Electronic Colloquium on Computational Complexity, Report No. 62 (2017)

Dual polynomials and communication complexity of XOR functions

Arkadev Chattopadhyay*[1] and Nikhil S. Mande†[1]

[1] *School of Technology and Computer Science, TIFR, Mumbai*

Abstract

We show a new duality between the *polynomial margin* complexity of f and the *discrepancy* of the function $f \circ \mathsf{XOR}$, called an XOR function. Using this duality, we develop polynomial based techniques for understanding the bounded error (BPP) and the weakly-unbounded error (PP) communication complexities of XOR functions. This enables us to show the following.

- A weak form of an interesting conjecture of Zhang and Shi[1] [41] asserts that for symmetric functions $f : \{0,1\}^n \to \{-1,1\}$, the weakly unbounded-error complexity of $f \circ \mathsf{XOR}$ is essentially characterized by the number of points i in the set $\{0,1,\ldots,n-2\}$ for which $D_f(i) \neq D_f(i+2)$, where D_f is the predicate corresponding to f. The number of such points is called the *odd-even* degree of f. We observe that a much earlier work of Zhang [40] implies that the PP complexity of $f \circ \mathsf{XOR}$ is $O(k \log n)$, where k is the odd-even degree of f. We show that the PP complexity of $f \circ \mathsf{XOR}$ is $\Omega(k/\log(n/k))$.
- We resolve a conjecture of Zhang [40] characterizing the Threshold of Parity circuit size of symmetric functions in terms of their odd-even degree.
- We obtain a new proof of the exponential separation between PP^{cc} and UPP^{cc} via an XOR function.
- We provide a characterization of the *approximate spectral norm* of symmetric functions, affirming a conjecture of Ada et al. [2] which has several consequences (cf. [2]). This also provides a new proof of the characterization of the bounded error complexity of symmetric XOR functions due to [41].

Additionally, we prove strong UPP lower bounds for $f \circ \mathsf{XOR}$, when f is symmetric and periodic with period $O(n^{1/2-\epsilon})$, for any constant $\epsilon > 0$. More precisely, we show that every such XOR function has unbounded error complexity $n^{\Omega(1)}$, unless f is constant or parity or its complement, in which case the complexity is just $O(1)$. As a direct consequence of this, we derive new exponential lower bounds on the size of depth-2 threshold circuits computing such XOR functions. Our UPP lower bounds do not involve the use of linear programming duality.

*Partially supported by a Ramanujan fellowship of the DST. arkadev.c@tifr.res.in

†Supported by a DAE fellowship. nikhil.mande@tifr.res.in

[1] The full conjecture has just been reported to be independently settled by Hatami and Qian [18]. However, their techniques are quite different and are not known to yield many of the results we obtain here.

ISSN 1433-8092

1 Introduction

We consider three well known models of randomized communication, in all of which Alice and Bob use only *private random coins*. Alice and Bob receive a pair of inputs $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ respectively. They want to jointly evaluate a function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ on the pair (X,Y) by using a communication protocol that minimizes the total *cost* in the worst case. The protocol is probabilistic with the requirement that $\Pr[\Pi(X,Y) = f(X,Y)] \geq 1/2 + \epsilon$, where $\epsilon > 0$. Each of the three models specifies a basic requirement on ϵ, and the goal of the players is to design an *efficient* protocol meeting this requirement that minimizes the cost. Further, each model has its own cost function. Protocols are efficient if their cost is poly-logarithmic in n, the length of the inputs to Alice and Bob.

Yao [39] introduced the model where the advantage ϵ needs to be a positive constant independent of the length of the inputs and the cost is the total number of bits communicated. The cost of the best protocol for computing a function in this model is called its *bounded error* complexity. Paturi and Simon [28] relaxed the requirement on advantage completely: ϵ only needs to be positive, but it can be otherwise decreasing arbitrarily with n. The complexity of f in this model is called its *unbounded error* complexity. Babai et al. [3] introduced a semi-relaxed model whose power is sandwiched between the two above models: while the correctness requirement is the same as that in the unbounded error case, low advantage is penalised by introducing a term in the cost function: the cost of a protocol is the sum of the total number of bits communicated *and* $\log(1/\epsilon)$. The complexity of a function in this semi-relaxed model is called its *weakly-unbounded error* complexity. The set of functions that have efficient bounded, weakly-unbounded and unbounded error protocols are called BPP^{cc}, PP^{cc} and UPP^{cc} respectively, closely borrowing terminology from standard Turing machine complexity classes.

Clearly, $\mathsf{BPP}^{cc} \subseteq \mathsf{PP}^{cc} \subseteq \mathsf{UPP}^{cc}$. Set-Disjointness, denoted by DISJ, separates BPP^{cc} from PP^{cc} due to [3] and the following simple PP protocol of logarithmic cost: Alice randomly chooses an index i in $[n]$ and sends the value of i and her ith bit to Bob. If both Alice and Bob have 1 as their ith bit, Bob outputs that they are not disjoint. Otherwise Alice and Bob output a random answer. Thus, the weakly-unbounded error complexity of DISJ, commonly considered to be a hard function [3, 29, 19], is exponentially smaller than its bounded error complexity.

There are fewer known strong lower bounds for the PP model than the bounded error one. This is partly explained by the fact that while techniques based on corruption and information theory yield lower bounds for bounded error model, the PP model is exactly characterized by the stronger measure of discrepancy [20]. Still, there are several functions for which discrepancy can be bounded. PP^{cc} was separated from UPP^{cc} in independent works of Sherstov [31] and Buhrman et al. [7]. Proving lower bounds for the unbounded error model, on the other hand, is even more difficult. The only known way for proving bounds here is lower bounding the *sign-rank* of the communication matrix [28]. The sign rank of a real matrix M with non-zero entries is the smallest number r such that there exists a matrix M' of rank r and the same dimension as that of M such that each of its entries has the same sign as the corresponding one in M. Clearly, there is a matrix rigidity-like flavor to this definition, perhaps explaining the difficulty of estimating this quantity well. In a beautiful and breakthrough work, Forster [14] managed to show that the Inner-Product (IP) function has high sign rank and consequently high unbounded-error complexity. The technique of Forster, relating the spectral norm of a matrix to its sign rank, forms the basis for the few subsequent works on lower bounds for explicit functions in the model. In particular, Razborov and Sherstov [30] and Sherstov [34] prove lower bounds on different functions making very interesting use of additional tools from approximation theory.

We consider a different class of functions in the unbounded error model. To explain this, let us introduce function composition. Given $f : \{0,1\}^n \to \{0,1\}$ and $g : \{0,1\}^{2b} \to \{0,1\}$, we denote by $f \circ g$ the following composed function: its input is naturally viewed as a $2 \times bn$ matrix consisting of n blocks each of which is a $2 \times b$ matrix. Alice and Bob get the first and second row respectively of this matrix. We define $(f \circ g)(w_1, \ldots, w_n) = f(g(w_1), \ldots, g(w_n))$, where w_i is the ith block. Here b is called the block length. Thus, DISJ is $\mathsf{NOR} \circ \mathsf{AND}$, with block length 1. Similarly, IP is the block size 1 function $\mathsf{XOR} \circ \mathsf{AND}$. Both have AND as the inner function but, as pointed out earlier, have widely different unbounded error complexity. What makes AND functions, that is functions of the form $f \circ \mathsf{AND}$ with block length 1, difficult?

