A Short List of Equalities Induces Large Sign Rank

Arkadev Chattopadhyay and Nikhil S. Mande
School of Technology and Computer Science
TIFR
Mumbai, India
Email: {arkadev.c, nikhil.mande}@tifr.res.in

***Abstract*—We exhibit a natural function F, that can be computed by just a linear sized decision list of 'Equalities', but whose sign rank is exponentially large. This yields the following two new unconditional complexity class separations. The first is an exponential separation between the depth-two threshold circuit classes Threshold-of-Majority and Threshold-of-Threshold, answering an open question posed by Amano and Maruoka [MFCS '05] and Hansen and Podolskii [CCC '10]. The second separation shows that the communication complexity class P^{MA} is not contained in UPP, strongly resolving a recent open problem posed by Göös, Pitassi and Watson [ICALP '16]. In order to prove our main result, we view F as an XOR function and develop a technique to lower bound the sign rank of such functions. This requires novel approximation theoretic arguments against polynomials of *unrestricted* degree. Further, our work highlights for the first time the class 'decision lists of exact thresholds' as a common frontier for making progress on longstanding open problems in Threshold circuits and communication complexity.**

***Keywords*-sign rank; XOR functions; communication complexity; circuit complexity; approximation theory**

I. Introduction

Sign rank is a delicate but powerful notion, which has a matrix rigidity-like flavor. It was introduced in the seminal work of Paturi and Simon [41]. The sign rank of a $\{-1, 1\}$ valued matrix M is defined to be the minimum rank of a real valued matrix each of whose entries agrees in sign with the corresponding entry of M. Sign rank has found numerous applications in computer science in areas like communication complexity, boolean circuit complexity, and computational learning theory. Paturi and Simon showed that the logarithm of the sign rank of a (communication) matrix is essentially equivalent to the unbounded-error 2-party communication complexity of the underlying function. Forster et al. [17] showed that proving lower bounds on the sign rank of a function gives lower bounds on the minimum size of any $\mathsf{THR} \circ \mathsf{MAJ}$ circuit computing it. Sign rank is known to be equivalent to dimension complexity, a geometric notion that is of fundamental importance in computational learning theory. Unfortunately, even proving lower bounds on the sign rank of a random function is non-trivial and was first done by Alon et al. [3]. On the other hand, proving strong lower bounds on the sign rank of an explicit function, IP, was a breakthrough achieved by Forster [16] fifteen years later. Since that work, there have relatively been just a few results proving strong sign rank lower bounds on explicit functions [48], [43], [10], [8]. While many basic questions about sign rank remain unanswered, new connections between it and other areas of mathematics keep showing up (see, for example, [4]).

We consider the following easily describable function F_n: The input, of length $n = 2m\ell$, is split into two disjoint parts, $X \in \{-1, 1\}^{m\ell}$ and $Y \in \{-1, 1\}^{m\ell}$. X and Y are each further divided into ℓ disjoint blocks $X_1, \ldots, X_\ell, Y_1, \ldots, Y_\ell$, of length m each. The function F_n outputs -1 iff the largest index $i \in [\ell]$ for which $X_i = Y_i$ holds is an odd index. For the purpose of this paper, we set $m = \ell^{1/3} + \log \ell$. It is not hard to see that F_n can be easily described as a decision list of Equalities.

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

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

The building block of F_n is Equality, which is a very simple function under various models of computation. It turns out F_n can be computed by a *depth-2* linear sized threshold formula. This simplicity of F_n, mainly in its depth complexity, enables us to settle two open problems. The first is a twenty-five year old (open since the work of Goldmann, Håstad and Razborov [18]) and very basic problem of understanding the relative power of weights in depth-2 threshold circuits. This application of our result is outlined in Section I-A. The second problem, posed much more recently by Göös, Pitassi and Watson [22], is a communication complexity class separation, outlined in Section I-B. Interestingly, our resolution of these two problems also serves to highlight an emerging barrier, that we call the 'sign rank barrier', against proving new lower bounds against depth-two threshold circuits and communication protocols just above the first level of the polynomial hierarchy.

A. Application: bottom weights can matter

Linear threshold functions (LTF's) form one of the most central classes of Boolean functions that are studied. Every

2575-8454/18/$31.00 ©2018 IEEE
DOI 10.1109/FOCS.2018.00014

IEEE computer society

such function is the halfspace induced by a real weight vector $\mathbf{w} \in \mathbb{R}^{n+1}$ denoted by $\mathsf{THR}_{\mathbf{w}}$ in the following way[1]: For each $x \in \{-1,1\}^n$, $\mathsf{THR}_{\mathbf{w}}(x) = \operatorname{sgn}\left(w_0 + \sum_{i=1}^{n} w_i x_i\right)$.

It is well known [40] that for every threshold gate with n inputs, there exists a threshold representation for it that uses only integer weights of magnitude at most $2^{O(n \log n)}$. The power of an LTF depends on the magnitude of the weights allowed. For instance, the Boolean function $\mathrm{GT}(x, y)$ that determines if the n-bit integer x is at least as large as the n-bit integer y is an example of an LTF that has no representation as an LTF with sub-exponentially small weights. Indeed in various areas, several questions and problems have been solved when the LTF's arising in the study are restricted to have small weights, but extending them to unrestricted weights are either open or have been solved after spending much research efforts. Examples of such areas are learning theory [33], [49], pseudorandom generators [46] , analysis of Boolean functions [26] and Boolean circuit complexity [14]. Understanding the relative power of large weights vs. small weights in the context of small-depth circuits having LTF's as gates has attracted attention by several works [5], [18], [51], [24], [25], [42], [23], [29], [19].

In this section, we describe the applicability of our main theorem in answering a longstanding open question that completes the picture of the role weights play in depth-2 threshold circuits. The class of all Boolean functions that can be computed by circuits of depth d and polynomial size, comprising gates computing LTF's (of polynomially bounded weights), is denoted by LT_d ($\widehat{LT}_d$). The seminal work of Minsky and Papert [38] showed that a simple function, Parity, is not in LT_1. While it is not very hard to verify that Parity is in $\widehat{LT}_2$, an outstanding problem is to exhibit an explicit function that is not in LT_2. This problem is now a well-identified frontier for research in circuit complexity.

