$, fit in the low depth threshold circuit hierarchy. We showed earlier in this chapter that a decision list of *exact threshold functions* cannot be computed by $\mathsf{THR} \circ \mathsf{MAJ}$. Some natural questions that arises from our work are as follows.

- Is $\mathsf{LDL} \subsetneq \mathsf{THR} \circ \mathsf{MAJ}$?
- Is $\mathsf{THR} \subseteq \widehat{\mathsf{LDL}}$? It is known that $\mathsf{THR} \subseteq \mathsf{MAJ} \circ \mathsf{MAJ}$ (see, for example, [GHR92, AM05, Hof96]). However it does not seem that any of the existing simulations of THR by $\mathsf{MAJ} \circ \mathsf{MAJ}$ can be easily modified to show $\mathsf{THR} \subseteq \widehat{\mathsf{LDL}}$.
- Is $\mathsf{PL}_1 \subseteq \mathsf{LDL}$? Is $\mathsf{PL}_1 \subseteq \widehat{\mathsf{LDL}}$?[1]
- Is $\widehat{\mathsf{LDL}} \subsetneq \mathsf{LDL}$?

6.8 References

The results presented in this chapter are based on joint work with Arkadev Chattopadhyay, Meena Mahajan and Nitin Saurabh [CMMS18].

[1] In a subsequent joint work with Arkadev Chattopadhyay and Suhail Sherif [CMS18], these were resolved in the negative. i.e. $\mathsf{PL}_1 \not\subseteq \mathsf{LDL}$.

Chapter 7

Summary and Conclusions

In Chapter 3, we showed a lifting theorem which lifts 'degree-hardness properties' of f to 'weight-hardness' properties of a lifted version of f, which we denote f^{op} (equivalently, f^{op} is just f lifted by a constant sized Indexing gadget). With some more work, the lifting theorem yielded lower bounds on the approximate weight and signed monomial complexity of symmetric functions, resolving conjectures posed by Ada, Fawzi and Hatami [AFH12] and Zhang [Zha92], respectively. We then showed an equivalence between the polynomial margin of a function f and the discrepancy of $f \circ \mathsf{XOR}$ and used this to reprove some known results and resolve a weak form of a conjecture by Zhang and Shi [ZS09]. The framework of our proofs is captured in Figure 7.1 (refer to Figure 3.1 for a more precise framework).

Some questions that remain open are listed below.

- We showed that $\mathsf{PP}(f \circ \mathsf{XOR})$ is tightly characterized by $m(f)$. One could ask whether the lower bound for $\mathsf{BPP}(f \circ \mathsf{XOR})$ in terms of $\mathsf{wt}_{1/3}(f)$ is tight. In

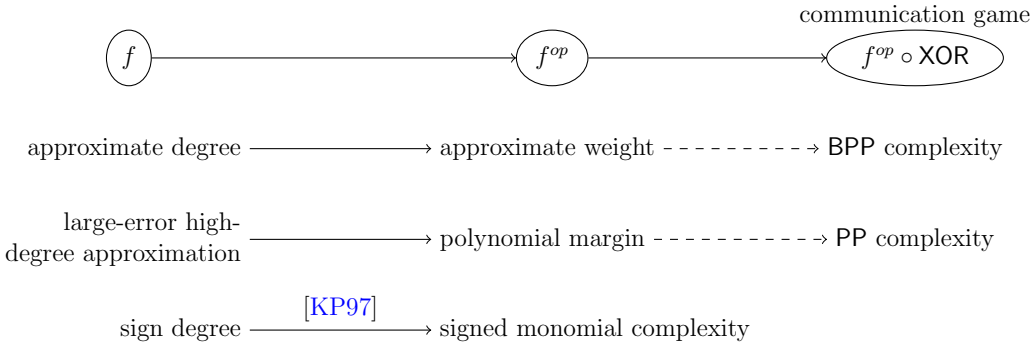

Figure 7.1: Framework of proofs in Chapter 3

other words, is it true that for every boolean function f,

$$\mathsf{BPP}(f \circ \mathsf{XOR}) \leq \mathrm{polylog}(\mathsf{wt}_{1/3}(f))? \tag{7.1}$$

In particular, we do not even know whether $\mathsf{BPP}(f \circ \mathsf{XOR}) \leq \mathrm{polylog}(\mathsf{wt}(f))$. Note that this is a special case of a conjecture by Grolmusz [Gro97], who conjectured that for *any* two-party function F, $\mathsf{BPP}(F) \leq \mathrm{polylog}(\mathsf{wt}(F))$.[1]