An important step towards understanding this was taken in the works of Sherstov [32, 33] and Shi and Zhu [37]. These papers reduced the task of proving lower bounds on the cost of both (quantum) bounded error and weakly unbounded error protocols for functions of the form $f \circ \mathsf{AND}$ 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 at least two directions. One was the adaptation of the technique to multi-party communication complexity in [10, 12, 24, 11], resulting in the first super-polynomial lower bounds for Disjointness in the hard NOF model. Using even more powerful approximation theoretic tools for polynomials, Sherstov [35] significantly improved these bounds. In another direction, Razborov and Sherstov [30] and Sherstov [34] further demonstrated the power of these dual polynomial based techniques by analyzing the unbounded error complexity of $f \circ \mathsf{AND}$ when f is a certain AC^0 function or it is symmetric. In short, dual polynomial techniques provide a systematic way of analyzing the communication complexity of AND functions. Besides these impressive developments, this approach relates to research on approximation theoretic questions on boolean functions, that are of independent interest (see for example [9, 38]).

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 (private coin) complexity is only $O(\log n)$, while its unbounded error complexity is just $O(1)$. In fact, in some contexts as discussed later in this work, proving even PP lower bounds for XOR functions seems more challenging than proving lower bounds for AND functions. Interestingly, Sherstov [31] used an XOR function introduced by Goldmann, Håstad and Razborov [16], to separate PP^{cc} from UPP^{cc}. Zhang and Shi [41] characterized the bounded error and quantum complexity of all symmetric XOR functions. Recently, Hatami, Hosseini and Lovett [17] nearly characterized the deterministic complexity of all XOR functions. Even more recently, after an initial version of this manuscript containing weaker results was submitted, Hatami and Qian [18] have just reported settling a conjecture of Zhang and Shi [41] on the unbounded error complexity of symmetric XOR functions. Both [41, 18] analyze XOR functions by finding simple reductions to appropriate AND functions. While such arguments are short, as commented by Ada et al. [2], it seems they do not provide new insights and techniques that can be applied more broadly to XOR functions.

In this work, we develop a dual polynomial based technique for analyzing XOR functions.

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^{cc} and UPP^{cc} via an XOR function. We further provide a new proof of the characterization of Zhang and Shi [41] 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 [25], as far we know, and as expressed in Ada et al. [2], it has not been used before this work in deriving explicit lower bounds on communication complexity.

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 [40]. Further, we characterize the approximate spectral norm of symmetric functions, confirming a conjecture of Ada et al. [2], which has various consequences (cf. [2]). We feel that these developments exhibit the potential of the dual polynomial based technique for proving lower bounds against XOR functions in general (that are not necessarily symmetric).

1.1 Our Results

In this section, we outline our main results.

1.1.1 Polynomial complexity measures of symmetric functions

In this section, we outline results we obtain by amplifying hardness of functions using the method of lifting as defined in Krause and Pudlák. We list applications of this 'hardness amplification' to symmetric functions, which resolve conjectures by Ada et al. [2] and Zhang [40].

For any function $f : \{-1,1\}^n \to \{-1,1\}$, let $f = \sum_{S \subseteq [n]} c_S \prod_{i \in S} x_i$ be the unique multilinear expansion of f. Define the *weight* of f, denoted by $wt(f)$ to be $\sum_{S \subseteq [n]} |c_S|$. [2]

[2]Note that this notion coincides with $||\hat{f}||_1$, the spectral norm of f. However, for the purposes of this paper, we shall

Definition 1.1 (Approximate weight)**.** *Define the* ϵ-approximate weight *of a function* $f : \{-1,1\}^n \to \{-1,1\}$*, denoted by* $wt_\epsilon(f)$ *to be the weight of a minimum weight polynomial such that for all* $x \in \{-1,1\}^n$*,* $|p(x) - f(x)| < \epsilon$*.* [3]

Definition 1.2. *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 1.3 (Margin)**.** *The margin of a function* $f : \{-1,1\}^n \to \{-1,1\}$ *is defined as follows.*

$$m(f) \triangleq \max_{p:wt(p)\leq 1} \left(\min_{x\in\{-1,1\}^n} p(x)f(x) \right)$$

Here, the maximum is only taken over those polynomials p *which sign represent* f *everywhere.*

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

Theorem 1.4. *Let* $F : \{-1,1\}^n \to \{-1,1\}^n$ *be any symmetric function.*

1. *If* $r(F) \geq 5$*, then there exists a universal constant* $c_1 > 0$ *such that*

$$\log(wt_{1/3}(F)) \geq c_1 \cdot r(F).$$

2. *If* $k = \deg_{oe}(F) \geq 16$*, then there exists a universal constant* c_2 *such that*

$$\mathrm{mon}_{\pm}(F) \geq 2^{c_2 \cdot k/\log(n/k)}$$

3. *If* $k = \deg_{oe}(F) \geq 16$*, then there exists a universal constant* c_3 *such that*

$$m(F) \leq \frac{1}{2^{c_3 \cdot k/\log(n/k)}}$$

We also use Part 1 of Theorem 1.4, to prove the following theorem, posed as a conjecture by Ada et al. [2].

Theorem 1.5 (Conjecture 1 in [2])**.** *For any symmetric function* $F : \{-1,1\}^n \to \{-1,1\}$*, there exist universal constants* $c_0, c_1 > 0$ *such that*

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

This has several consequences (cf. [2]), which we do not state here. We also resolve the following conjecture by Zhang [40].

Theorem 1.6 (Conjecture 1 in [40])**.** *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* $\log^{O(1)} n$*.*

1.1.2 PP complexity

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

Our main tool for analyzing the discrepancy of XOR functions is a tight relationship (upto constant factors) between $\mathrm{disc}(f \circ \mathsf{XOR})$ and $m(f)$. We derive this using linear programming duality.

Theorem 1.7 (Polynomial Margin-Discrepancy theorem)**.** *Let* $f \to \{-1,1\}^n \to \{-1,1\}$*.*

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

use the former notation.

[3] This notion coincides with the notion of the ϵ-*approximate spectral norm* of f, denoted by $||\hat{f}||_{1,\epsilon}$, as defined in [2].

The proof of Theorem 1.7 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 lower bounding the discrepancy of $f \circ \mathsf{XOR}$ with 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 [32]. This theorem has yielded a methodical way of proving discrepancy bounds for $f \circ \mathsf{PM}$ by lower bounding the *sign degree* of f, where PM denotes the pattern matrix gadget, and is defined formally in Section 2. This has led to much progress in understanding the communication complexity of AND functions (for example, [10, 12, 33, 34]). 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.

The first 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 Shi and Zhang.

Theorem 1.8. *Let $F : \{-1,1\}^{4n} \to \{-1,1\}$ be any symmetric function, and let $r \geq 4$ be its odd-even degree. Then, there exists a universal constant $c > 0$ such that $\mathsf{PP}(F \circ \mathsf{XOR}) \geq cr/\log(n/r)$ where $\mathsf{PP}(F \circ \mathsf{XOR})$ denotes the PP complexity of $F \circ \mathsf{XOR}$.*

To prove the above, Theorem 1.7 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^{cc} from UPP^{cc}. We do this by proving that an XOR function, almost identical to the GHR function (cf. [16]) has exponentially small discrepancy. It is well known that this function has very efficient UPP protocols. We define the GHR function formally in Section 2.4.

Theorem 1.9.

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

1.1.3 BPP complexity

Using linear programming duality and the generalized discrepancy method (Theorem 2.20), we give a simple alternate proof of the following result from [25].