By contrast, the relatively early work of Hajnal et al. [23] established the fact that the Inner-Product modulo 2 function (denoted by IP), that is easily seen to be in $\widehat{LT}_3$, is not in $\widehat{LT}_2$. It turns out that there is a natural class sitting between $\widehat{LT}_2$ and LT_2, denoted by $\mathsf{THR} \circ \mathsf{MAJ}$, where the top THR gate has unrestricted weights, but the weights of the bottom MAJ gates are restricted to be only polynomially large. Goldmann et al. [18] proved several interesting results, which implied the following structure.

$$\widehat{LT}_2 \overset{[18]}{=} \mathsf{MAJ} \circ \mathsf{THR} \overset{[18]}{\subsetneq} \mathsf{THR} \circ \mathsf{MAJ} \subseteq LT_2 \overset{[18]}{\subseteq} \widehat{LT}_3.$$

In a breakthrough work, Forster [16] showed that IP has sign rank $2^{\Omega(n)}$ for the natural partition of input variables. This yielded an exponential separation between $\mathsf{THR} \circ \mathsf{MAJ}$ and $\widehat{LT}_3$. This meant that at least one of the two containments $\mathsf{THR} \circ \mathsf{MAJ} \subseteq LT_2$ and $LT_2 \subseteq \widehat{LT}_3$ is strict. One might believe that the first containment is actually an equality, motivated by the fact that Goldmann et al. showed that in a related setting, $\mathsf{MAJ} \circ \mathsf{MAJ} = \mathsf{MAJ} \circ \mathsf{THR}$. Alman and Williams [2] recently showed interesting upper bounds on the 'probabilistic sign-rank' for functions in LT_2. In contrast, Amano and Maruoka [5] and Hansen and Podolskii [24] state that *separating* $\mathsf{THR} \circ \mathsf{MAJ}$ from $\mathsf{THR} \circ \mathsf{THR} = LT_2$ would be an important step for shedding more light on the structure of depth-2 boolean circuits. However, as far as we know, there was no clear target function identified for the purpose of separating the two classes. No progress on this question was made until our work.

We show that indeed $\mathsf{THR} \circ \mathsf{MAJ} \subsetneq \mathsf{THR} \circ \mathsf{THR}$ and the function F_n achieves the desired separation. To see why it does, we first note that F_n can be conveniently expressed as a composed function in the following way: consider a simple adaptation of the well-known $\mathsf{ODD\text{-}MAX\text{-}BIT}$ function, which we denote by OMB_ℓ^0. The function OMB_ℓ^0 outputs -1 precisely if the rightmost bit that is set to 1 occurs at an odd index. It is simple to observe that it is a linear threshold function:

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

Let $f_m \circ g_n : \{-1,1\}^{m \times n} \to \{-1,1\}$ be the composed function on mn input bits, where each of the m input bits to the outer function f is obtained by applying the inner function g to a block of n bits. It is not hard to verify that $F_n = \mathsf{OMB}_\ell^0 \circ \mathsf{OR}_{\ell^{1/3} + \log \ell} \circ \mathsf{XOR}_2$.

We first observe that F_n can be computed by linear sized $\mathsf{THR} \circ \mathsf{THR}$ formulas. For each $x \in \{-1,1\}^n$, let $\mathsf{ETHR}_{\mathbf{w}}(x) = -1 \iff w_0 + w_1 x_1 + \cdots + w_n x_n = 0$. Thus, ETHR gates are also called exact threshold gates. By first observing that every function computed by a formula of the form $\mathsf{THR} \circ \mathsf{OR}$ can also be computed by a formula of the form $\mathsf{THR} \circ \mathsf{AND}$ with a linear blow-up in size, it follows that F_n can be computed by linear size formulas of the form $\mathsf{THR} \circ \mathsf{AND} \circ \mathsf{XOR}_2$. Note that each $\mathsf{AND} \circ \mathsf{XOR}_2$ is computable by an ETHR gate. Hence, F_n is in $\mathsf{THR} \circ \mathsf{ETHR}$, a class that Hansen and Podolskii [24] showed is identical to the class $\mathsf{THR} \circ \mathsf{THR}$. Thus, our main theorem (Theorem I.1) and the above observation yield the following circuit class separation.

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

The message of Theorem I.2 may be contrasted with previous knowledge as follows: While weights at the bottom do not matter if the top is light, they do matter if the top is heavy. Further, Theorem I.2 also explains for the first

[1]Throughout this paper, we consider the input and output domains to be $\{-1,1\}^n$ and $\{-1,1\}$, rather than $\{0,1\}^n$ and $\{0,1\}$ respectively. -1 is identified with 'True', and 1 with 'False'.

time why current lower bound methods fail to get traction with $\mathsf{THR} \circ \mathsf{THR}$. Interestingly, it also suggests some new paths along which progress can be made. This is discussed in Section VIII.

B. Application: communication complexity

Göös [20] pointed out that F_n can be used to demonstrate another complexity class separation, this time in communication complexity. Complexity classes in communication complexity were first introduced in the seminal work of Babai, Frankl and Simon [6] as an analogue to the standard Turing complexity classes. While *unconditionally* understanding the relative power of (non)determinism and randomness in the context of Turing machines seems well beyond current techniques, Babai et al. had hoped that making progress in the mini-world of communication protocols would be less difficult. Indeed, we understand a lot more in the latter world. For instance, the class P^{cc} is *strictly* contained in both BPP^{cc} and NP^{cc}, while BPP^{cc} and NP^{cc}[2] are provably different. A major goal, set by Babai et al., is to prove lower bounds against the polynomial hierarchy for which the simple function IP has long been identified as a target. Unfortunately, it even remains open to exhibit a function that is not in the second level of the hierarchy. Our result explains this lack of progress by showing that a total function, conceivably well below the second level, has large sign rank.

Functions whose communication matrix of dimension $2^n \times 2^n$ have sign rank upper bounded by a quasi-polynomial in n were shown in [41] to correspond exactly to the complexity class UPP. This is the strongest communication complexity class against which we know how to prove explicit lower bounds. Razborov and Sherstov [43] proved that PH (in fact, $\Pi_2\mathsf{P}$) contains functions with large sign rank, rendering the sign rank technique essentially useless to prove lower bounds against the second level. A natural question is to understand until where does the sign rank method suffice to prove lower bounds.

Indeed, there is a rich landscape of communication complexity classes below the second level as discussed in a recent, almost exhaustive survey by Göös, Pitassi and Watson [22]. To motivate our contributions, we informally define MA protocols. Merlin, an all powerful prover, has access to Alice and Bob's inputs. He sends a (purported) proof string to Alice and Bob, who then run a randomized protocol to verify the proof. The protocol accepts an input iff the verification goes through. We say the protocol computes a function F if for all inputs to Alice and Bob, the probability of outputting the correct answer is at least $2/3$. The cost of the protocol on an input is the sum of the length of Merlin's proof string and the number of bits communicated between Alice and Bob. A function is said to be in the complexity class MA if there is such a protocol computing it with polylogarithmic worst-case cost (in the size of the input). For example, the function $\mathsf{OR} \circ \mathsf{EQ}$ can be seen to be in MA as follows: Merlin sends Alice and Bob the index of an input to the OR gate (if it exists) where EQ outputs -1, and Alice and Bob run an efficient randomized protocol for EQ to verify this. The class MA is a natural generalization of NP, and has received a lot of attention, starting with the work of [32]. It is known that MA is strictly contained in UPP.

One could similarly define AM, but its power remains much less understood. Göös et al. [22] conjectured that the (potentially incomparable) classes $\mathsf{AM} \cap \mathsf{coAM}$ and $\mathsf{S}_2\mathsf{P}$ contain functions of large sign rank. In a very recent work, Bouland et al. [8] showed that there is a *partial function* in $\mathsf{AM} \cap \mathsf{coAM}$ which has large sign rank, (partially) resolving the first conjecture.[3] We provide a strong confirmation of the second conjecture.