- It has been established that 'degree-hardness' properties of f lift to hardness of communication complexity of $f \circ \mathsf{PM}$ in the BPP and PP models [She11a]. It is believed that if f has large sign degree, then $\mathsf{UPP}(f \circ \mathsf{PM})$ is large [RS10]. We showed in Chapter 3 that a corresponding lifting theorem for XOR functions (in the BPP and PP models) holds when we start with *weight-hardness* properties rather than degree-hardness of f. A natural open question that arises is follows. Is it true that for any boolean function f,

$$\mathsf{UPP}(f \circ \mathsf{XOR}) \geq \Omega(\log \mathsf{mon}_{\pm}(f))?$$

In Chapter 4, we exhibited a function with large sign rank. The function was efficiently describable as a decision list of Equalities. The simplicity of the function along with the largeness of its sign rank yielded the circuit class separation $\mathsf{THR} \circ \mathsf{MAJ} \subsetneq \mathsf{THR} \circ \mathsf{THR}$ and the communication class separation $\mathsf{P}^{\mathsf{MA}} \not\subseteq \mathsf{UPP}$. The containment $\mathsf{THR} \circ \mathsf{MAJ} \subsetneq \mathsf{THR} \circ \mathsf{THR}$ was open since the work of Goldmann et al. [GHR92], and was later explicitly posed by Amano and Maruoka [AM05] and Hansen and Podolskii [HP10]. The communication class separation $\mathsf{P}^{\mathsf{MA}} \not\subseteq \mathsf{UPP}$ implies $\mathsf{S_2P} \not\subseteq \mathsf{UPP}$, resolving an open question posed by Göös, Pitassi and Watson [GPW18].

A well-identified frontier in circuit complexity is to prove explicit lower bounds against $\mathsf{THR} \circ \mathsf{THR}$. A natural program to take a step in this direction that arises from our work is to prove lower bounds against *decision lists of exact threshold functions*. We additionally exploited a weakness of the class of decision lists of *linear* threshold functions observed by Turán and Vatan [TV97] and showed that $\mathsf{MAJ} \circ \mathsf{XOR}$ is not in this class, completely resolving an open question posed by Turán and Vatan [TV97]. We also proved a sign rank lower bound for certain symmetric XOR functions. In conclusion, we list some open questions and possible future directions of work.

[1] In a subsequent joint work with Arkadev Chattopadhyay and Suhail Sherif [CMS18], Grolmusz's conjecture and Equation (7.1) were shown to be false.

- Prove lower bounds, for functions in NP, against decision lists of exact thresholds. It is interesting to note that such lower bounds are known against decision lists of *Equalities* (since they are in AC^0, for instance).

- We exhibited a function in $\mathsf{THR} \circ \mathsf{THR}$ that requires size $2^{\Omega(n^{1/4})}$ to be computed by $\mathsf{THR} \circ \mathsf{MAJ}$ circuits. Can one improve this separation?

We showed in Chapter 5 that the PP_k complexity of GHR_k^N is $\Omega\left(\sqrt{N}\right)$ for $k = O(\log N)$. As mentioned in Section 5.1.2, Sherstov [She16c] shows existence of functions with $\Omega(N)$ cost in PP_k but that have efficient UPP_k protocols. In general, current techniques do not allow us to go beyond $\log N$ players to prove lower bounds for the cost of even deterministic protocols. This remains one of the most interesting problems in NOF complexity. However, let us remark that for many of the functions used in the literature (see for example [Gro94, BGKL03, ACFN15, CS14]), there are surprisingly efficient protocols when $k > \log N$. Moreover these protocols are typically deterministic and either simultaneous or barely interactive. On the other hand, we do not immediately see an efficient randomized interactive protocol for GHR_k^N at $k > \log N$.

- Is GHR_k^N a hard function even for $k > \log N$?

- Can one exhibit an explicit function in UPP_k that requires $\Omega(N)$ PP_k cost?

- Recall our Margin-Discrepancy equivalence from Theorem 3.1.4. This implied that $\mathsf{PP}(f \circ \mathsf{XOR}) = \Theta\left(\log \frac{1}{m(f)}\right)$. Recall that GHR_k^N could be expressed as $f \circ \mathsf{XOR}$, where $m(f)$ is exponentially small. Thus, one might hope that the Margin-Discrepancy equivalence continues to hold for multi-party communication. Interestingly, this belief is false! The function $\mathsf{MOD}_4 \circ \mathsf{XOR}_3$ was shown to be easy for deterministic 3-player protocols [ACFN15]. Is there a neat characterization of discrepancy of XOR functions in the NOF model?

- Hansen and Podolskii [HP15] showed that $(\log N)^{\omega(1)}$ unbounded-error lower bounds for functions in the 3-party NOF model imply super-polynomial lower bounds against $\mathsf{THR} \circ \mathsf{THR}$. This is another possible line of attack towards proving strong $\mathsf{THR} \circ \mathsf{THR}$ lower bounds.

