Electronic Colloquium on Computational Complexity, Report No. 176 (2018)

The Log-Approximate-Rank Conjecture is False

Arkadev Chattopadhyay*[1], Nikhil S. Mande†[1], and Suhail Sherif‡[1]

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

Abstract

We construct a simple and total XOR function F on $2n$ variables that has only $O(\sqrt{n})$ spectral norm, $O(n^2)$ approximate rank and $n^{O(\log n)}$ approximate nonnegative rank. We show it has polynomially large randomized bounded-error communication complexity of $\Omega(\sqrt{n})$. This yields the first exponential gap between the logarithm of the approximate rank and randomized communication complexity for total functions. Thus F witnesses a refutation of the Log-Approximate-Rank Conjecture (LARC) which was posed by Lee and Shraibman [LS09] as a very natural analogue for randomized communication of the still unresolved Log-Rank Conjecture for deterministic communication. The best known previous gap for any total function between the two measures is a recent 4th-power separation by Göös, Jayram, Pitassi and Watson [GJPW17].

Additionally, our function F refutes Grolmusz's Conjecture [Gro97] and a variant of the Log-Approximate-Nonnegative-Rank Conjecture, suggested recently by Kol, Moran, Shpilka and Yehudayoff [KMSY14], both of which are implied by the LARC. The complement of F has exponentially large approximate nonnegative rank. This answers a question of Lee [Lee12] and Kol et al. [KMSY14], showing that approximate nonnegative rank can be exponentially larger than approximate rank. The function F also falsifies a conjecture about parity measures of Boolean functions made by Tsang, Wong, Xie and Zhang [TWXZ13]. The latter conjecture implied the Log-Rank Conjecture for XOR functions. Our result further implies that at least one of the following statements is true: (a) The Quantum-Log-Rank Conjecture is false; (b) The *total function* F exponentially separates quantum communication complexity from its classical randomized counterpart.

*Partially supported by a Ramanujan fellowship of the DST. arkadev.c@tifr.res.in. Part of the work was done while participating in the program "Lower bounds in computational complexity" at the Simons Institute for the Theory of Computing.

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

‡Supported by a DAE fellowship. suhail.sherif@tifr.res.in

ISSN 1433-8092

1 Introduction

One of the most classical open problems in communication complexity is the Log-Rank Conjecture (LRC), introduced by Lovász and Saks [LS88]. Informally, it states that the deterministic communication complexity of a function F, a fundamental interactive complexity measure, is equivalent up to polynomial factors to a basic algebraic measure, namely the logarithm of the rank of the communication matrix of F, denoted by M_F. Despite receiving intense research focus, and some interesting recent progress [GL14, Lov16], the problem still remains wide open with surprising connections to other fields (see the recent survey by Lovett [Lov14]). The best upper bound on communication complexity has been improved, after decades, to square-root of the rank by Lovett [Lov16]. The best lower bound was improved, also recently, to $\Omega(\log^2 \mathsf{rank}(M_F))$ by Göös, Pitassi and Watson [GPW15].

If one expects linear algebraic measures to characterize deterministic communication complexity, it is natural to expect matrix analytic measures to characterize randomized communication complexity. It is well known that a deterministic communication protocol of cost c for F decomposes M_F into at most 2^c rank-one matrices (in fact combinatorial rectangles). Similarly, an ϵ-error randomized protocol for F decomposes the acceptance probability matrix Π of the protocol into at most 2^c rank-one matrices. The fact that the acceptance probability matrix point-wise approximates M_F to within ϵ, shows that the ϵ-approximate rank of M_F, denoted by $\mathsf{rank}_\epsilon(M_F)$ (see Definition 2.5), is at most 2^c. Motivated by this observation, Lee and Shraibman [LS09] made the following conjecture about ten years ago analogous to the original LRC.