In order to state our result, let us consider the complexity class P^{MA} that is contained in $\mathsf{S}_2\mathsf{P}$. A function is in P^{MA} if it can be computed by deterministic protocols of polylogarithmic cost, where Alice and Bob have oracle access to any function in MA. The function F_n under the natural input partition (recall that it can be expressed as a decision list of *Equalities*) can be efficiently solved by P^{MA} protocols by an appropriate binary search, and querying an $\mathsf{OR} \circ \mathsf{EQ}$ oracle at each step.

We thus prove the following as a consequence of our main theorem.

Theorem I.3. *The function F_n witnesses the following communication complexity class separation.*

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

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

On the other hand, it is known that $\mathsf{P}^{\mathsf{NP}} \subsetneq \mathsf{UPP}$ and $\mathsf{MA} \subsetneq \mathsf{PP} \subsetneq \mathsf{UPP}$. These facts combined with Theorem I.3 shows that P^{MA} is right on the frontier between what we understand and what we do not. Thus, proving lower bounds against P^{MA} protocols emerges as a natural program for advancing the set of currently known techniques, given our work. Future directions are further discussed in Section VIII.

C. Related Work

Long after Forster [16] showed that upper bounding the spectral norm of a $\{-1, 1\}$ valued matrix suffices to show sign rank lower bounds, Sherstov [48] introduced an innovative method that designed a passage to a suitable

[2] Henceforth, we often drop *cc* from the superscript for convenience since we deal exclusively with communication complexity classes.

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

approximation problem via LP duality. This basic framework was again used to prove sign rank lower bounds, using additional tools from approximation theory, by Razborov and Sherstov [43], and subsequently by Bun and Thaler [10] and Bouland et al. [8]. All of these works [48], [43], [10], [8] rely on the passage, invented by Sherstov [48] to an approximation theoretic problem involving *low-degree* polynomials. This passage is made possible by exploiting the elegant spectral properties of communication matrices of the target functions, following the basic pattern matrix method of Sherstov [47].

Unfortunately, it seems difficult to embed a pattern matrix in a function in $\mathsf{THR} \circ \mathsf{THR}$. Consequently, we come up with a different type of function, F_n, that is an *XOR function*. Proving lower bounds on communication complexity of XOR functions, in general, has received a lot of attention recently [39], [53], [52], [27], [31]. However, there seem to be just two previous works that prove a lower bound on the sign rank of an XOR function, due to Hatami and Qian [28] and subsequently but independently by Ada et al. [1]. This result characterizes the sign rank of functions of the form $f \circ \mathsf{XOR}$ when f is symmetric. In contrast, our target function F_n is not a symmetric XOR function. Moreover, both the works [28] and [1] obtain their result using neat reductions from pattern matrices of symmetric functions, which had been analyzed by Sherstov [48]. Such a reduction for a function in $\mathsf{THR} \circ \mathsf{THR}$ is unknown, and plausibly impossible. This forces us to use a first-principle based argument for bounding the sign rank of an XOR function. Such functions also have nice spectral properties that are however different from those of pattern matrices. More specifically, the approximation-theoretic problem that one is led to in this case involves polynomials with *unrestricted* degree but low Fourier weight. A similar flavored but simpler problem had been tackled in a recent work of the authors [12], which characterized the *discrepancy* of XOR functions. Roughly speaking, in that work, the primal program constrained a distribution μ such that f correlates poorly with all parities w.r.t μ. However, there was no smoothness constraint imposed on μ in [12], which is what we are constrained to have in this work. Analyzing this combination of high degree parity constraints and the smoothness constraints is the main new technical challenge that our work addresses.

Our function F_n is simpler than the earlier functions in other ways. It is just a decision list of 'Equalities' that is therefore, both in the boolean circuit class $\mathsf{THR} \circ \mathsf{THR}$ and the communication complexity class P^{MA}. It is precisely this property of F_n that allows us to simultaneously answer two open questions.

D. Our Techniques

We strive to prove a lower bound on the sign rank of a function $F \in \mathsf{THR} \circ \mathsf{THR}$. We are guided by a communication complexity theoretic interpretation of sign rank, due to Paturi and Simon [41]. They introduced a model of two-party randomized communication, called the unbounded error model. In this model, Alice and Bob are only required to give the right answer with probability strictly greater than 1/2 on every input. Paturi and Simon showed that the sign rank of the communication matrix of F essentially characterizes its unbounded error complexity.

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

Taking this a step further, one could hope that the bottom gates could be any function that is hard to compute with such tiny error ϵ. The simplest such canonical function is Equality (denoted by EQ). Therefore, a plausible target is $\mathsf{THR} \circ \mathsf{EQ}$. This still turns out to be in $\mathsf{THR} \circ \mathsf{THR}$ as $\mathsf{EQ} \in \mathsf{ETHR}$. Moreover, EQ has a nice composed structure. It is just $\mathsf{AND} \circ \mathsf{XOR}$, which lets us re-express our target as $F = \mathsf{THR} \circ \mathsf{AND} \circ \mathsf{XOR}$, for some top THR that is 'suitably' hard. At this point, we view F as an XOR function whose outer function, f, needs to have sufficiently good analytic properties for us to prove that $f \circ \mathsf{XOR}$ has high sign rank.

We are naturally drawn to the work of Razborov and Sherstov [43] for inspiration as they bound the sign rank of a three-level composed function as well. They showed that $\mathsf{AND} \circ \mathsf{OR} \circ \mathsf{AND}_2$ has high sign rank. They exploited the fact that this function embeds a *pattern matrix* inside it, which has nice convenient spectral properties as observed in [47]. These spectral properties dictate them to look for an *approximately smooth orthogonalizing* distribution w.r.t which the outer function $f = \mathsf{AND} \circ \mathsf{OR}$ has zero correlation with small degree parities. This naturally gives rise to a linear program, that seeks to maximize the *smoothness* of the distribution under the constraints of low-degree orthogonality. The main technical challenge that Razborov and Sherstov overcome is the analysis of the dual of this LP using and building appropriate approximation theoretic tools. We follow this general framework of analyzing the dual of a suitable LP. However, as we are forced to work with an XOR function, there are new challenges that crop up. This is understandable, for if we take the same outer function of $\mathsf{AND} \circ \mathsf{OR}$, then the resulting XOR function has small sign rank. Indeed, this remains true even if one were to harden

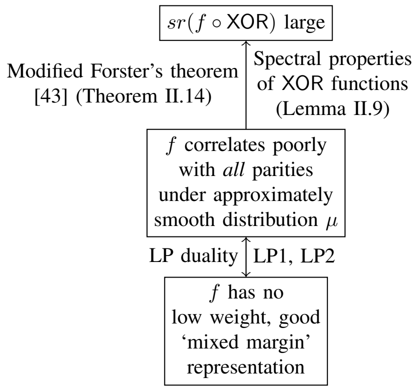

Figure 1. Approximation theoretic hardness of f implies large sign rank of $f \circ \mathsf{XOR}$ (Theorem III.1).

the outer function to $\mathsf{MAJ} \circ \mathsf{OR}$. This is simply because $\mathsf{OR} \circ \mathsf{XOR}$ is non-equality (NEQ), and a simple efficient UPP protocol for $\mathsf{MAJ} \circ \mathsf{NEQ}$ exists.