There are several future directions that stem from our lower bounds against linear decision lists in Chapter 6. These were discussed in Section 6.7.

List of my Publications

[1] Arkadev Chattopadhyay, Meena Mahajan, Nikhil S. Mande, and Nitin Saurabh. Lower Bounds for Linear Decision Lists. Manuscript, 2018.

[2] Arkadev Chattopadhyay and Nikhil S. Mande. Separation of Unbounded-Error Models in Multi-Party Communication Complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, 2016 https://eccc.weizmann.ac.il/report/2016/095 To appear in Theory of Computing.

[3] Arkadev Chattopadhyay and Nikhil S. Mande. Dual Polynomials and Communication Complexity of XOR Functions. *Electronic Colloquium on Computational Complexity (ECCC)*, 2017. https://eccc.weizmann.ac.il/report/2017/062 (Extended version of 'A Lifting Theorem with Applications to Symmetric Functions').

[4] Arkadev Chattopadhyay and Nikhil S. Mande. A Lifting Theorem with Applications to Symmetric Functions. https://doi.org/10.4230/LIPIcs.FSTTCS.2017.23 *FSTTCS*, 2017.

[5] Arkadev Chattopadhyay and Nikhil S. Mande. A Short List of Equalities Induces Large Sign Rank *CElectronic Colloquium on Computational Complexity (ECCC)*, 2017. https://eccc.weizmann.ac.il/report/2017/083 To appear in FOCS, 2018.

Bibliography

[ACFN15] Anil Ada, Arkadev Chattopadhyay, Omar Fawzi, and Phuong Nguyen. The NOF Multiparty Communication Complexity of Composed Functions. *Computational Complexity*, 24(3):645–694, 2015. Preliminary version in the *39th International Colloquium on Automata, Languages and Programming (ICALP 2012)*.

[AFH12] Anil Ada, Omar Fawzi, and Hamed Hatami. Spectral Norm of Symmetric Functions. In *Proceedings of the 16th International Workshop on Randomization and Computation (RANDOM 2012)*, pages 338–349, 2012.

[AFK17] Anil Ada, Omar Fawzi, and Raghav Kulkarni. On the Spectral Properties of Symmetric Functions. *CoRR*, 2017. arXiv:1704.03176.

[AFR85] Noga Alon, Peter Frankl, and Vojtech Rödl. Geometrical Realization of Set Systems and Probabilistic Communication Complexity. In *Proceedings of the 26th Annual IEEE Symposium on Foundations of Computer Science (FOCS 1985)*, pages 277–280, 1985.

[AM05] Kazuyuki Amano and Akira Maruoka. On the Complexity of Depth-2 Circuits with Threshold Gates. In *Proceedings of the 30th International Symposium on the Mathematical Foundations of Computer Science (MFCS 2005)*, pages 107–118, 2005.

[AMY16] Noga Alon, Shay Moran, and Amir Yehudayoff. Sign rank versus VC dimension. In *Proceedings of the 29th Annual Conference on Computational Learning Theory (COLT 2016)*, pages 47–80, 2016.

[AW17] Josh Alman and R. Ryan Williams. Probabilistic rank and matrix rigidity. In *Proceedings of the 49th Annual ACM Symposium on Theory of Computing (STOC 2017)*, pages 641–652, 2017.

[BBG14] Eric Blais, Joshua Brody, and Badih Ghazi. The Information Complexity of Hamming Distance. In *Proceedings of the 18th International Workshop on Randomization and Computation (RANDOM 2014)*, pages 465–489, 2014.

[BCH+16] Adam Bouland, Lijie Chen, Dhiraj Holden, Justin Thaler, and Prashant Nalini Vasudevan. On the Power of Statistical Zero Knowledge.

In *Proceedings of the 58th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2017)*, pages 708–719, 2016.

[BDPW10] Paul Beame, Matei David, Toniann Pitassi, and Philipp Woelfel. Separating Deterministic from Randomized Multiparty Communication Complexity. *Theory of Computing*, 6(1):201–225, 2010. Preliminary version in the *34th International Colloquium on Automata, Languages and Programming (ICALP 2007)*.

[Bei94] Richard Beigel. Perceptrons, PP, and the Polynomial Hierarchy. *Computational Complexity*, 4:339–349, 1994. Preliminary version in the *7th Annual Structure in Complexity Theory Conference (Structures 1992)*.

[BFS86] László Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory (preliminary version). In *Proceedings of the 27th Annual IEEE Symposium on Foundations of Computer Science (FOCS 1986)*, pages 337–347, 1986.

[BGKL03] László Babai, Anna Gál, Peter G. Kimmel, and Satyanarayana V. Lokam. Communication Complexity of Simultaneous Messages. *SIAM J. Comput.*, 33(1):137–166, 2003.