Conjecture 1.1 (Log-Approximate-Rank Conjecture [LS09])**.** *There exists a universal constant α, such that the randomized $(1/3)$-error communication complexity of every total Boolean function F is $O\left(\log^\alpha \mathsf{rank}_{1/3}(M_F)\right)$.*

There are multiple reasons to be interested in the Log-Approximate-Rank Conjecture (LARC). In particular, the LARC implies a variety of other intriguing conjectures (see Figure 1), each of which has received significant individual attention. First, Gavinsky and Lovett [GL14] showed that the deterministic communication complexity of any function F is at most a multiplicative factor of $\log^2(\mathsf{rank}(M_F))$ away from the randomized complexity of F. Thus, the LARC implies the LRC and is a tempting generalization of the latter. Indeed in their survey, Lee and Shraibman [LS09] observed that all lower bounds that had been obtained on the randomized communication complexity of any F were within a quadratic factor of the logarithm of the approximate rank. The quadratic separation is witnessed by the classical linear lower bound [KS92, Raz92] on the randomized complexity of the Set-Disjointness function and the breakthrough tight bound of $\Theta(\sqrt{n})$ on the quantum complexity [BCW98, Raz03] of the same function. Very recently, the lower bound on the parameter α in the LARC was improved to 4 by Göös et al. [GJPW17] by exhibiting another function. Lovett in his survey [Lov14] recommends making progress towards the LARC as a natural future research direction. A related notion to rank is the nonnegative rank of a matrix M with nonnegative entries, denoted by $\mathsf{rank}^+(M)$. It is the smallest number of rank-1 matrices, each with nonnegative entries, needed such that their sum equals M. Yannakakis [Yan91] proved that the nonnegative rank of the slack matrix is essentially equivalent to its extension complexity. This has enabled a lot of recent exciting progress in combinatorial optimization (see, for example, [FMP$^+$15, KMR17]). More relevant to this work, we note that the deterministic communication complexity of a function F is long known [Lov90] to be bounded from above by a polynomial in the logarithm of the nonnegative rank. Motivated by this and other reasons, Kol et al. [KMSY14] proposed a weakening of the LARC, wherein they replace $\mathsf{rank}_{1/3}(M_F)$ with $\max\left\{\mathsf{rank}^+_{1/3}(M_F), \mathsf{rank}^+_{1/3}(M_{\overline{F}})\right\}$ (see Definition 2.5). This they call the Log-Approximate-Nonnegative-Rank Conjecture, which we abbreviate as the LANRC. They also consider the following strengthening of the LANRC.

Conjecture 1.2 (Strong Form of the Log-Approximate-Nonnegative-Rank Conjecture [KMSY14])**.** *There is some universal constant α^+, such that the randomized $(1/3)$-error communication complexity of every total Boolean function F is $O\left(\log^{\alpha^+} \mathsf{rank}^+_{1/3}(M_F)\right)$.*

In a much earlier work, Grolmusz [Gro97] made a seemingly different conjecture that he called the randomized analogue of the Log-Rank conjecture. For any real-valued function $f : \{0,1\}^n \to \mathbb{R}$, let $\left|\left|\widehat{f}\right|\right|_1$ denote the 1-norm of the Fourier transform of f, also known in the literature as the spectral norm of f. Interestingly, several recent papers, both in theoretical computer science and additive combinatorics, have been studying the structure of functions of low spectral norm (see, for example, [GS08, AFH12, TWXZ13, STV17, San18]).

Conjecture 1.3 (Grolmusz [Gro97])**.** *There exists a constant $\beta > 0$, such that the $(1/3)$-error randomized communication complexity of every total Boolean function F is at most $O\left(\log^{\beta}\left|\left|\widehat{F}\right|\right|_1\right)$.*

As Figure 1 shows, Grolmusz's conjecture also follows[1] from the LARC, by observing that functions of small spectral norm can be well approximated point-wise by a sparse real-valued function (see Appendix A for a full proof). Further, it can be seen using a result of Gavinsky and Lovett [GL14] that Grolmusz's conjecture itself implies the Log-Rank conjecture for XOR functions, i.e. functions of the form $f \circ \mathsf{XOR}$, where f is an arbitrary Boolean function. This is an important class of functions for which the LRC remains open despite recent efforts [TWXZ13, TXZ16, STV17].

In this work, we construct a simple and total Boolean XOR function on $O(n)$ bits, that has spectral norm $O(\sqrt{n})$, but whose randomized communication complexity is $\Theta\left(\sqrt{n}\right)$. This immediately yields a strong refutation not only of the LARC, but also of both the strong form of the LANRC and Grolmusz's Conjecture at the same time. In particular, our function has approximate nonnegative rank bounded by $n^{O\log(n)}$. Further, our work has interesting consequences for parity measures of Boolean functions and rules out an approach to prove the LRC for XOR functions.