Figure 1 describes a general passage from the problem of lower bounding the sign rank of a function $f \circ \mathsf{XOR}$ to a sufficient problem of proving an approximation-theoretic hardness property of f, namely f has no good 'mixed margin' representation by low-weight polynomials. Theorem III.1 states the precise connection between the approximation-theoretic property of f and the sign rank of $f \circ \mathsf{XOR}$. This passage is made possible by using well-known spectral properties of XOR functions and LP duality. This is similar to earlier works [43], [48], [10], [8], where the spectral properties of pattern matrices were analyzed. The key difference between our work and theirs is in the nature of the approximation-theoretic problem that we end up with. While all these previous works had to rule out good *low-degree* representations, our Theorem III.1 stipulates us to rule out good *low-weight* representations of otherwise *unrestricted degree*.

Our main technical contribution is Theorem IV.1 which shows that the function $\mathsf{OMB}^0 \circ \mathsf{OR}$ is inapproximable by low-weight polynomials of *unrestricted degree*, in a sense which we elaborate on below. We prove this by a novel combination of ideas, sketched in Figure 2, that differs entirely from analysis in earlier works. One may view this result as a hardness amplification result, albeit specific to the function OMB^0. We start with the function OMB^0 which has no low weight 'worst-case margin' representation when the degree of the approximating polynomial is bounded [7]. We show that on composition with large fan-in OR gates, the function $\mathsf{OMB}^0 \circ \mathsf{OR}$ becomes '*mixed-margin*' inapproximable by low-weight polynomials, even with *unrestricted degree*. We believe this result to be of independent interest in the area of analysis of Boolean functions and approximation theory.

The first step in our method is to borrow an averaging idea from Krause and Pudlák [36] to show the following: A low-weight good approximation of $g \circ \mathsf{OR}_m$ by a polynomial p over the parity (Fourier) basis implies that there exists a low-weight polynomial q over the OR basis which approximates g as well as p approximates $g \circ \mathsf{OR}_m$, save an additive loss of at most 2^{-m}. This transformation to q is very useful because although it is still unrestricted in degree, it is over the OR basis, that is vulnerable to random restrictions. Indeed, in the next step, we hit q with random restrictions to reduce its degree. At this point, we extract a low weight *and* low-degree polynomial r that still approximates g_{rest}, the restriction of g. We now appeal to interesting properties of the ODD-MAX-BIT function by setting $g = \mathsf{OMB}^0$. First, we observe that OMB^0 on l bits, under random restrictions, retains its hardness as it contains OMB^0 on $l/8$ bits with high probability. Next, we show that OMB^0 does not have good low-degree approximations by appealing to classical approximation-theoretic tools, suitably modifying the arguments of Buhrman et al. [9] and Beigel [7]. This provides us with the required contradiction. Figure 2 provides an overview of the steps in our proof of Theorem IV.1.

II. Preliminaries

In this section, we provide some necessary preliminaries.

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

Definition II.2 (Threshold functions)**.** *A function $f : \{-1,1\}^n \to \{-1,1\}$ is called a* linear threshold function *if there exist integer weights $a_0, a_1, \ldots, a_n$ such that for all inputs $x \in \{-1,1\}^n$, $f(x) = \mathrm{sgn}(a_0 + \sum_{i=1}^n a_i x_i)$. Let* THR *denote the class of all such functions.*

Definition II.3 (Exact threshold functions)**.** *A function $f : \{-1,1\}^n \to \{-1,1\}$ is called an* exact threshold function *if there exist reals $w_1, \ldots, w_n, t$ such that $f(x) = -1 \iff \sum_{i=1}^n w_i x_i = t$. Let* ETHR *denote the class of exact threshold functions.*

Hansen and Podolskii [24] showed the following.

Theorem II.4 (Hansen and Podolskii [24])**.** *If a function $f : \{-1,1\}^n \to \{-1,1\}$ can be represented by a* $\mathsf{THR} \circ \mathsf{ETHR}$ *formula of size s, then it can be represented by a* $\mathsf{THR} \circ \mathsf{THR}$ *formula of size $2s$.*

One may refer to the full version of this paper [13] for a proof. In fact, Hansen and Podolskii [24] showed that the circuit class $\mathsf{THR} \circ \mathsf{THR}$ is identical to the circuit class $\mathsf{THR} \circ \mathsf{ETHR}$. However, we do not require the full generality of their result.

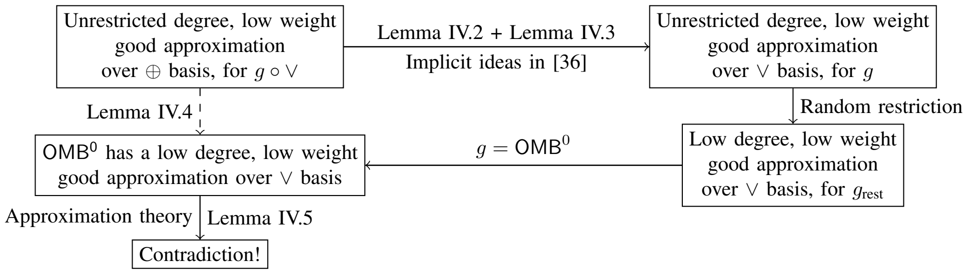

Figure 2. Approximation theoretic analysis (Theorem IV.1)

The next lemma states that any function computable by a $\mathsf{THR} \circ \mathsf{OR}$ formula can be computed by a $\mathsf{THR} \circ \mathsf{AND}$ formula without a blowup in the size.

Lemma II.5. *Suppose $f : \{-1,1\}^n \to \{-1,1\}$ can be computed by a $\mathsf{THR} \circ \mathsf{OR}$ formula of size s. Then, f can be computed by a $\mathsf{THR} \circ \mathsf{AND}$ formula of size s.*

Refer to the full version of this paper [13] for a proof of Lemma II.5, and for basics of Fourier analysis.

Lemma II.6 (Folklore)**.** *For any function $f : \{-1,1\}^n \to \mathbb{R}$,*

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

Fact II.7 (Plancherel's identity)**.** *For any functions $f, g : \{-1,1\}^n \to \mathbb{R}$,*

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

Definition II.8 (Signed monomial complexity)**.** *The* signed monomial complexity *of a boolean 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 f.

The following lemma characterizes the spectral norm of the communication matrix of XOR functions.

Lemma II.9 (Folklore)**.** *Let $f : \{-1,1\}^n \to \mathbb{R}$ be any real valued function and let M denote the communication matrix of $f \circ \mathsf{XOR}$. Then,*

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

where $||M||$ denotes the operator (spectral) norm of M.

We require the following well-known lemma by Minsky and Papert [38].

Lemma II.10 (Minsky and Papert [38])**.** *Let $p : \{-1,1\}^n \to \mathbb{R}$ be any symmetric real polynomial of degree d. Then, there exists a univariate polynomial q of degree at most d, such that for all $x \in \{-1,1\}^n$,*

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

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

We require the following approximation-theoretic lemma by Ehlich and Zeller [15] and Rivlin and Cheney [45].