[BNS92] László Babai, Noam Nisan, and Mario Szegedy. Multiparty Protocols, Pseudorandom Generators for Logspace, and Time-Space Trade-Offs. *J. Comput. Syst. Sci.*, 45(2):204–232, 1992. Preliminary version in the *21st Annual ACM Symposium on Theory of Computing (STOC 1989)*.

[Bol86] Béla Bollobás. *Combinatorics: set systems, hypergraphs, families of vectors, and combinatorial probability*. Cambridge University Press, 1986.

[Bou05] Jean Bourgain. Estimation of certain exponential sums arising in complexity theory. *Comptes Rendus Mathematique*, 340(9):627–631, 2005.

[Bru90] Jehoshua Bruck. Harmonic Analysis of Polynomial Threshold Functions. *SIAM J. Discrete Math.*, 3(2):168–177, 1990.

[BT15a] Mark Bun and Justin Thaler. Dual lower bounds for approximate degree and Markov-Bernstein inequalities. *Inf. Comput.*, 243:2–25, 2015. Preliminary version in the *42nd International Colloquium on Automata, Languages and Programming (ICALP 2015)*.

[BT15b] Mark Bun and Justin Thaler. Hardness Amplification and the Approximate Degree of Constant-Depth Circuits. In *Proceedings of the 42nd International Colloquium on Automata, Languages and Programming (ICALP 2015)*, pages 268–280, 2015.

[BT16] Mark Bun and Justin Thaler. Improved Bounds on the Sign-Rank of AC^0. In *Proceedings of the 43rd International Colloquium on Automata, Languages and Programming (ICALP 2016)*, pages 37:1–37:14, 2016.

[BT17] Mark Bun and Justin Thaler. A Nearly Optimal Lower Bound on the Approximate Degree of AC^0. In *Proceedings of the 58th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2017)*, pages 1–12, 2017.

[BVdW07] Harry Buhrman, Nikolai K. Vereshchagin, and Ronald de Wolf. On Computation and Communication with Small Bias. In *Proceedings of the 22nd Annual IEEE Conference on Computational Complexity (CCC 2007)*, pages 24–32, 2007.

[CA08] Arkadev Chattopadhyay and Anil Ada. Multiparty Communication Complexity of Disjointness. *Electronic Colloquium on Computational Complexity (ECCC)*, 2008. http://eccc.hpi-web.de/eccc-reports/2008/TR08-002/index.html.

[CFL83] Ashok K. Chandra, Merrick L. Furst, and Richard J. Lipton. Multi-Party Protocols. In *Proceedings of the 15th Annual ACM Symposium on Theory of Computing (STOC 1983)*, pages 94–99, 1983.

[CGPT06] Arkadev Chattopadhyay, Navin Goyal, Pavel Pudlák, and Denis Thérien. Lower bounds for circuits with MOD_m gates. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006)*, pages 709–718, 2006.

[Cha07] Arkadev Chattopadhyay. Discrepancy and the Power of Bottom Fan-in in Depth-three Circuits. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2007)*, pages 449–458, 2007.

[Cha09] Arkadev Chattopadhyay. *Circuits, Communication and Polynomials*. PhD thesis, McGill University, 2009.

[CM16] Arkadev Chattopadhyay and Nikhil Mande. A Separation of Unbounded-Error Models in Multiparty Communication Complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, 2016. http://eccc.hpi-web.de/report/2016/095, To appear in Theory of Computing.

[CM17a] Arkadev Chattopadhyay and Nikhil S. Mande. Dual polynomials and communication complexity of XOR functions. *CoRR*, 2017. http://arxiv.org/abs/1704.02537, (Extended version of 'A Lifting Theorem with Applications to Symmetric Functions').

[CM17b] Arkadev Chattopadhyay and Nikhil S. Mande. A Lifting Theorem with Applications to Symmetric Functions. In *37th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2017, December 11-15, 2017, Kanpur, India*, pages 23:1–23:14, 2017.

[CM17c] Arkadev Chattopadhyay and Nikhil S. Mande. A Short List of Equalities Induces Large Sign Rank. *CoRR*, 2017. http://arxiv.org/abs/1705.02397, To appear in FOCS, 2018.

[CMMS18] Arkadev Chattopadhyay, Meena Mahajan, Nikhil S. Mande, and Nitin Saurabh. Lower bounds for linear decision lists. Manuscript, 2018.

[CMS18] Arkadev Chattopadhyay, Nikhil S. Mande, and Suhail Sherif. The Log-Approximate-Rank Conjecture is False. *Electronic Colloquium on Computational Complexity (ECCC)*, 2018. https://eccc.weizmann.ac.il/report/2018/176.