Theorem 1.10. *For any function $f : \{-1,1\}^n \to \{-1,1\}$, there exists a universal constant $c > 0$ such that*

$$R_{1/3}(f \circ \mathsf{XOR}) \geq c \log\left(wt_{1/3}(f) - 4\right).$$

Remark 1.11. *In fact, lower bounds on $wt_{1/3}(f)$ yield lower bounds on the bounded error quantum communication complexity of $f \circ \mathsf{XOR}$).*

Although Theorem 1.10 was known from [25], to the best of our knowledge, ours is the first work to use this technique to prove lower bounds for explicit functions.

Using Part 1 of Theorem 1.4 in conjunction with Theorem 1.10 provides an alternate proof of the following result of Zhang and Shi [41].

Theorem 1.12 ([41])**.** *Let $F : \{-1,1\}^n \to \{-1,1\}$ be any symmetric function. Then, $R_{1/3}(F \circ \mathsf{XOR}) = \Omega(r(F))$.*

Blais et al. [4] also provided an alternate proof to Theorem 1.12 by showing a lower bound on the information complexity of symmetric XOR functions (this however, does not imply quantum lower bounds).

1.1.4 UPP complexity

We consider the UPP complexity $f \circ \mathsf{XOR}$ when f is symmetric and periodic. More precisely,

Definition 1.13 (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.*

We now state our main result regarding unbounded error communication below:

Theorem 1.14. *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,*

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

where $\mathsf{UPP}(f)$ denotes the unbounded error communication complexity of f.

Remark 1.15. *A very recent result of Hatami and Qian [18] subsumes Theorem 1.14. However, their result is based on a simple reduction to symmetric AND functions, whose unbounded error complexity has been tightly characterized by Sherstov [34] using sophisticated tools from approximation theory. Our result, on the other hand, is based on first principles using Fourier analysis of boolean functions.*

The above implies that the XOR function corresponding to a symmetric and periodic f with period $O(n^{1/2-\epsilon})$, for some constant $\epsilon > 0$, has unbounded-error complexity $n^{\Omega(1)}$ as long as f is neither constant nor Parity nor its complement.

A well known consequence of proving unbounded error lower bounds against f is a lower bound for the size of depth-2 circuits of the form $\mathsf{THR} \circ \mathsf{L}_{\mathsf{comm}}$ computing f where $\mathsf{L}_{\mathsf{comm}}$ denotes the class of functions with low deterministic communication complexity. As a result, Theorem 1.14 implies that in particular, $\mathsf{MOD}_m^A \circ \mathsf{XOR}$ is not in polynomial sized $\mathsf{THR} \circ \mathsf{SYM}$ circuits, which we formally state in Theorem 6.11. This generalizes a result of Zhang [40] and Krause and Pudlák [22] who showed, among other things, that $\mathsf{MOD}_p^{\{0\}}$ cannot be computed by polynomial sized Threshold of Parity circuits.

1.2 Proof outline

Our proof strategy is depicted in Figure 1. First, we use a nice idea due to Krause and Pudlák [22], 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 easily adapted to show a more general result. In particular, our Lemma 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 1.7 and 1.10, 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 [21], 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) that are relevant for us, the complexity of g is upper bounded by that of h. We observe (Lemma 3.5 and Lemma 3.4) that if f is a symmetric (linear threshold) function, then there exists a symmetric (linear threshold) function F such

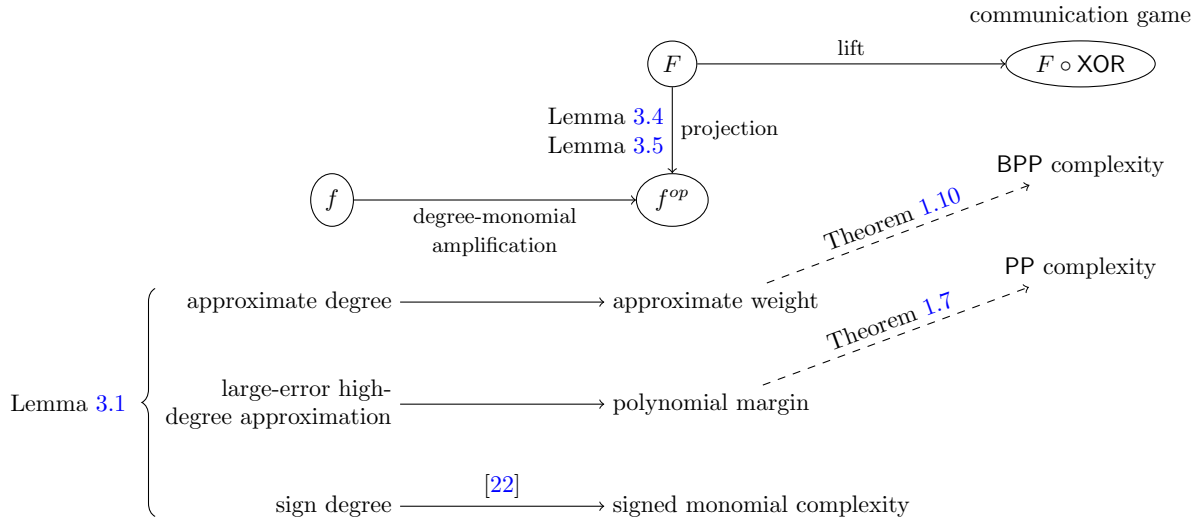

Figure 1: General framework

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.

1.2.1 PP complexity

We prove two main results regarding PP complexity by upper bounding margin complexity. The first is to reprove an exponential separation of PP protocols from those of 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. However, precisely this feature of F makes it difficult to prove a strong PP lower bound. Goldmann et al. [16] used an ingenious specialized argument directly establishing that the discrepancy is small. We, on the other hand, use Theorem 1.7 which directs us in proving that F must have small margin complexity. The challenge here is to prove a strong unrestricted degree margin complexity 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 2.5, 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 1 to show that f^{op} has a small upper bound on the (unrestricted degree) margin complexity. We then use our monomial projection lemma for threshold functions, Lemma 3.4, 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 the UTHR is needed only to provide an *explicit* exponential separation of PP and UPP.

As a second application of our framework to PP complexity, we prove Theorem 1.8, which states that the PP complexity of $F \circ \mathsf{XOR}$ is essentially the odd-even degree of F when it is symmetric. The main challenge here is to work with the notion of odd-even degree, which has no immediate algebraic interpretation as opposed to sign degree. Lemma 3.5 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 polynomial hardness amplification Lemma 3.1 shows that the margin of f^{op} must be small if the base function f has large sign degree.

1.2.2 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 2.11. 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 1.1. In Section 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 1.10 that $\log wt_{1/3}(F)$ is a lower bound on the bounded error communication complexity of $F \circ \mathsf{XOR}$.

The general framework of Figure 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 3.5 provides such a monomial projection in which the combinatorial quantity $r(F)$ corresponds to another combinatorial quantity $\Gamma(f)$, which is defined in Section 2. Paturi's Theorem [27] shows that $\Gamma(f)$ characterizes the approximate degree of f. The polynomial hardness amplification of Figure 1, via Lemma 3.1, implies that f^{op}, and therefore F, has large approximate weight. This already proves Theorem 1.5 which was conjectured by Ada et al. [2]. Moreover, Theorem 1.10 implies the hardness of $F \circ \mathsf{XOR}$ against bounded error protocols.

1.2.3 UPP complexity

We remark here that, although a very recent independent result of Hatami and Qian [18] subsumes 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 [34] on symmetric AND functions. Interestingly, our UPP lower bounds are not obtained via linear programming duality, as opposed to our PP and BPP lower bounds.