Lemma II.11 ([15], [45])**.** *The following holds true for any real valued $\alpha > 0$ and integer $k > 0$. Let $p : \mathbb{R} \to \mathbb{R}$ be a univariate polynomial of degree $d < \sqrt{k/4}$, such that $p(0) \geq \alpha$, and $p(i) \leq 0$ for all $i \in [k]$. Then, there exists $i \in [k]$ such that $p(i) < -2\alpha$.*

Definition II.12 (OR polynomials)**.** *Define a function $p : \{-1,1\}^n \to \mathbb{R}$ of the form $p(x) = \sum_{S \subseteq [n]} a_S \bigvee_{i \in S} x_i$ to be an* OR polynomial*. Define the weight of p (in the OR basis) to be $\sum_{S \subseteq [n]} |a_S|$, and its degree to be $\max_{S \subseteq [n]}\{|S| : a_S \neq 0\}$.*

Remark II.13. *In the above definition, 'OR monomials' are defined as follows.*

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

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

Define the *sign rank* of a real valued matrix $A = [A_{ij}]$, denoted by $sr(A)$ to be the least rank of a real matrix $B = [B_{ij}]$ such that $A_{ij}B_{ij} > 0$ for all (i,j) such that $A_{ij} \neq 0$.

We require the following generalization of Forster's theorem [16] by Razborov and Sherstov [43].

Theorem II.14 (Razborov and Sherstov [43])**.** *Let $A = [A_{xy}]_{x \in X, y \in Y}$ be a real valued matrix with $s = |X||Y|$ entries, such that $A \neq 0$. For arbitrary parameters $h, \gamma > 0$, if all but h of the entries of A satisfy $|A_{xy}| \geq \gamma$, then*

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

Forster et al. [17] showed that functions with efficient $\mathsf{THR} \circ \mathsf{MAJ}$ representations have small sign rank.

Lemma II.15 (Forster et al. [17]). *Let $F : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ be a boolean function computed by a* $\mathsf{THR} \circ \mathsf{MAJ}$ *circuit of size s. Then,*

$$sr(M_F) \leq sn,$$

where M_F denotes the communication matrix of F.

For the purpose of this paper, we abuse notation, and use $sr(F)$ and $sr(M_F)$ interchangeably, to denote the sign rank of M_F.

In the model of communication we consider, 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 are given access to *private* randomness and wish to compute a given function $F : \mathcal{X} \times \mathcal{Y} \to \{-1,1\}$. We will use $\mathcal{X} = \mathcal{Y} = \{-1,1\}^n$ for the purposes of this paper. Alice and Bob communicate using a randomized protocol which has been agreed upon in advance. The cost of the protocol is the maximum number of bits communicated in the worst-case input and coin toss outcomes. A protocol Π computes F with advantage ϵ if the probability of F agreeing with Π is at least $1/2 + \epsilon$ for all inputs. We denote the cost of the best such protocol to be $R_\epsilon(F)$. Note here that we deviate from standard notation (used in [37], for example). Unbounded-error communication complexity was introduced by Paturi and Simon [41], and is defined as $\mathsf{UPP}(F) = \inf_{\epsilon>0}(R_\epsilon(F))$. This measure gives rise to the natural communication complexity class UPP^{cc}, defined as $\mathsf{UPP}^{cc}(F) \equiv \{F : \mathsf{UPP}(F) = \mathrm{polylog}(n)\}$.

Paturi and Simon [41] showed an equivalence between $\mathsf{UPP}(F)$ and the sign rank of M_F.

Theorem II.16 (Paturi and Simon [41]). *For any function $F : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$,*

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

III. Sign rank to polynomial approximation

In this section, we prove how a certain approximation-theoretic hardness property of f implies that the sign rank of $f \circ \mathsf{XOR}$ is large, as outlined in Figure 1.

Let $f : \{-1,1\}^n \to \{-1,1\}$ be any function, $\delta > 0$ be a parameter, and X be any subset of $\{-1,1\}^n$. We consider the following linear program, which has exactly the same structure as in (LP1) in [48] except for one crucial difference described below:

(LP1)

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

The first constraint in (LP1) specifies that correlation of f against *all parities* need to be small w.r.t a distribution μ. Note that in [48], this constraint was only imposed for *low degree* parities. This difference between the two linear programs forces us to entirely change the analysis of the dual from the one in [48]. As discussed earlier in Section I-D, this analysis is one of our main technical innovations. The second last constraint enforces the fact that μ is 'δ-smooth' over the set X. As we had indicated earlier in Section I-D, these constraints make analyzing the LP challenging.

Standard manipulations (as in [12], for example) and strong linear programming duality reveal that the optimum of (LP1) equals the optimum of (LP2). Let OPT denote the optima of these programs.

(LP2)

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

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

We now show that upper bounding OPT for any function f yields sign rank lower bounds against $f \circ \mathsf{XOR}$. The proof idea is depicted in Figure 1. The reader may refer to the appendix for boolean Fourier analysis basics.

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

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

Proof:

By (LP1), there exists a distribution μ on $\{-1,1\}^n$ such that $\mu(x) \geq \frac{\delta}{2^n}$ for all $x \in X$, and $\max_{S \subseteq [n]} |\widehat{f\mu}(S)| \leq \frac{OPT}{2^n}$. By Lemma II.9,

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

Each $x \in X$ contributes to 2^n entries of $M_{f\mu \circ \mathsf{XOR}}$ whose absolute value is at least δ. Plugging values in Theorem II.14, we obtain

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

which proves the desired sign rank lower bound. ∎

IV. Hardness of approximating $\mathsf{OMB}^0_\ell \circ \mathsf{OR}_m$

Below is our main technical result, capturing the essence of Figure 2, which says that no dual polynomial exists with a large optimum value for (LP2) when $f = \mathsf{OMB}^0_\ell \circ \bigvee_{\ell^{1/3}+\log \ell} : \{-1,1\}^{\ell^{4/3}+\ell \log \ell} \to \{-1,1\}$, even when the smoothness parameter δ is as high as 1/4.

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

Theorem IV.1 can be viewed as a hardness amplification theorem as follows. Our base function is OMB^0, which is known to be hard to approximate in the worst case by low-degree sign-representing polynomials [7], [9]. We show that a lifted version of this function, $\mathsf{OMB}^0_\ell \circ \mathsf{OR}_m$, cannot be approximated well under a significantly weaker notion of approximation where we permit any approximating polynomial to have the following additional power.

- Unrestricted degree but low weight.
- It need not sign represent $\mathsf{OMB}^0_\ell \circ \mathsf{OR}_m$, but a certain linear combination of their worst-case and average-case margin is small (see (LP2)).

We prove Theorem IV.1 towards the end of this section. In the remaining part of this section, we outline the various tools that go into proving Theorem IV.1, following the schematic in Figure 2. The proofs of all lemmas and claims in this section can be found in the full version of this paper [13].

We first use an idea from Krause and Pudlák [36] which enables us to work with polynomial approximations for g, given a polynomial approximation for $g \circ \bigvee_m$. We use the following notation for the following two lemmas. For any set $I \subseteq [\ell] \times [m]$, define $J \subseteq [\ell]$ to be the projection of I on $[\ell]$; $i \in J \iff \exists j, x_{i,j} \in I$. For any $y \in \{-1,1\}^\ell$, let μ_y denote the uniform distribution over all inputs $x \in \{-1,1\}^{m\ell}$ such that $\bigvee_m(x) = y$. Lemma IV.2 and Lemma IV.3 represent the first implication in Figure 2. The first tool we use is an approximation of monomials (in the parity basis) by OR functions, with a small error.