[COS17] Xi Chen, Igor Carboni Oliveira, and Rocco A. Servedio. Addition is exponentially harder than counting for shallow monotone circuits. In *Proceedings of the 49th Annual ACM Symposium on Theory of Computing (STOC 2017)*, pages 1232–1245, 2017.

[CS14] Arkadev Chattopadhyay and Michael E. Saks. The Power of Super-logarithmic Number of Players. In *Proceedings of the 18th International Workshop on Randomization and Computation (RANDOM 2014)*, pages 596–603, 2014.

[CSS16] Ruiwen Chen, Rahul Santhanam, and Srikanth Srinivasan. Average-Case Lower Bounds and Satisfiability Algorithms for Small Threshold Circuits. In *Proceedings of the 31st Annual Computational Complexity Conference (CCC 2016)*, pages 1:1–1:35, 2016.

[DKO14] Andrew Drucker, Fabian Kuhn, and Rotem Oshman. On the power of the congested clique model. In *Proceedings of the 32nd Annual ACM Symposium on Principles of Distributed Computing (PODC 2014)*, pages 367–376, 2014.

[dW10] Ronald de Wolf. A note on quantum algorithms and the minimal degree of ϵ-error polynomials for symmetric functions. *Quantum Information & Computation*, 8(10):943–950, 2010.

[EZ64] Hartmut Ehlich and Karl Zeller. Schwankung von polynomen zwischen gitterpunkten. *Mathematische Zeitschrift*, 86(1):41–44, 1964.

[FF81] Peter Frankl and Zoltán Füredi. A short proof for a theorem of Harper about Hamming-spheres. *Discrete Mathematics*, 34(3):311–313, 1981.

[FKL+01] Jürgen Forster, Matthias Krause, Satyanarayana V. Lokam, Rustam Mubarakzjanov, Niels Schmitt, and Hans Ulrich Simon. Relations Between Communication Complexity, Linear Arrangements, and Computational Complexity. In *Proceedings of the 21st International Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2001)*, pages 171–182, 2001.

[For02] Jürgen Forster. A linear lower bound on the unbounded error probabilistic communication complexity. *J. Comput. Syst. Sci.*, 65(4):612–625, 2002. Preliminary version in the *16th Annual IEEE Conference on Computational Complexity (CCC 2001)*.

[GHR92] Mikael Goldmann, Johan Håstad, and Alexander A. Razborov. Majority Gates vs. General Weighted Threshold Gates. *Computational Complexity*, 2:277–300, 1992. Preliminary version in the *7th Annual Structure in Complexity Theory Conference (Structures 1992)*.

[GK98] Mikael Goldmann and Marek Karpinski. Simulating Threshold Circuits by Majority Circuits. *SIAM J. Comput.*, 27(1):230–246, 1998. Preliminary version in the *25th Annual ACM Symposium on Theory of Computing (STOC 1993)*.

[Göö17] Mika Göös. Private Communication, 2017.

[GPW18] Mika Göös, Toniann Pitassi, and Thomas Watson. The Landscape of Communication Complexity Classes. *Computational Complexity*, 27(2):245–304, 2018. Preliminary version in the *43rd International Colloquium on Automata, Languages and Programming (ICALP 2016)*.

[Gro94] Vince Grolmusz. The BNS Lower Bound for Multi-Party Protocols in Nearly Optimal. *Inf. Comput.*, 112(1):51–54, 1994.

[Gro97] Vince Grolmusz. On the Power of Circuits with Gates of Low L_1 Norms. *Theor. Comput. Sci.*, 188(1-2):117–128, 1997.

[GT91] Hans Dietmar Gröger and György Turán. On Linear Decision Trees Computing Boolean Functions. In *Proceedings of the 18th International Colloquium on Automata, Languages and Programming (ICALP 1991)*, pages 707–718, 1991.

[Har66] L. H. Harper. Optimal numberings and isoperimetric problems on graphs. *Journal of Combinatorial Theory*, 1:385–393, 1966.

[HG91] Johan Håstad and Mikael Goldmann. On the Power of Small-Depth Threshold Circuits. *Computational Complexity*, 1:113–129, 1991. Preliminary version in the *31st Annual IEEE Symposium on Foundations of Computer Science (FOCS 1990)*.

[HHL18] Hamed Hatami, Kaave Hosseini, and Shachar Lovett. Structure of Protocols for XOR Functions. *SIAM J. Comput.*, 47(1):208–217, 2018. Preliminary version in the *57th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2016)*.

[HKM12] Prahladh Harsha, Adam R. Klivans, and Raghu Meka. An invariance principle for polytopes. *J. ACM*, 59(6):29:1–29:25, 2012. Preliminary

version in the *42nd Annual ACM Symposium on Theory of Computing (STOC 2010)*.