The starting point of our work in proving UPP lower bounds is a modification of Forster's theorem [14] by Forster et al. [15] who relate 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. Refer to Theorem 2.23 for details. We then note in Lemma 2.24 that the spectral norm of $f \circ \mathsf{XOR}$ is just a scaling 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. Thus, one cannot use Theorem 2.23 directly. Next, we prove in Theorem 6.3 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 6.3. 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. 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 (Theorem 1.14).

2 Preliminaries

We provide the necessary preliminaries in this section.

Note that in the following definitions, we interchangeably use the view of the input variables being $\{-1,1\}$ valued, and $\{0,1\}$ valued. For most of our results regarding the discrepancy of XOR functions, we view the input variables as $\{-1,1\}$ valued, whereas we view the inputs as $\{0,1\}$ valued while dealing with the unbounded error model. In general, 0 corresponds to 1, and 1 corresponds to -1 in the two

views. We denote the *Hamming weight* of a string $x \in \{0,1\}^n$ ($\{-1,1\}^n$) to be the number of variables set to 1 (-1) in x.

2.1 Types of functions

A function $f : \{-1,1\}^n \to \{-1,1\}$ is called 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. The value taken by a symmetric function on an input 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 such that there are i many variables in x taking the value -1. 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 $\deg_{oe}(f)$, to be $|i \in \{0,1,\ldots,n-2\} : D_f(i) \neq D_f(i+2)|$.

Definition 2.1 (XOR functions)**.** *A function $F : \{0,1\}^n \times \{0,1\}^n \to \{-1,1\}$ is said to be an* XOR *function if there exists a function $f : \{0,1\}^n \to \{0,1\}$ such that $F(x_1,\ldots,x_n,y_1,\ldots,y_n) = f(x_1 \oplus y_1,\ldots,x_n \oplus y_n)$ for all $x_1,\ldots,x_n,y_1,\ldots y_n \in \{0,1\}$. We use the notation $F = f \circ$ XOR.*

Definition 2.2 (Threshold functions)**.** *Define a function $f : \{0,1\}^n \to \{-1,1\}$ to be a linear threshold function if there exist integer weights $a_1,\ldots,a_n$ such that for all inputs $x = (x_1,\ldots,x_n) \in \{0,1\}^n$, $f(x) = sgn\left(\sum_{i=1}^n a_i x_i\right)$. Let* THR *denote the class of all such functions. Let* MAJ *denote the class of linear threshold functions whose weights are polynomially bounded in n.*

Definition 2.3 (Universal threshold)**.** *Define a class of threshold functions, $U_{l,k} : \{\{-1,1\}^k\}^l \to \{0,1\}$ 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 2.4 (Minsky and Papert [26])**.** *$U_{l,k}$ 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.

2.2 Fourier analysis

Consider the vector space of functions from $\{0,1\}^n$ to $\mathbb{R}$, equipped with the following inner product.

$$\langle f,g\rangle = \mathbb{E}_{x\in\{0,1\}^n} f(x)g(x) = \frac{1}{2^n}\sum_{x\in\{0,1\}^n} f(x)g(x)$$

Define characters χ_S for every $S \subseteq [n]$ by $\chi_S(x) = (-1)^{\sum_{i\in S} x_i}$. The set $\{\chi_S : S \subseteq [n]\}$ forms an orthonormal basis for this vector space. Thus, every $f : \{0,1\}^n \to \mathbb{R}$ can be uniquely written as $f = \sum_{S\subseteq[n]} \widehat{f}(S)\chi_S$ where

$$\widehat{f}(S) = \langle f,\chi_S\rangle = \mathbb{E}_{x\in\{0,1\}^n} f(x)\chi_S(x) \tag{1}$$

2.3 Polynomials

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 [37, 33, 38] for example. Note that we do not restrict the weight of the approximating polynomial in this case.

$$\varepsilon_d(f) \triangleq \min_{p:deg(p)\leq d}\left(\max_{x\in\{-1,1\}^n} |p(x) - f(x)|\right) \tag{2}$$

Sherstov [36] proved that there exists a linear threshold function which cannot be approximated well, even by large degree polynomials.

Theorem 2.5 ([36], 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.

Definition 2.6 (Approximate degree). *For any function $f : \{-1,1\}^n \to \{-1,1\}$ and polynomial $p : \{-1,1\}^n \to \mathbb{R}$, we say that p approximates f to error ϵ if for all $x \in \{-1,1\}^n$, $|p(x) - f(x)| \leq \epsilon$. The ϵ-approximate degree of f, denoted $\widetilde{\deg}_\epsilon(f)$ is the minimum degree of a polynomial p which approximates f to error ϵ.*

Definition 2.7 (Signed monomial complexity). *The* signed monomial complexity *of a function $f : \{-1,1\}^n \to \{-1,1\}$, denoted by $\mathrm{mon}_\pm(f)$ is the minimum number of monomials required by a polynomial p to sign represent f on all inputs.*

Note that the signed monomial complexity of a function f exactly corresponds to the minimum size Threshold of Parity circuit computing it.

Theorem 2.8 ([40]). *Let $f : \{-1,1\}^n \to \{-1,1\}$ be a symmetric boolean function such that $\deg_{oe}(f) = \log^{O(1)} n$. Then, f can be computed by a quasi-polynomial size Threshold of Parity circuit.*

The following is a result by Paturi [27] which gives us tight bounds on the approximate degree of symmetric functions.

Theorem 2.9 ([27]). *For any symmetric function $f : \{0,1\}^n \to \{-1,1\}$, define the quantity $\Gamma(f) = \min\{|2k - n + 1| : D_f(k) \neq D_f(k+1)$ and $0 \leq k \leq n-1\}$. Then,*

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

Definition 2.10. *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 2.11 (Threshold weight). *Define the* threshold weight *of a function $f : \{-1,1\}^n \to \{-1,1\}$, denoted by $wt_\pm(f)$ 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 [21], [33], [8]), 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 2.11 because of its following relationship with the polynomial margin, which can be easily verified.

Lemma 2.12. *For any function $f : \{-1,1\}^n \to \{-1,1\}$,*

$$m(f) = \frac{1}{wt_\pm(f)}$$

The following theorem was proved by Ada et al. [2], which characterizes the weight of a symmetric function.

Theorem 2.13 ([2]). *For any symmetric function $f : \{-1,1\}^n \to \{-1,1\}$,*

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

2.4 Communication complexity

We now recall some notions from communication complexity.

In the models of communication of our interest, two players, say Alice and Bob, are given inputs $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ for some finite input sets $\mathcal{X}, \mathcal{Y}$, they have access to *private* randomness and they wish to compute a given function $f : \mathcal{X} \times \mathcal{Y} \to \{-1, 1\}$. Unless mentioned otherwise, we use $\mathcal{X} = \mathcal{Y} = \{0, 1\}^n$. Alice and Bob 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 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 [23]. Define the following measures of complexity of f.

$$\mathsf{PP}(f) = \min_{\epsilon} \left(R_\epsilon(f) + \log\left(\frac{1}{\epsilon}\right)\right)$$

and

$$\mathsf{UPP}(f) = \min_{\epsilon}(R_\epsilon(f)).$$

The latter quantity was introduced by Paturi and Simon [28], and we call it the unbounded error communication complexity of f. The former adds a penalty term depending on the advantage, and was proposed by Babai et al. [3]. We refer to this cost as the weakly-unbounded error communication complexity of f. These measures give rise to the following communication complexity classes [3].