Lemma IV.2. *Let ℓ, m be positive integers such that $m > \log \ell$. For any set $I \subseteq [\ell] \times [m], y \in \{-1,1\}^\ell$,*

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

The proof of Lemma IV.2 appears in the proof of Lemma 2.3 in [36].

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

Lemma IV.3. *Let ℓ, m be positive integers such that $m > \log \ell$, and $g : \{-1,1\}^\ell \to \{-1,1\}$ be any function. Define $f = g \circ \bigvee_m : \{-1,1\}^{m\ell} \to \{-1,1\}, \Delta \in \mathbb{R}, e_x \geq 0\ \forall x \in X$, where X denotes the set of all inputs x in $\{-1,1\}^{m\ell}$ such that $\bigvee_m(x) = -1^\ell$, and let p be a real polynomial such that*

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

Then, there exists an OR *polynomial q, of weight at most $wt(p)$, such that for all $y \in \{-1,1\}^\ell$, $q(y)g(y) \geq \Delta - wt(p)\,(2\ell \cdot 2^{-m})$ and $q(-1^\ell)g(-1^\ell) \geq \Delta + \frac{\sum_{x \in X} e_x}{|X|} - wt(p)\,(2\ell \cdot 2^{-m})$.*

Next, we use random restrictions which reduces the degree of the approximating OR polynomial, at the cost of a small error. In particular, we consider the case when $g = \mathsf{OMB}^0_l$. This represents the dashed implication in Figure 2.

Lemma IV.4. *Let ℓ, m be any positive integers such that $m > \log \ell$. Let $g_\ell = \mathsf{OMB}^0_\ell : \{-1,1\}^\ell \to \{-1,1\}, f = g_\ell \circ \bigvee_m$, and $\Delta, \{e_x \geq 0 : x \in X\}$ (where X is defined as in Lemma IV.3), and p be a real polynomial such that*

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

Then, for any integer $d > 0$, there exists an OR *polynomial $r : \{-1,1\}^{\ell/8} \to \mathbb{R}$, of degree d and weight at most $wt(p)$, such that for all $y \in \{-1,1\}^{\ell/8}$, $r(y)g_{\ell/8}(y) \geq \Delta - wt(p)\left(2\ell \cdot 2^{-m} + 2^{-(d-1)}\right)$ and $r(-1^{\ell/8}) \geq \Delta + \frac{\sum_{x \in X} e_x}{|X|} - wt(p)\left(2\ell \cdot 2^{-m} + 2^{-(d-1)}\right)$.*

A. Hardness of OMB^0

The following lemma states that approximating OMB^0 well by a low-weight polynomial p is not possible unless the degree of p is large. This captures the last implication in Figure 2.

Lemma IV.5. *Suppose $p : \{-1,1\}^n \to \mathbb{R}$ is a polynomial of degree $d < \sqrt{n/4}$ and $a > 0, b \in \mathbb{R}$ be reals such that $p(-1^n) \geq a$ and $\mathsf{OMB}^0_n(x)p(x) \geq b$ for all $x \in \{-1,1\}^n$. Define $p_{max} = \max_{i \in \{0,\ldots,\lfloor n/10d^2 \rfloor\}} \{2^i a + \left(3 \cdot 2^i - 3\right) b\}$. Then, there exists $x \in \{-1,1\}^n$ such that $|p(x)| \geq p_{max}$.*

A simple consequence of the above lemma is that the weight of a polynomial p (in either the OR basis, or the parity basis) satisfying the assumptions of Lemma IV.5 is at least p_{max}. This property of p suffices for our need.

The proof of Lemma IV.5 follows an iterative argument, making repeated use of Lemma II.11, inspired by the arguments of Beigel [7] and Buhrman et al. [9].

Remark IV.6. *We remark here that this strengthens the result of Beigel [7], who proved that any good approximation by a low-degree* sign-representing *polynomial for* OMB^0 *must have large weight. Our approximating polynomial is not constrained to be sign representing (b might be negative in Lemma IV.5). In fact, it might disagree in sign on all points but* -1^n.

We now prove our main technical result, following the schematic depicted in Figure 2.

Proof of Theorem IV.1:

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

Lemma IV.4 (set $m = \ell^{1/3} + \log \ell$) shows the existence of an OR polynomial r, on $\ell/8$ variables, of degree $\ell^{1/3}$ and weight 1, such that for all $y \in \{-1,1\}^{\ell/8}$, $r(y)\mathsf{OMB}^0_{\ell/8}(y) \geq \Delta_{\mathsf{OPT}} - 2\cdot 2^{-\ell^{1/3}} - 2\cdot 2^{-\ell^{1/3}}$ and $r(-1^{\ell/8}) \geq \Delta + \frac{\sum_{x\in X}\xi_{x,\mathsf{OPT}}}{|X|} - 2\cdot 2^{-\ell^{1/3}} - 2\cdot 2^{-\ell^{1/3}}$. Observe that

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

r satisfies the assumptions of Lemma IV.5 with $d = \deg(r) = \ell^{1/3} < \sqrt{\ell/32}$ (since any OR polynomial of degree d can be represented by a polynomial of degree at most d), $a = \Delta_{\mathsf{OPT}} + \frac{\sum_{x\in X}\xi_{x,\mathsf{OPT}}}{|X|} - 4\cdot 2^{-\ell^{1/3}}$, and $b = \Delta_{\mathsf{OPT}} - 4\cdot 2^{-\ell^{1/3}}$. Note that a is non-negative because $a = \Delta_{\mathsf{OPT}} + \frac{\sum_{x\in X}\xi_{x,\mathsf{OPT}}}{|X|} - 4\cdot 2^{-\ell^{1/3}} \geq 2^{-\frac{\ell^{1/3}}{81}} - 4\cdot 2^{-\ell^{1/3}} \geq 0$.

Set $k = \ell^{1/3}/80$ for the remaining of this proof. By Lemma IV.5, there exists an $x \in \{-1,1\}^{\ell/8}$ such that

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

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

V. Proof of Main Theorem

We are now ready to prove our sign rank lower bound.

Theorem V.1 (Restatement of Theorem I.1). *Let* $f = \mathsf{OMB}^0_\ell \circ \bigvee_{\ell^{1/3}+\log\ell} : \{-1,1\}^{\ell^{4/3}+\ell\log\ell} \to \{-1,1\}$. *Then, for sufficiently large values of* ℓ, $sr(f\circ\mathsf{XOR}) \geq 2^{\frac{\ell^{1/3}}{81}-3}$.

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

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

■

VI. Applications

In this section, we list a few applications of Theorem I.1.

A. A separation of depth-2 threshold circuit classes

We are now ready to prove Theorem I.2, which gives us a lower bound on the size of $\mathsf{THR}\circ\mathsf{MAJ}$ circuits computing $F_n = \mathsf{OMB}^0_\ell \circ \bigvee_{\ell^{1/3}+\log\ell} \circ \mathsf{XOR}_2$, and resolving an open question posed in [5], [24] by yielding an exponential separation between the circuit classes $\mathsf{THR}\circ\mathsf{MAJ}$ and $\mathsf{THR}\circ\mathsf{THR}$.