[HMP+93] A. Hajnal, W. Maas, P. Pudlák, M. Szegedy, and G. Turán. Threshold Circuits of Bounded Depth. *J. Comput. Syst. Sci.*, 46(2):129–154, 1993.

[Hof96] Thomas Hofmeister. A Note on the Simulation of Exponential Threshold Weights. In *Computing and Combinatorics, Second Annual International Conference, COCOON '96, Hong Kong, June 17-19, 1996, Proceedings*, pages 136–141, 1996.

[HP10] Kristoffer Arnsfelt Hansen and Vladimir V. Podolskii. Exact Threshold Circuits. In *Proceedings of the 25th Annual IEEE Conference on Computational Complexity (CCC 2010)*, pages 270–279, 2010.

[HP15] Kristoffer Arnsfelt Hansen and Vladimir V. Podolskii. Polynomial threshold functions and Boolean threshold circuits. *Inf. Comput.*, 240:56–73, 2015. Preliminary version in the *38th International Symposium on the Mathematical Foundations of Computer Science (MFCS 2013)*.

[HQ17] Hamed Hatami and Yingjie Qian. The Unbounded-Error Communication Complexity of symmetric XOR functions. *CoRR*, 2017. arXiv:1704.00777.

[IW10] Russell Impagliazzo and Ryan Williams. Communication Complexity with Synchronized Clocks. In *Proceedings of the 25th Annual IEEE Conference on Computational Complexity (CCC 2010)*, pages 259–269, 2010.

[Kla03] Hartmut Klauck. Rectangle Size Bounds and Threshold Covers in Communication Complexity. In *Proceedings of the 18th Annual IEEE Conference on Computational Complexity (CCC 2003)*, pages 118–134, 2003.

[Kla07] Hartmut Klauck. Lower Bounds for Quantum Communication Complexity. *SIAM J. Comput.*, 37(1):20–46, 2007. Preliminary version in the *42nd Annual IEEE Symposium on Foundations of Computer Science (FOCS 2001)*.

[KLL+15] Iordanis Kerenidis, Sophie Laplante, Virginie Lerays, Jérémie Roland, and David Xiao. Lower Bounds on Information Complexity via Zero-Communication Protocols and Applications. *SIAM J. Comput.*, 44(5):1550–1572, 2015. Preliminary version in the *53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS 2012)*.

[KMSY18] Sampath Kannan, Elchanan Mossel, Swagato Sanyal, and Grigory Yaroslavtsev. Linear Sketching over $\mathbb{F}_2$. In *Proceedings of the 33rd Annual Computational Complexity Conference (CCC 2018)*, pages 8:1–8:37, 2018.

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, 1997.

[KOS04] Adam R. Klivans, Ryan O'Donnell, and Rocco A. Servedio. Learning intersections and thresholds of halfspaces. *J. Comput. Syst. Sci.*, 68(4):808–840, 2004. Preliminary version in the *43rd Annual IEEE Symposium on Foundations of Computer Science (FOCS 2002)*.

[KP97] Matthias Krause and Pavel Pudlák. On the Computational Power of Depth-2 Circuits with Threshold and Modulo Gates. *Theor. Comput. Sci.*, 174(1-2):137–156, 1997. Preliminary version in the *26th Annual ACM Symposium on Theory of Computing (STOC 1994)*.

[KP98] Matthias Krause and Pavel Pudlák. Computing Boolean Functions by Polynomials and Threshold Circuits. *Computational Complexity*, 7(4):346–370, 1998. Preliminary version in the *36th Annual IEEE Symposium on Foundations of Computer Science (FOCS 1995)*.

[Kra06] Matthias Krause. On the computational power of Boolean decision lists. *Computational Complexity*, 14(4):362–375, 2006. Preliminary version in the *19th Symposium on Theoretical Aspects of Computer Science (STACS 2002)*.

[KS92] Bala Kalyanasundaram and Georg Schnitger. The Probabilistic Communication Complexity of Set Intersection. *SIAM J. Discrete Math.*, 5(4):545–557, 1992. Preliminary version in the *2nd Annual Structure in Complexity Theory Conference (Structures 1987)*.

[KS06] Adam R. Klivans and Rocco A. Servedio. Toward Attribute Efficient Learning of Decision Lists and Parities. *Journal of Machine Learning Research*, 7:587–602, 2006. Preliminary version in the *17th Annual Conference on Computational Learning Theory (COLT 2004)*.

[KW16] Daniel M. Kane and Ryan Williams. Super-linear gate and super-quadratic wire lower bounds for depth-two and depth-three threshold circuits. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing (STOC 2016)*, pages 633–643, 2016.