Definition 2.14.

$$\mathsf{PP}^{cc}(f) \equiv \{f : \mathsf{PP}(f) = \text{polylog}(n)\}$$
$$\mathsf{UPP}^{cc}(f) \equiv \{f : \mathsf{UPP}(f) = \text{polylog}(n)\}$$

Define the discrepancy of a rectangle $S \times T$ under a distribution λ on $\{-1, 1\}^n \times \{-1, 1\}^n$ as follows.

Definition 2.15 (Discrepancy)**.**

$$\text{disc}_\lambda(S \times T, f) = \sum_{(x,y)\in S\times T} f(x, y)\lambda(x, y)$$

The discrepancy of f under a distribution λ is defined as

$$\text{disc}_\lambda(f) = \max_{S\subseteq[n],T\subseteq[n]} \text{disc}_\lambda(S \times T, f)$$

and the discrepancy of f is defined to be

$$\text{disc}(f) = \min_{\lambda} \text{disc}_\lambda(f)$$

Klauck [20] proved that discrepancy and PP complexity are equivalent notions.

Theorem 2.16 (Klauck [20])**.** *For any function $f : \{-1, 1\}^n \to \{-1, 1\}$,*

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

In [16], Goldmann et al. exhibited a distribution under which the one way communication complexity of $U_{4n,n} \circ \mathsf{XOR}$ is large. Sherstov [31] noted that the same proof can be used to show that $\text{disc}(U_{4n,n} \circ \mathsf{XOR}) \leq O\left(\frac{\sqrt{n}}{2^{n/2}}\right)$.

Remark 2.17. *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 [33]. 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.

Theorem 2.18 ([33] Thm 1.5)**.** *Let $F = f \circ \mathsf{PM}$ for a given function $f : \{0,1\}^n \to \{-1,1\}$. Then*

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

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 2.19. *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 paper, we only consider the case when each block is of size 2.*

The following theorem, first proposed by Klauck [20], provides a tool for proving bounded error communication lower bounds for functions. Its proof may be found in [11, 12], for example.

Theorem 2.20 (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.*

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

For notational convenience, we use the notation $U(f)$ to represent $\mathsf{UPP}(f \circ \mathsf{XOR})$. 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.

Paturi and Simon [28] showed an equivalence between $\mathsf{UPP}(f)$ and a quantity called the sign rank of M_f where M_f denotes the communication matrix of f. Define the sign rank of a real matrix M with no 0 entries as follows.

Definition 2.21 (Sign Rank)**.**

$$sr(M) = \min_A \{ rk(A) : sgn(A_{ij}) = sgn(M_{ij}) \}$$

We overload notation and use $sr(f)$ to denote $sr(M_f)$.

Theorem 2.22 (Paturi and Simon [28])**.**

$$\mathsf{UPP}(f) = \log sr(A) \pm O(1)$$

Finding an explicit matrix with superlogarithmic sign rank remained a challenge until a breakthrough result of Forster [14], who proved that the sign rank of any $N \times N$ Hadamard matrix is at least $\Omega(\sqrt{N})$. This implied an asymptotically tight lower bound for the unbounded error communication complexity of the inner product (modulo 2) function. We use a generalization of Forster's theorem by Forster et al. [15].

Theorem 2.23 (Forster et al. [15])**.** *Let $M_{m \times N}$ be a real matrix with no 0 entries. Then,*

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

where $||M||$ denotes the spectral norm of the matrix M.

Thus, it suffices to prove upper bounds on the spectral norm of the communication matrix of a function in order to prove unbounded error lower bounds for that function. Let us now state a lemma characterizing the spectral norm of the communication matrix of XOR functions.

Lemma 2.24 (Folklore). *Let $f : \{0,1\}^n \times \{0,1\}^n \to \mathbb{R}$ be any real valued function and let M denote the communication matrix of $f \circ \mathsf{XOR}$. Then,*

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

Although this is farily 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 $z_T \in \{-1,1\}^{2^n}$ defined by $(z_T)_y = \chi_T(y)$.

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 z_T is an eigenvector of M with eigenvalue $2^n\widehat{f}(T)$. Consider the xth coordinate of Mz_T.

$$\begin{aligned}(Mz_T)_x &= \sum_{y \in \{0,1\}^n} \sum_{S \subseteq [n]} \widehat{f}(S)\chi_S(x)\chi_S(y)\chi_T(y) \\ &= \sum_S \widehat{f}(S)\chi_S(x) \sum_{y \in \{0,1\}^n} \chi_{S \triangle T}(y) \\ &= \widehat{f}(T)\chi_T(x)2^n\end{aligned}$$

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.

□

3 Lifting functions

In this section we first show how we 'lift' functions as introduced by Krause and Pudlák [22]. 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. [2] and Zhang [40].

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.

$$f^{op}(x_1, \ldots, x_n, y_1, \ldots, y_n, z_1, \ldots, z_n) = f(u_1, \ldots, u_n) \tag{3}$$

where for all i, $u_i = (x_i \wedge z_i) \vee (y_i \wedge \bar{z}_i)$. 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 [22].

The following lemma translates hardness properties of f into other hardness properties of f^{op}. The proof of this lemma is based on ideas from [22].

Lemma 3.1. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be any function.*

1. *If $\varepsilon_d(f) > 1 - 2^{-d}$ for some $d \geq 2$, then $m(f^{op}) \leq 2^{-c'd}$ for any constant $0 < c' < 1 - \frac{1}{d}$.*
2. *$\mathrm{mon}_{\pm}(f^{op}) \geq 2^{\deg_{\pm}(f)}$.*
3. *$wt_{1/3}(f^{op}) \geq 2^{c \cdot \widetilde{\deg}_{2/3}(f)}$ for any constant $c < 1 - 1/\widetilde{\deg}_{2/3}(f)$.*

Proof. We first prove part 1.

Let p be a polynomial of weight 1 representing f^{op} with margin at least $\frac{1}{2^{c'd}}$ for a fixed positive constant $0 < c' < 1 - \frac{1}{d}$, 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 $|S| \geq d$,

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

Now since $wt(p) = 1$, we have

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

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 $\frac{1}{2^{c'd}} - \frac{1}{2^d}$.
- 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 $\frac{1}{2^{c'd}} - \frac{1}{2^d}$. This leads to a contradiction when $c' < 1 - \frac{1}{d}$, since we assumed that $\varepsilon_d(f) > 1 - \frac{1}{2^d}$.

We omit the proofs of the other two statements as they follow along extremely similar lines. □

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 3.2 (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 $x_1, \ldots, x_m$.*

The following lemma is an easy consequence of definitions.

Lemma 3.3. *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*

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

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 [21].

Lemma 3.4. *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 $f(x) = sgn\left(\sum_{i=1}^{n} w_i x_i\right)$. Note that

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

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

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

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

Lemma 3.5. *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.

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

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 $x_1,\ldots,x_n,y_1,\ldots,y_n,z_1,\ldots,z_n$ and $n-b$ relevant variables with value 1. Say $(x_1,\ldots,x_n,y_1,\ldots,y_n,-x_1z_1,\ldots,-x_nz_n,y_1z_1,\ldots,y_nz_n)$ contains a many -1's. Then,

$$\begin{aligned} 4n - 2a = \sum_{i=1}^{n} x_i + y_i - x_i z_i + y_i z_i = \sum_{i=1}^{n} x_i(1-z_i) + y_i(1+z_i) = 2n - 4b \\ \implies a = 2b + n \end{aligned}$$

Thus,

$$\begin{aligned} g(x_1,\ldots,x_n,y_1,\ldots,y_n,z_1,\ldots,z_n) &= D_F(2b+n) = D_f(b) \\ &= f^{op}(x_1,\ldots,x_n,y_1,\ldots,y_n,z_1,\ldots,z_n) \end{aligned}$$

The last equality follows from Equation 3. □

In fact, the proof of Lemma 3.5 can be seen to imply the following lemma.