The *parity kill number* of f is defined as

$$\mathsf{C}_{\oplus,\min}(f) := \min\{\text{co-dim}(S) | S \text{ is an affine subspace on which } f \text{ is constant}\}.$$

Tsang et al. [TWXZ13] conjectured that the parity kill number of any non-constant f is bounded above by a polynomial in the logarithm of $\left|\left|\widehat{f}\right|\right|_1$.

Conjecture 1.4 (Parity Kill Number Conjecture, Conjecture 25 in [TWXZ13])**.** *There is an absolute constant θ such that for any non-constant $f : \{-1,1\}^n \to \{-1,1\}$, $\mathsf{C}_{\oplus,\min}(f) = O\left(\log^{\theta}(\left|\left|\widehat{f}\right|\right|_1)\right)$.*

Conjecture 1.4 implies, via results of Tsang et al. [TWXZ13], the Log-Rank Conjecture for XOR functions. Establishing the LRC for this class is particularly appealing as it boils down to neat Fourier analytic questions of independent interest. Tsang et al. could not find a counter-example to the above for even $\theta = 1$. Later, O'Donnell et al. [OWZ+14] constructed a function for which θ needs to be at least 1.58. In contrast, the Boolean function SINK, which is our main construction, strongly refutes Conjecture 1.4. We show that $\mathsf{C}_{\oplus,\min}(\mathsf{SINK})$ is exponentially larger than $\log\left(\left|\left|\widehat{\mathsf{SINK}}\right|\right|_1\right)$.

1.1 Our Results and Intuition

The difficulty in coming up with counter-examples for either the LRC or the LARC is that most functions in use in communication complexity are block-composed functions of the form $f \circ g$, where g is a convenient obfuscating gadget. A recent success story in the field is that of lifting theorems (see, for example, [GPW15, CKLM17, GPW17]) which lift decision-tree complexity of f to the communication complexity of $f \circ g$, suitably multiplied by a measure of g that very closely approximates the communication complexity of g. We believe the main reasons such composed functions so far have not generated any counter-example for the LRC/LARC, is that the analogue of the log-rank type of conjectures in the decision-tree world is long known to be true. In particular, both the exact and approximate (real) degree of a function f is polynomially related to its decision tree complexity.

This is the reason we turn to XOR functions, i.e. functions of the form $f \circ \mathsf{XOR}$. Here, it is expected that instead of the ordinary decision tree complexity of f, its *parity decision tree* (PDT) complexity[2]

[1] up to a polylogarithmic factor in the input size

[2] Refer to Section 2 for formal definitions of measures used in the remaining part of this section.

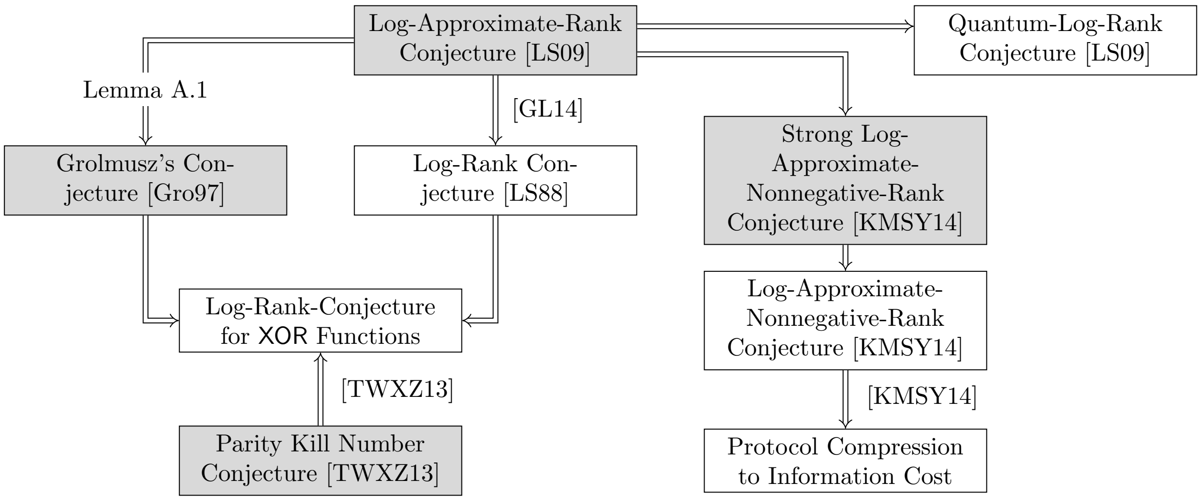

Figure 1: Implications between various interesting conjectures. Shaded conjectures are disproved in this work, and the rest remain unresolved.