Proof of Theorem I.2.: First, we show that F_n is computable by linear-sized $\mathsf{THR}\circ\mathsf{THR}$ formulas. Let $n = 2\ell^{4/3} + 2\ell\log\ell$ denote the number of input bits to $F_n = \mathsf{OMB}^0_\ell \circ \bigvee_{\ell^{1/3}+\log\ell} \circ \mathsf{XOR}_2$. By Lemma II.5, F_n can be computed by a $\mathsf{THR}\circ\mathsf{AND}\circ\mathsf{XOR}_2$ formula of size $2\ell^{4/3} + 2\ell\log\ell$. Hence $F_n \in \mathsf{THR}\circ\mathsf{ETHR} = \mathsf{THR}\circ\mathsf{THR}$, by Theorem II.4.

Next, we show a lower bound on the size of any $\mathsf{THR}\circ\mathsf{MAJ}$ circuit computing F_n. Suppose $\mathsf{OMB}^0_l \circ \bigvee_{l^{1/3}+\log l} \circ \mathsf{XOR}_2$ could be represented by a $\mathsf{THR}\circ\mathsf{MAJ}$ circuit of size s. By Lemma II.15 and Theorem V.1,

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

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

■

B. Communication complexity class separations

In this section, we show explicit separations between certain communication complexity classes, resolving an open question posed in [22]. This application of our main result was brought to our attention by Göös [20]. Formal definitions of communication complexity classes of interest may be found in the full version of this paper [13].

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

Proof: It follows from Theorem V.1 and Theorem II.16. ■

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

It is not hard to see that there is an MA protocol for $\bigvee_\ell \circ \mathsf{EQ}_{\ell^{1/3}+\log \ell}$ of cost polylogarithmic in ℓ. Using this, and a binary search, one can obtain a P^{MA} upper bound for F_n (see the full version of this paper [13] for a formal description of the protocol).

Along with Theorem VI.1, this yields the following result.

Theorem VI.2.

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

It is known that $\mathsf{P}^{\mathsf{MA}} \subseteq \mathsf{S2P}$, and $\mathsf{P}^{\mathsf{MA}} \subseteq \mathsf{BPP}^{\mathsf{NP}}$ (see, e.g., [22] for references for such containments, and an excellent overview on the landscape of communication complexity classes).

Thus, Theorem VI.2 yields

$$\mathsf{S2P} \not\subseteq \mathsf{UPP} \quad \text{and} \quad \mathsf{BPP}^{\mathsf{NP}} \not\subseteq \mathsf{UPP}.$$

The first non-inclusion resolves an open question posed in [22]. To the best of our knowledge, ours is the first explicit total function to witness the second non-inclusion. We remark here that Bouland et al. [8] used a partial function to witness the same separation.

VII. An upper bound

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

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

Proof: As noted in the previous section, F_n is expressible as a circuit of the form $\mathsf{THR}_\ell \circ \mathsf{EQ}_{\ell^{1/3}+\log \ell}$. A natural unbounded error protocol for F_n is to sample an input to the top threshold with probability proportional to its weight, and solve the corresponding Equality deterministically. The cost associated with sampling an input to the threshold is $\log \ell$, and the cost of solving an Equality deterministically is $\ell^{1/3} + \log \ell$, which is at most $2\ell^{1/3}$ for large enough values of ℓ. Since $n = \ell^{4/3} + \ell \log \ell > \ell^{4/3}$, the cost of the unbounded error protocol is $O(n^{1/4})$. By Theorem II.16, F_n has sign rank $2^{O(n^{1/4})}$. ■

VIII. Conclusion

We exhibit the first function known to be computable efficiently (in fact in linear size) by depth-2 Threshold circuits, but which has exponentially large sign rank. This result solves two open problems in one go: the first is a basic and old open problem, arising from the classical work of Goldmann, Håstad and Razborov [18] from the early nineties, of determining the power of weights in depth-2 Threshold circuits. Can such circuits be efficiently simulated by depth-2 circuits in which the bottom gates are restricted to have small weights? Goldmann et al. showed that they can be if we allow only small weights to appear at the top gate in the circuit we want to simulate. We prove that in general, such efficient simulations are impossible. This, along with previous work, yields the following fine structure of depth-2 Threshold circuit classes.

$$\widehat{LT}_2 \overset{[18]}{=} \mathsf{MAJ} \circ \mathsf{THR} \overset{[18]}{\subsetneq} \underbrace{\mathsf{THR} \circ \mathsf{MAJ} \subsetneq LT_2}_{\text{This work}} \overset{[18]}{\subseteq} \widehat{LT}_3.$$

Our work provides the first formal explanation of why current techniques have failed so far to prove lower bounds against $\mathsf{THR} \circ \mathsf{THR}$ circuits. It also suggests following directions along which progress can be made on this longstanding problem:

Our function is just a short decision list of Equalities. While it is not hard to show that decision lists of Equalities cannot compute IP,[4] can we prove strong lower bounds on the size of *decision lists of exact thresholds* for computing an explicit function in NP? This class of circuits is a sub-class of $\mathsf{THR} \circ \mathsf{THR}$ that is arguably natural. Our main result shows that this sub-class already inherits the curse of large sign rank. This raises the challenge of proving lower bounds on their size as a natural next step.

On a second front, our main result shows that the communication complexity class P^{MA} has functions with large sign rank, strongly resolving an open problem posed recently by Göös et al. [22]. This is in contrast to the known facts that every function in P^{NP} and MA have small sign rank. As the sign rank lower bound technique remains the strongest known technique for proving lower bounds against communication protocols (including quantum protocols), it suggests that new techniques need to be developed for proving bounds against P^{MA}. Indeed, there are specialized techniques for proving lower bounds against the class P^{NP} (see [30], [21]). Can they be generalized to P^{MA}? In particular, note that every function expressible as a short decision list of exact thresholds is in P^{MA}. Proving lower bounds on the length of such decision lists for computing an explicit function is also

[4] Since they are in AC^0 for instance.

a natural first step for proving lower bounds against P^{MA} communication protocols.

Our work puts the spotlight on the basic and simple computational model of 'decision lists of exact thresholds' that is capable of very efficiently computing a function of large sign rank. Proving lower bounds on the size of such decision lists is a necessary step for proving lower bounds against both $\mathsf{THR} \circ \mathsf{THR}$ circuit size and P^{MA} communication cost.

Acknowledgments

A.C. was partially supported by a Ramanujan fellowship of the DST, and N.S.M. was partially supported by a TCS fellowship. We are grateful to Kristoffer Hansen for bringing to our attention the question of separating the classes $\mathsf{THR}{\circ}\mathsf{MAJ}$ and $\mathsf{THR}{\circ}\mathsf{THR}$ at the Summer School on Lower Bounds, held in Prague in 2015. We also thank Michal Koucký for organizing and inviting us to the workshop. We thank Mika Göös for pointing out to us the consequences of our results in communication complexity class separations (Section VI-B), and Justin Thaler, Ryan Williams and Jaikumar Radhakrishnan for invaluable feedback on an earlier version of this manuscript.

References

[1] Anil Ada, Omar Fawzi, and Raghav Kulkarni. On the spectral properties of symmetric functions. *Arxiv*, 2017.