Lemma 3.6. *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 $b \in \{0,1,\ldots,n\}$, F takes the value $D_f(b)$, and F takes arbitrary values on inputs of Hamming weight not in $\{2b+n : b \in \{0,1,\ldots,n\}\}$. Then, f^{op} is a monomial projection of F.*

3.3 Consequences for symmetric functions

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

We first prove Theorem 1.4.

Proof of Theorem 1.4.

- 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 $\log wt_{1/3}(F') \geq c'r$ 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 3.5, 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 2.9, $\widetilde{\deg}_{2/3}(f) = \Theta(r)$.

  Using Lemma 3.1 and Lemma 3.3, we obtain that there exists a universal constant $c_1 > 0$ such that

$$\log(wt_{1/3}(F)) \geq \log(wt_{1/3}(F')) \geq \log(wt_{1/3}(f^{op})) \geq c_1 r \tag{4}$$

- Consider any symmetric function $F : \{-1,1\}^{4n} \to \{-1,1\}$ such that $\deg_{oe}(F) \geq 4j$ 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 3.5, Lemma 3.1 and Lemma 3.3, 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 1.5, resolving a conjecture of Ada et al. [2].

Proof of Theorem 1.5. It follows as a direct consequence of Part 1 of Theorem 1.4 and the upper bound in Theorem 2.13. □

Finally, we prove Theorem 1.6 here, settling a conjecture of Zhang [40].

Proof of Theorem 1.6. The upper bound follows from Theorem 2.8. It suffices to show a lower bound for when $\deg_{oe}(f) \geq 16$. The lower bound follows from Part 2 of Theorem 1.4 in this case.

□

4 Discrepancy of XOR functions

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

4.1 Margin-discrepancy equivalence

In this section, we prove Theorem 1.7, 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 1.7. 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 $m(F) \leq 4\mathrm{disc}(F)$. Let λ denote a distribution under which $\mathrm{disc}_\lambda(F) = \mathrm{disc}(F)$, 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.

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

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

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

Let us first write a linear 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

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

Thus,

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

Here, the first inequality follows from the definition of $\mu^{\oplus}$, and the following equality follows from Lemma 2.24. This proves the claim.

□

We remark here that Linial and Shraibman [24] 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 base function. However, since we do not use this notion in the rest of this paper, 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 follows.

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

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 [24] showed that the margin of a sign matrix is equivalent to its discrepancy up to a constant factor.

Theorem 4.1 ([24] Thm 3.1)**.** *For every sign matrix A,*

$$\mathrm{disc}(A) \leq m(A) \leq 8\mathrm{disc}(A)$$

We now note that Theorem 1.7 implies the first inequality of Theorem 4.1 for the special case of XOR functions.

Claim 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 $wt(p') = 1$, $||u_T||_2 = ||v_T||_2 = 1$. 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,

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

□

4.2 A new separation of PP^{cc} from UPP^{cc}

In this section, we show here how to obtain an alternate proof that the GHR function has large PP complexity. It is well known that $\mathsf{GHR} \in \mathsf{UPP}^{cc}$.

Proof of Theorem 1.9. Theorem 2.5 and Lemma 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 and Lemma 3.4 then show existence of a linear threshold function $f' : \{-1,1\}^{4n} \to \{-1,1\}$ such that $m(f') \leq 2^{-cn}$. Using Theorem 1.7 and Theorem 2.16, 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 2.4, 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 2.4, 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 1.7 and Theorem 2.16, we have

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

□

4.3 XOR is harder than PM

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 inverse exponential discrepancy, but $\oplus \circ \mathsf{XOR}$ has extremely large discrepancy.

Theorem 4.3. *Let $f : \{-1,1\}^n \to \{-1,1\}$. Then,*

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

Proof. Consider $f \circ \mathsf{PM}$ and substitute $d = n$ in Theorem 2.18 to obtain

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

By Theorem 1.7, $\mathrm{disc}(f \circ \mathsf{XOR}) < \delta \implies m(f) < 4\delta$. 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,

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

□

4.4 Symmetric functions with large odd-even degree

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

Proof of Theorem 1.8. Using Theorem 1.7 and Part 3 of Theorem 1.4, 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 1.8. □

5 Bounded error communication complexity of XOR functions

In this section, we analyze the bounded error communication complexity of XOR functions.

Proof of Theorem 1.10. We write a linear program which captures the best error a weight w polynomial can achieve in approximating a given function f.

Variables	$\epsilon, \{\alpha_S : S \subseteq [n]\}$		
Minimize	ϵ		
s.t.	$\left\lvert f(x) - \sum_{S \subseteq [n]} \alpha_S \chi_S(x) \right\rvert$	$\leq \epsilon$	$\forall x \in \{-1,1\}^n$
	$\sum_{S \subseteq [n]} \lvert\alpha_S\rvert$	$\leq w$	
	$\epsilon \geq 0$		
	$\alpha_S \in \mathbb{R}$		$\forall S \subseteq [n]$

By manipulations similar to those in Section 4.1, we obtain the following dual program.

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

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 $wt_{1/3}(f) = w'$. 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 $\lvert \sum_x \mu'(x)\chi_S(x)\rvert \leq \frac{3}{w'}$ (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

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

This, along with Theorem 2.20 proves the following.

$$R_{7/15}(f \circ \mathsf{XOR}) \geq \log w' - 4$$

By standard error reduction, we obtain Theorem 1.10. □

Using Part 1 of Theorem 1.4 and Theorem 1.10, we obtain a new proof of Theorem 1.12.

6 Sign rank of XOR functions

In this section, we analyze the unbounded error communication complexity of XOR functions.

6.1 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 6.1. *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 [40] 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 [5, 13, 1]. The notation we use is that from [1].

Definition 6.2. *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 6.1.

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.

$$\mathsf{MOD}_m^A(x) = 1 - \frac{2}{m}\sum_{k\in A}\left(\sum_{a=0}^{m-1} \mathsf{EXP}_m^{a,k}(x)\right) = \begin{cases} -1 & |x| \equiv k (\bmod\ m) \text{ for some } k \in A \\ 1 & \text{otherwise} \end{cases} \tag{5}$$

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 (1),

$$\begin{aligned} \widehat{\mathsf{MOD}_m^A}(S) &= \mathbb{E}_{x\in\{0,1\}^n}\left[\mathsf{MOD}_m^A(x)\chi_S(x)\right] \\ &= \mathbb{E}_{x\in\{0,1\}^n}\left[\chi_S(x)\right] - \frac{2}{m}\sum_{k\in A}\sum_{a=0}^{m-1}\mathbb{E}_{x\in\{0,1\}^n}\left[\mathsf{EXP}_m^{a,k}(x)\chi_S(x)\right] \end{aligned} \tag{6}$$

where the second equality follows from Equation (5) and linearity of expectation. Recall from Definition 6.2 that $\mathsf{EXP}_m^{a,b}(x) = \omega^{a((\sum_{j=1}^n x_j)-b)}$. 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 $a \in \{1,\ldots,m-1\}$,