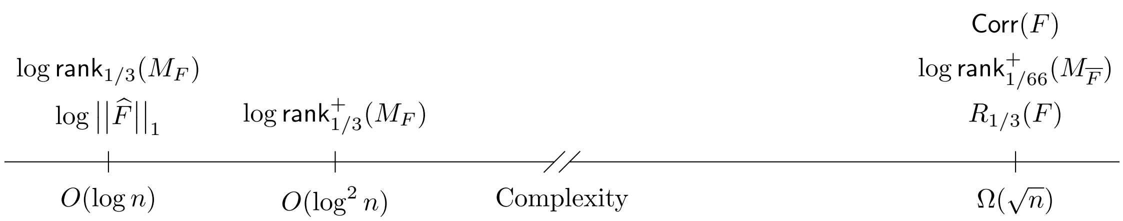

Figure 2: Various complexity measures of our function $F := \mathsf{SINK} \circ \mathsf{XOR}$ on $2n$ bits. Prior to this work, no exponential separation between any pair of these measures was known for a total function.

should lift to the communication complexity of $f \circ \mathsf{XOR}$. Much less is understood about parity decision tree complexity. It is natural to expect that the role of degree of a function f, should be played by its Fourier sparsity, i.e. the size of the support set of the Fourier transform of f, denoted by $\left|\left|\widehat{f}\right|\right|_0$. The analogue of the Log-Rank conjecture for PDT's states that the PDT complexity of f should be at most $\log^{O(1)}(\left|\left|\widehat{f}\right|\right|_0)$. This is tantalizingly open, despite recent efforts [TWXZ13, STV17, TXZ16].

A natural analogue of the LARC for PDT's would state that the randomized PDT complexity of f is bounded from above by a polynomial in the logarithm of its approximate Fourier sparsity. The ϵ-approximate sparsity of f, denoted by $\left|\left|\widehat{f}\right|\right|_{0,\epsilon}$, is the smallest number s such that there exists an s-sparse real function g that point-wise (on the Boolean cube) ϵ-approximates f. Our main construction comes up with a function that refutes this analogue of LARC for PDT's.

We define a function $\mathsf{SINK} : \{0,1\}^n \to \{0,1\}$ where the input of length $n := \binom{m}{2}$ specifies the orientation of the edges of the complete graph on m vertices. The function outputs 1 if there is a vertex that is a sink in the given orientation of edges, and 0 otherwise. A formal definition is given below.

Definition 1.5 (SINK)**.** *Consider a tournament (see Definition 2.12) on m vertices defined by the $\binom{m}{2}$ variables $x_{i,j}$ for $i < j \in [m]$ in the following way: $x_{i,j} = 1 \implies v_i \to v_j$ is the direction of the (v_i, v_j) edge, and $x_{i,j} = 0 \implies v_i \leftarrow v_j$ is the direction. The function SINK computes whether or not there is a sink in the graph. In other words,*

$$\mathsf{SINK}(x) = 1 \iff \exists\, i \in [m] \text{ such that all edges adjacent to } v_i \text{ are incoming.}$$

A simple but key fact about SINK is that the value of each edge variable rules out exactly one of its endpoints from being a sink. This ensures that at most one vertex can be a sink. This results in the following Fourier analytic measures being low for SINK.

Claim 1.6 (Part of Theorem 1.10)**.**

1. $||\widehat{\mathsf{SINK}}||_1 \leq m$.
2. $||\widehat{\mathsf{SINK}}||_{0,1/3} = O(m^4)$.

The same key fact makes it simple to observe that SINK can be computed by an ordinary deterministic decision tree that makes at most $2m$ queries. However, it turns out that even a *randomized parity* decision tree cannot do substantially better. The intuition is that a linear query, unless localized to the edges incident to a specific vertex of the graph, cannot get much information about the status of that vertex. Thus, one expects that after t linear queries, the PDT algorithm will only have information about $O(t)$ vertices, keeping the status of the remaining $m - O(t)$ vertices virtually undetermined. A careful analysis formalizing this intuition results in two strong lower bounds.

Theorem 1.7.