[2] Josh Alman and R. Ryan Williams. Probabilistic rank and matrix rigidity. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017,*, pages 641–652, 2017.

[3] Noga Alon, Peter Frankl, and Vojtech Rödl. Geometrical realization of set systems and probabilistic communication complexity. In *26th Annual Symposium on Foundations of Computer Science, 1985*, pages 277–280, 1985.

[4] Noga Alon, Shay Moran, and Amir Yehudayoff. Sign rank versus VC dimension. In *Proceedings of the 29th Conference on Learning Theory, COLT 2016*, pages 47–80, 2016.

[5] Kazuyuki Amano and Akira Maruoka. On the complexity of depth-2 circuits with threshold gates. In *30th International Symposium on Mathematical Foundations of Computer Science, 2005, Proceedings*, pages 107–118, 2005.

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

[7] Richard Beigel. Perceptrons, PP, and the polynomial hierarchy. *Computational Complexity*, 4:339–349, 1994.

[8] Adam Bouland, Lijie Chen, Dhiraj Holden, Justin Thaler, and Prashant Nalini Vasudevan. On the power of statistical zero knowledge. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS , 2017*, pages 708–719, 2017.

[9] 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.

[10] Mark Bun and Justin Thaler. Improved bounds on the sign-rank of AC^0. In *43rd International Colloquium on Automata, Languages, and Programming, ICALP, 2016*, pages 37:1–37:14, 2016.

[11] Mark Bun and Justin Thaler. A nearly optimal lower bound on the approximate degree of AC^0. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS, 2017*, pages 1–12, 2017.

[12] Arkadev Chattopadhyay and Nikhil S. Mande. Dual polynomials and communication complexity of XOR functions. *Arxiv*, 2017.

[13] Arkadev Chattopadhyay and Nikhil S. Mande. Weights at the bottom matter when the top is heavy. *Electronic Colloquium on Computational Complexity (ECCC)*, 24:83, 2017.

[14] Xi Chen, Igor Carboni Oliveira, and Rocco A. Servedio. Addition is exponentially harder than counting for shallow monotone circuits. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC, 2017*, pages 1232–1245, 2017.

[15] Hartmut Ehlich and Karl Zeller. Schwankung von polynomen zwischen gitterpunkten. *Mathematische Zeitschrift*, 86(1):41–44, 1964.

[16] Jürgen Forster. A linear lower bound on the unbounded error probabilistic communication complexity. *J. Comput. Syst. Sci.*, 65(4):612–625, 2002.

[17] 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 *FSTTCS 2001: Foundations of Software Technology and Theoretical Computer Science, 21st Conference, 2001, Proceedings*, pages 171–182, 2001.

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

[19] Mikael Goldmann and Marek Karpinski. Simulating threshold circuits by majority circuits. *SIAM J. Comput.*, 27(1):230–246, 1998.

[20] Mika Göös. Private Communication, 2017.

[21] Mika Göös, Pritish Kamath, Toniann Pitassi, and Thomas Watson. Query-to-communication lifting for P^NP. In *32nd Computational Complexity Conference, CCC, 2017*, pages 12:1–12:16, 2017.

[22] Mika Göös, Toniann Pitassi, and Thomas Watson. The landscape of communication complexity classes. *Computational Complexity*, 27(2):245–304, 2018.

[23] A. Hajnal, W. Maas, P. Pudlák, M. Szegedy, and G. Turán. Threshold circuits of bounded depth. *J. Comput. Syst. Sci.*, 46(2):129–154, 1993.

[24] Kristoffer Arnsfelt Hansen and Vladimir V. Podolskii. Exact threshold circuits. In *Proceedings of the 25th Annual IEEE Conference on Computational Complexity, CCC, 2010*, pages 270–279, 2010.

[25] Kristoffer Arnsfelt Hansen and Vladimir V. Podolskii. Polynomial threshold functions and boolean threshold circuits. *Inf. Comput.*, 240:56–73, 2015.

[26] Prahladh Harsha, Adam R. Klivans, and Raghu Meka. An invariance principle for polytopes. *J. ACM*, 59(6):29:1–29:25, 2012.

[27] Hamed Hatami, Kaave Hosseini, and Shachar Lovett. Structure of protocols for XOR functions. *SIAM J. Comput.*, 47(1):208–217, 2018.

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

[29] Thomas Hofmeister. A note on the simulation of exponential threshold weights. In *Computing and Combinatorics, Second Annual International Conference, COCOON, 1996*, pages 136–141, 1996.

[30] Russell Impagliazzo and Ryan Williams. Communication complexity with synchronized clocks. In *Proceedings of the 25th Annual IEEE Conference on Computational Complexity, CCC, 2010*, pages 259–269, 2010.

[31] Sampath Kannan, Elchanan Mossel, Swagato Sanyal, and Grigory Yaroslavtsev. Linear sketching over F_2. In *33rd Computational Complexity Conference, CCC, 2018*, pages 8:1–8:37, 2018.

[32] Hartmut Klauck. Rectangle size bounds and threshold covers in communication complexity. In *18th Annual IEEE Conference on Computational Complexity (Complexity 2003)*, pages 118–134, 2003.

[33] Adam R. Klivans, Ryan O'Donnell, and Rocco A. Servedio. Learning intersections and thresholds of halfspaces. *J. Comput. Syst. Sci.*, 68(4):808–840, 2004.

[34] Adam R. Klivans and Rocco A. Servedio. Toward attribute efficient learning of decision lists and parities. *Journal of Machine Learning Research*, 7:587–602, 2006.

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

[36] Matthias Krause and Pavel Pudlák. Computing boolean functions by polynomials and threshold circuits. *Computational Complexity*, 7(4):346–370, 1998.

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

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

[39] Ashley Montanaro and Tobias Osborne. On the communication complexity of XOR functions. *CoRR*, abs/0909.3392, 2009.

[40] S. Muroga. *Threshold Logic and its Applications*. Wiley-Interscience, 1971.

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

[42] Alexander A. Razborov. On small depth threshold circuits. In *Third Scandinavian Workshop on Algorithm Theory (SWAT)*, pages 42–52, 1992.

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

[44] Ronald L. Rivest. Learning decision lists. *Machine Learning*, 2(3):229–246, 1987.

[45] Theodore J Rivlin and Elliott W Cheney. A comparison of uniform approximations on an interval and a finite subset thereof. *SIAM Journal on numerical Analysis*, 3(2):311–320, 1966.

[46] Rocco A. Servedio and Li-Yang Tan. Fooling intersections of low-weight halfspaces. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS, 2017*, pages 824–835, 2017.

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

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

[49] Alexander A. Sherstov. Optimal bounds for sign-representing the intersection of two halfspaces by polynomials. *Combinatorica*, 33(1):73–96, 2013.

[50] Alexander A. Sherstov. The power of asymmetry in constant-depth circuits. In *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS, 2015*, pages 431–450, 2015.

[51] K. I. Siu and J. Bruck. On the power of threshold circuits with small weights. *SIAM J. Discrete Math.*, 4(3):423–435, 1991.

[52] Shengyu Zhang. Efficient quantum protocols for XOR functions. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA, 2014*, pages 1878–1885, 2014.

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