[LLZ11] Ming Lam Leung, Yang Li, and Shengyu Zhang. Tight Bounds on Communication Complexity of Symmetric XOR Functions in One-Way and SMP Models. In *Theory and Applications of Models of Computation - 8th Annual Conference, TAMC 2011, Tokyo, Japan, May 23-25, 2011. Proceedings*, pages 403–408, 2011.

[LS88] László Lovász and Michael E. Saks. Lattices, Möbius Functions and Communication Complexity. In *Proceedings of the 29th Annual IEEE Symposium on Foundations of Computer Science (FOCS 1988)*, pages 81–90, 1988.

[LS09a] Troy Lee and Adi Shraibman. Disjointness is Hard in the Multiparty Number-on-the-Forehead Model. *Computational Complexity*, 18(2):309–336, 2009. Preliminary version in the *23rd Annual IEEE Conference on Computational Complexity (CCC 2008)*.

[LS09b] Troy Lee and Adi Shraibman. Lower Bounds in Communication Complexity. *Foundations and Trends in Theoretical Computer Science*, 3(4):263–398, 2009.

[LS09c] Nathan Linial and Adi Shraibman. Learning Complexity vs Communication Complexity. *Combinatorics, Probability & Computing*, 18(1-2):227–245, 2009. Preliminary version in the *23rd Annual IEEE Conference on Computational Complexity (CCC 2008)*.

[LZ10] Troy Lee and Shengyu Zhang. Composition Theorems in Communication Complexity. In *Proceedings of the 37th International Colloquium on Automata, Languages and Programming (ICALP 2010)*, pages 475–489, 2010.

[LZ13] Yang Liu and Shengyu Zhang. Quantum and randomized communication complexity of XOR functions in the SMP model. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:10, 2013.

[MO09] Ashley Montanaro and Tobias Osborne. On the communication complexity of XOR functions. *CoRR*, abs/0909.3392, 2009.

[MP69] Marvin Minsky and Seymour Papert. *Perceptrons*. MIT Press, 1969.

[Mur71] S. Muroga. *Threshold Logic and its Applications*. Wiley-Interscience, 1971.

[NS94] Noam Nisan and Mario Szegedy. On the Degree of Boolean Functions as Real Polynomials. *Computational Complexity*, 4:301–313, 1994. Preliminary version in the *24th Annual ACM Symposium on Theory of Computing (STOC 1992)*.

[Pat92] Ramamohan Paturi. On the Degree of Polynomials that Approximate Symmetric Boolean Functions (Preliminary Version). In *Proceedings of the 24th Annual ACM Symposium on Theory of Computing (STOC 1992)*, pages 468–474, 1992.

[Pat10] Mihai Patrascu. Towards polynomial lower bounds for dynamic problems. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing (STOC 2010)*, pages 603–610, 2010.

[PS86] Ramamohan Paturi and Janos Simon. Probabilistic Communication Complexity. *J. Comput. Syst. Sci.*, 33(1):106–123, 1986. Preliminary version in the *25th Annual IEEE Symposium on Foundations of Computer Science (FOCS 1984)*.

[Raz92a] Alexander A. Razborov. On small depth Threshold circuits. In *Third Scandinavian Workshop on Algorithm Theory (SWAT)*, pages 42–52, 1992.

[Raz92b] Alexander A. Razborov. On the Distributional Complexity of Disjointness. *Theor. Comput. Sci.*, 106(2):385–390, 1992. Proceedings of the 17th International Colloquium on Automata, Languages and Programming (ICALP 1990).

[Raz00] Ran Raz. The BNS-Chung criterion for multi-party communication complexity. *Computational Complexity*, 9(2):113–122, 2000.

[Raz03] Alexander A. Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya: Mathematics*, 67(1):145, 2003.

[RC66] Theodore J Rivlin and Elliott W Cheney. A comparison of uniform approximations on an interval and a finite subset thereof. *SIAM Journal on numerical Analysis*, 3(2):311–320, 1966.

[Riv87] Ronald L. Rivest. Learning Decision Lists. *Machine Learning*, 2(3):229–246, 1987.

[RS10] Alexander A. Razborov and Alexander A. Sherstov. The Sign-Rank of AC^0. *SIAM J. Comput.*, 39(5):1833–1855, 2010. Preliminary version in the *49th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2008)*.

[RY15] Anup Rao and Amir Yehudayoff. Simplified Lower Bounds on the Multiparty Communication Complexity of Disjointness. In *Proceedings of the 30th Annual Computational Complexity Conference (CCC 2015)*, pages 88–101, 2015.

[SB91] K. I. Siu and J. Bruck. On the Power of Thrshold Circuits with Small Weights. *SIAM J. Discrete Math.*, 4(3):423–435, 1991.

[She08] Alexander A. Sherstov. Halfspace Matrices. *Computational Complexity*, 17(2):149–178, 2008. Preliminary version in the *22nd Annual IEEE Conference on Computational Complexity (CCC 2007)*.