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

Thus, in Equation (6), 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 (6) gives us

$$\left|\widehat{\mathsf{MOD}_m^A}(S)\right| \leq \left|\mathbb{E}_{x\in\{0,1\}^n}\left[\chi_S(x)\right]\right| + \left|\frac{2}{m}\sum_{k\in A}\sum_{a=0}^{m-1}\mathbb{E}_{x\in\{0,1\}^n}\left[\mathsf{EXP}_m^{a,k}(x)\chi_S(x)\right]\right| \tag{7}$$

$$\leq \frac{2(m-1)^2}{m}\cdot\left(\cos\left(\frac{\pi}{2m}\right)\right)^n \leq 2m\left(\cos\left(\frac{\pi}{2m}\right)\right)^n \tag{8}$$

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 (6) 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 (6) and using the above observations, we get

$$\begin{aligned}
\left|\widehat{\mathsf{MOD}_m^A}(\emptyset)\right| &\leq \left|\mathbb{E}_{x\in\{0,1\}^n}\left[\chi_\emptyset(x)\right] - \frac{2}{m}\sum_{k\in A}\mathbb{E}_{x\in\{0,1\}^n}\left[\mathsf{EXP}_m^{0,k}(x)\chi_\emptyset(x)\right]\right| \\
&\quad + \left|\frac{2}{m}\sum_{k\in A}\sum_{a=1}^{m-1}\mathbb{E}_{x\in\{0,1\}^n}\left[\mathsf{EXP}_m^{a,k}(x)\chi_\emptyset(x)\right]\right| && (9) \\
&\leq \left|1 - 2\frac{|A|}{m}\right| + 2m\left(\cos\left(\frac{\pi}{2m}\right)\right)^n && (10) \\
&\leq 1 - \frac{2}{m} + 2m\left(\cos\left(\frac{\pi}{2m}\right)\right)^n && \text{since } A \neq \emptyset, [m]
\end{aligned}$$

□

6.2 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. As an application of this theorem, we show that $\mathsf{UPP}(\mathsf{MOD}_m \circ \mathsf{XOR}) = n^{\Omega(1)}$ for values of odd m up to $O(n^{1/2-\epsilon})$. Theorem 2.22 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 2.23 and Theorem 2.24, we get

Corollary 1. *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,*

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

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

Note that we cannot use the outer function to be MOD_p (for a constant p) in Corollary 1, 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 6.3. *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, $sr(f \circ \mathsf{XOR}) \geq \frac{\delta}{c}$.*

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 1, we have

$$sr(f \circ \mathsf{XOR}) = sr(f' \circ \mathsf{XOR}) \geq \frac{1}{\max\limits_{S\notin\mathcal{S}}\left|\widehat{f'}(S)\right|} \cdot \min_x |f'(x)| \geq \frac{\delta}{c} \tag{11}$$

□

Let us first recall the Complete Quadratic function, whose Fourier coefficients were analyzed by Bruck [6]. Define $\mathsf{CQ} : \{0,1\}^n \to \{-1,1\}$ by

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

Lemma 6.4 ([6])**.** *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 $S \subseteq [n]$.*

Theorem 6.5. *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 6.3, use $\mathcal{S} = \emptyset$. The values obtained using Claim 6.1 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,

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

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

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

Thus, using the equivalence between sign rank and unbounded error communication complexity from Theorem 2.22,

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

□

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 6.3 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 6.8, 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 6.9), 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 1.13.

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 6.6. *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 require the following simple, yet powerful lemma, the proof of which we omit.

Lemma 6.7 (Folklore)**.** *For any functions $f, g : \{0,1\}^n \to \mathbb{R}$*

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

Proof of Lemma 6.6. Since $\mathsf{MOD}^{A'}_{p,n} = \mathsf{MOD}^{A}_{m,n} \oplus \mathsf{MOD}^{A+i}_{m,n}$, applying Lemma 6.7 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 6.8. *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 6.4,
$$U(\mathsf{CQ}) \geq n/2$$
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 6.6 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 regarding unbounded error complexity (Theorem 1.14), which says that any function of the type $\mathsf{MOD}^{A}_{m} \circ \mathsf{XOR}$ for any non-simple A is hard in the unbounded error communication model 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 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 1.14.

Claim 6.9. *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 6.9 implies Theorem 1.14. Recall that Theorem 6.5 gave us

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

This, along with Claim 6.8 and Claim 6.9, 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 6.9.

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^A_m and MOD^a_m 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, $\mathsf{MOD}^A_m \oplus \mathsf{MOD}^{A+p}_m = \mathsf{MOD}^{x\oplus y}_p$. By Lemma 6.6,
   $$U(\mathsf{MOD}^A_{m,n}) \geq \frac{U(\mathsf{MOD}^{x\oplus y}_{p,n-m})}{2}$$
   (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 $\mathsf{MOD}^a_m = \mathsf{MOD}^x_p$.
   (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, $\mathsf{MOD}^b_p = \mathsf{MOD}^A_m \oplus \mathsf{MOD}^{A+1}_m$. Use Lemma 6.6 and conclude
      $$U(\mathsf{MOD}^A_{m,n}) \geq \frac{U(\mathsf{MOD}^b_{p,n-m})}{2}$$
      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 $\mathsf{MOD}^A_m \oplus \mathsf{MOD}^{A+k}_m = \mathsf{MOD}^{x\oplus y}_k$. By Lemma 6.6,
   $$U(\mathsf{MOD}^A_{m,n}) \geq \frac{U(\mathsf{MOD}^{x\oplus y}_{k,n-m})}{2}$$
   By the induction hypothesis, the claim is true for $\mathsf{MOD}^{x\oplus y}_{k,n-m}$. It is easy to see that this implies the claim for $\mathsf{MOD}^A_{m,n}$.
   (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^a_m is the same as MOD^x_k.
   (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 6.6 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 $k \equiv 2 (\mathrm{mod}\ 4)$, 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 6.6 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,

$$U(\mathsf{MOD}_m^a) = U(\mathsf{MOD}_4^X) = \Omega(n)$$

□

6.3 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 upper bounded by essentially $\deg_{oe}(f)$. Our proof follows along the lines of Zhang [40] who shows that a symmetric function with small odd-even degree has a small Threshold of Parity circuit representation.

Theorem 6.10. *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 $wt(q_1 \cdot q_2) \leq wt(q_1) \cdot wt(q_2)$ and $wt(q_1+q_2) \leq wt(q_1)+wt(q_2)$. Thus,

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

Note that all the coefficients of p are integer valued. Thus, the polynomial $p' = \frac{p}{wt(p)}$ is a polynomial of weight 1, which sign represents f with margin at least $\frac{1}{wt(p)}$. By Theorem 1.7 and Theorem 2.16,

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

□

6.4 Circuits

In this section, we show how to obtain a size lower bound on a restricted class of threshold circuits computing $\mathsf{MOD}_m^A \circ \mathsf{XOR}$ for non simple A. Forster et al. [15] noted that sign rank lower bounds also yield lower bounds against $\mathsf{THR} \circ \mathsf{MAJ}$ circuits. In fact, it yields lower bounds for the class $\mathsf{THR} \circ \mathsf{L}_{\mathsf{comm}}$ where $\mathsf{L}_{\mathsf{comm}}$ denotes any gate with low deterministic communication complexity. We show the following.