[She09a] Alexander A. Sherstov. Approximate Inclusion-Exclusion for Arbitrary Symmetric Functions. *Computational Complexity*, 18(2):219–247, 2009. Preliminary version in the *23rd Annual IEEE Conference on Computational Complexity (CCC 2008)*.

[She09b] Alexander A. Sherstov. Separating AC^0 from Depth-2 Majority Circuits. *SIAM J. Comput.*, 38(6):2113–2129, 2009. Preliminary version in the *39th Annual ACM Symposium on Theory of Computing (STOC 2007)*.

[She11a] Alexander A. Sherstov. The Pattern Matrix Method. *SIAM J. Comput.*, 40(6):1969–2000, 2011. Preliminary version in the *40th Annual ACM Symposium on Theory of Computing (STOC 2008)*.

[She11b] Alexander A. Sherstov. The unbounded-error communication complexity of symmetric functions. *Combinatorica*, 31(5):583–614, 2011. Preliminary version in the *49th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2008)*.

[She13a] Alexander A. Sherstov. The Intersection of Two Halfspaces Has High Threshold Degree. *SIAM J. Comput.*, 42(6):2329–2374, 2013. Preliminary version in the *50th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2009)*.

[She13b] Alexander A. Sherstov. Optimal bounds for sign-representing the intersection of two halfspaces by polynomials. *Combinatorica*, 33(1):73–96, 2013. Preliminary version in the *42nd Annual ACM Symposium on Theory of Computing (STOC 2010)*.

[She14] Alexander A. Sherstov. Communication Lower Bounds Using Directional Derivatives. *J. ACM*, 61(6):34:1–34:71, 2014. Preliminary version in the *45th Annual ACM Symposium on Theory of Computing (STOC 2013)*.

[She15] Alexander A. Sherstov. The Power of Asymmetry in Constant-Depth Circuits. In *Proceedings of the 56th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2015)*, pages 431–450, 2015.

[She16a] Alexander A. Sherstov. Private Communication, 2016.

[She16b] Alexander A. Sherstov. The Multiparty Communication Complexity of Set Disjointness. *SIAM J. Comput.*, 45(4):1450–1489, 2016. Preliminary version in the *44th Annual ACM Symposium on Theory of Computing (STOC 2012)*.

[She16c] Alexander A. Sherstov. On multiparty communication with large versus unbounded error. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:138, 2016. To appear in Theory of Computing.

[Spa08] Robert Spalek. A Dual Polynomial for OR. *CoRR*, 2008.

[ST17] Rocco A. Servedio and Li-Yang Tan. Fooling Intersections of Low-Weight Halfspaces. In *Proceedings of the 58th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2017)*, pages 824–835, 2017.

[SZ09] Yaoyun Shi and Yufan Zhu. Quantum communication complexity of block-composed functions. *Quantum Information & Computation*, 9(5):444–460, 2009.

[Tha16] Justin Thaler. Lower Bounds for the Approximate Degree of Block-Composed Functions. In *Proceedings of the 43rd International Colloquium on Automata, Languages and Programming (ICALP 2016)*, pages 17:1–17:15, 2016.

[TV97] György Turán and Farrokh Vatan. Linear decision lists and partitioning algorithms for the construction of neural networks. In *Foundations of Computational Mathematics*, pages 414–423. Springer, 1997.

[UT11] Kei Uchizawa and Eiji Takimoto. Lower Bounds for Linear Decision Trees via an Energy Complexity Argument. In *Proceedings of the 36th International Symposium on the Mathematical Foundations of Computer Science (MFCS 2011)*, pages 568–579, 2011.

[UT15] Kei Uchizawa and Eiji Takimoto. Lower Bounds for Linear Decision Trees with Bounded Weights. In *SOFSEM 2015: Theory and Practice of Computer Science - 41st International Conference on Current Trends in Theory and Practice of Computer Science, Pec pod Sněžkou, Czech Republic, January 24-29, 2015. Proceedings*, pages 412–422, 2015.

[Yao79] Andrew Chi-Chih Yao. Some Complexity Questions Related to Distributive Computing (Preliminary Report). In *Proceedings of the 11th Annual ACM Symposium on Theory of Computing (STOC 1979)*, pages 209–213, 1979.

[Zha92] Zhi-Li Zhang. Complexity of symmetric functions in perceptron-like models. Master's thesis, University of Massachusetts at Amherst, 1992.

[Zha14] Shengyu Zhang. Efficient quantum protocols for XOR functions. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 1878–1885, 2014.

[ZS09] Zhiqiang Zhang and Yaoyun Shi. Communication complexities of symmetric XOR functions. *Quantum Information & Computation*, 9(3):255–263, 2009.