Theorem 6.11. *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. The rank of the communication matrix of each C gate is at most c, thus the sign rank of a function computed by a $\mathsf{THR} \circ C$ circuit is at most sc, where s is the size of the circuit. Theorem 1.14 and Theorem 2.22 tells us that $sr(\mathsf{MOD}_m^A \circ \mathsf{XOR}) \geq 2^{\Omega\left(\frac{n-km}{jm}\right) - \frac{2j \log j}{m}}$, where $m = j2^k$, and j is either 4 or odd. Thus,

$$sc \geq 2^{\Omega\left(\frac{n-km}{jm}\right) - \frac{2j \log j}{m}} \implies s \geq 2^{\Omega\left(\frac{n-km}{jm}\right) - \frac{2j \log j}{m} - c}$$

□

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 Pudlak [22], and of Zhang [40] 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 $\mathsf{MOD}_m^A \notin \mathsf{THR} \circ \mathsf{PAR}$.

7 Acknowledgements

We thank anonymous reviewers for providing invaluable comments regarding the presentation of parts of this paper. We thank Justin Thaler for pointers regarding the connection between margin and threshold weight, and bringing the recent paper of Hatami and Qian [18] to our notice.

References

[1] Anil Ada, Arkadev Chattopadhyay, Omar Fawzi, and Phuong Nguyen. The NOF multiparty communication complexity of composed functions. *Computational Complexity*, 24(3):645–694, 2015.

[2] Anil Ada, Omar Fawzi, and Hamed Hatami. Spectral norm of symmetric functions. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques - 15th International Workshop, APPROX 2012, and 16th International Workshop, RANDOM 2012, Cambridge, MA, USA, August 15-17, 2012. Proceedings*, pages 338–349, 2012.

[3] László Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory (preliminary version). In *27th Annual Symposium on Foundations of Computer Science, Toronto, Canada, 27-29 October 1986*, pages 337–347, 1986.

[4] Eric Blais, Joshua Brody, and Badih Ghazi. The information complexity of hamming distance. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2014, September 4-6, 2014, Barcelona, Spain*, pages 465–489, 2014.

[5] Jean Bourgain. Estimation of certain exponential sums arising in complexity theory. *Comptes Rendus Mathematique*, 340(9):627–631, 2005.

[6] Jehoshua Bruck. Harmonic analysis of polynomial threshold functions. *SIAM J. Discrete Math.*, 3(2):168–177, 1990.

[7] Harry Buhrman, Nikolay Vereshchagin, and Ronald de Wolf. On computation and communication with small bias. In *Proceedings of the Twenty-Second Annual IEEE Conference on Computational Complexity*, CCC '07, pages 24–32. IEEE Computer Society, 2007.

[8] Mark Bun and Justin Thaler. Hardness amplification and the approximate degree of constant-depth circuits. In *Automata, Languages, and Programming - 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part I*, pages 268–280, 2015.

[9] Mark Bun and Justin Thaler. A nearly optimal lower bound on the approximate degree of ac^0. *CoRR*, abs/1703.05784, 2017.

[10] Arkadev Chattopadhyay. Discrepancy and the power of bottom fan-in in depth-three circuits. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2007), October 20-23, 2007, Providence, RI, USA, Proceedings*, pages 449–458, 2007.

[11] Arkadev Chattopadhyay. *Circuits, Communication and Polynomials*. PhD thesis, McGill University, 2009.

[12] Arkadev Chattopadhyay and Anil Ada. Multiparty communication complexity of disjointness. *Electronic Colloquium on Computational Complexity (ECCC)*, 15(002), 2008.

[13] Arkadev Chattopadhyay, Navin Goyal, Pavel Pudlák, and Denis Thérien. Lower bounds for circuits with mod_m gates. In *47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006), 21-24 October 2006, Berkeley, California, USA, Proceedings*, pages 709–718, 2006.

[14] Jürgen Forster. A linear lower bound on the unbounded error probabilistic communication complexity. In *Proceedings of the 16th Annual IEEE Conference on Computational Complexity, Chicago, Illinois, USA, June 18-21, 2001*, pages 100–106, 2001.

[15] 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 *FST TCS 2001: Foundations of Software Technology and Theoretical Computer Science, 21st Conference, Bangalore, India, December 13-15, 2001, Proceedings*, pages 171–182, 2001.

[16] Mikael Goldmann, Johan Håstad, and Alexander A. Razborov. Majority gates VS. general weighted threshold gates. *Computational Complexity*, 2:277–300, 1992.

[17] H. Hatami, K. Hosseini, and S. Lovett. Structure of protocols for XOR functions. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:44, 2016.

[18] Hamed Hatami and Yingjie Qian. The unbounded-error communication complexity of symmetric xor functions. *Arxiv*, 2017.

[19] Bala Kalyanasundaram and Georg Schnitger. The probabilistic communication complexity of set intersection. *SIAM J. Discrete Math.*, 5(4):545–557, 1992.

[20] Hartmut Klauck. Lower bounds for quantum communication complexity. *SIAM J. Comput.*, 37(1):20–46, 2007.

[21] Matthias Krause. On the computational power of boolean decision lists. *Computational Complexity*, 14(4):362–375, 2006.

[22] 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.

[23] Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, 1997.

[24] Troy Lee and Adi Shraibman. Disjointness is hard in the multiparty number-on-the-forehead model. *Computational Complexity*, 18(2):309–336, 2009.

[25] Troy Lee and Adi Shraibman. Lower bounds in communication complexity. *Foundations and Trends in Theoretical Computer Science*, 3(4):263–398, 2009.

[26] Marvin Minsky and Seymour Papert. *Perceptrons - an introduction to computational geometry*. MIT Press, 1987.

[27] 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, May 4-6, 1992, Victoria, British Columbia, Canada*, pages 468–474, 1992.

[28] Ramamohan Paturi and Janos Simon. Probabilistic communication complexity. *J. Comput. Syst. Sci.*, 33(1):106–123, 1986.

[29] Alexander A. Razborov. On the distributional complexity of disjointness. *Theor. Comput. Sci.*, 106(2):385–390, 1992.

[30] Alexander A. Razborov and Alexander A. Sherstov. The sign-rank of AC^0. *SIAM J. Comput.*, 39(5):1833–1855, 2010.

[31] Alexander A. Sherstov. Halfspace matrices. *Computational Complexity*, 17(2):149–178, 2008.

[32] Alexander A. Sherstov. Separating AC^0 from depth-2 majority circuits. *SIAM J. Comput.*, 38(6):2113–2129, 2009.

[33] Alexander A. Sherstov. The pattern matrix method. *SIAM J. Comput.*, 40(6):1969–2000, 2011.

[34] Alexander A. Sherstov. The unbounded-error communication complexity of symmetric functions. *Combinatorica*, 31(5):583–614, 2011.

[35] Alexander A. Sherstov. Communication lower bounds using directional derivatives. *J. ACM*, 61(6):34:1–34:71, 2014.

[36] Alexander A. Sherstov. On multiparty communication with large versus unbounded error. *Electronic Colloquium on Computational Complexity (ECCC)*, 2016.

[37] Yaoyun Shi and Yufan Zhu. Quantum communication complexity of block-composed functions. *Quantum Information & Computation*, 9(5):444–460, 2009.

[38] Justin Thaler. Lower bounds for the approximate degree of block-composed functions. In *43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy*, pages 17:1–17:15, 2016.

[39] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *Proceedings of the 11h Annual ACM Symposium on Theory of Computing, April 30 - May 2, 1979, Atlanta, Georgia, USA*, pages 209–213, 1979.

[40] Zhi-Li Zhang. Complexity of symmetric functions in perceptron-like models. Master's thesis, University of Massachusetts at Amherst, 1992.

[41] Zhiqiang Zhang and Yaoyun Shi. Communication complexities of symmetric XOR functions. *Quantum Information & Computation*, 9(3):255–263, 2009.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il