1. $\mathsf{C}_{\oplus,\min}(\mathsf{SINK}) = \lceil 2m/3 \rceil$.
2. *Any PDT computing* SINK *has size at least* $2^{2m/3}$.

The first part above shows that the parity kill number of SINK is exponentially larger than the logarithm of its spectral norm, strongly refuting a conjecture of Tsang et al. [TWXZ13] (Conjecture 1.4). The best known previous separation, due to O'Donnell et al. [OWZ+14], was sub-quadratic. The second part of Theorem 1.7 easily follows from the first, but has the following interesting consequence. Combining the work of Tsang et al. with their own results, Shpilka, Tal and Volk [STV17] showed that any Boolean function f on n bits has a deterministic parity decision tree of size $2^A n^{2A}$, where $A = ||\widehat{f}||_1$. This, in particular, showed that every function can be decomposed as a sum of at most $2^A n^{2A}$ many indicator functions of affine spaces, yielding an exponential improvement upon the earlier doubly exponential bound given by the structure theorem of Green and Sanders [GS08]. Shpilka et al. hoped that their upper bound on PDT size could be further significantly improved to $\mathrm{poly}(n, A)$. The second part of Theorem 1.7, on the other hand, shows that the bound of Shpilka et al. is indeed tight in its dependence on the spectral norm of f.

Our next result shows that SINK remains hard for even randomized parity decision trees. Let $\mathsf{RPDT}_\epsilon(g)$ denote the ϵ-error randomized parity decision tree complexity of the Boolean function g.

Theorem 1.8. $\mathsf{RPDT}_{1/3}(\mathsf{SINK}) = \Theta(m)$.

We prove this lower bound as a robust version of Part 1 of Theorem 1.7. We give a balanced distribution under which any nearly 0-monochromatic affine subspace must have large co-dimension. The existence of such a distribution is then shown to imply an RPDT lower bound.

Theorem 1.8 and Claim 1.6 together already refute the analogue of the LARC for randomized parity decision trees. This makes the XOR function $\mathsf{SINK} \circ \mathsf{XOR}$ a natural candidate for refuting the LARC for randomized communication complexity.

While proving a randomized lifting theorem for XOR functions remains an interesting open problem (a deterministic analogue of this was established recently by Hatami, Hosseini and Lovett [HHL18]), we are able to confirm that hardness lifting does take place for SINK.

More precisely, we are interested in the composed function $\mathsf{SINK} \circ \mathsf{XOR} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, where $n := \binom{m}{2}$.

Theorem 1.9. $R_{1/3}(\mathsf{SINK} \circ \mathsf{XOR}) = \Theta(m)$.

To lift the RPDT hardness to communication hardness, we use the natural lift of the distribution used to prove Theorem 1.8. We then follow ideas from Gavinsky [Gav16] and use it in conjunction with Shearer's lemma to conclude an $\Omega(m)$ lower bound on the corruption (see Definition 2.13) of $\mathsf{SINK} \circ \mathsf{XOR}$, which in turn gives us the randomized communication lower bound.

In contrast, through well-known connections, the Fourier simplicity of SINK recorded in Claim 1.6 also results in several analytic measures being low for $\mathsf{SINK} \circ \mathsf{XOR}$.

Theorem 1.10.

1. $\left|\left|\widehat{\mathsf{SINK} \circ \mathsf{XOR}}\right|\right|_1 = \left|\left|\widehat{\mathsf{SINK}}\right|\right|_1 \leq m$.

2. $\mathsf{rank}_{1/3}(M_{\mathsf{SINK} \circ \mathsf{XOR}}) \leq \left|\left|\widehat{\mathsf{SINK}}\right|\right|_{0,1/3} = O(m^4)$.

3. $\mathsf{rank}^+_{1/3}(M_{\mathsf{SINK} \circ \mathsf{XOR}}) = m^{O(\log m)}$.

Theorems 1.9 and 1.10 together immediately yield a refutation of Grolmusz's Conjecture (Conjecture 1.3), the strong form of the LANRC (Conjecture 1.2) and therefore the LARC (Conjecture 1.1).

Since corruption lower bounds approximate nonnegative rank, our proof also shows that the approximate nonnegative rank of the complement of $\mathsf{SINK} \circ \mathsf{XOR}$ is large, showing for the first time that the property of having small approximate nonnegative rank is not closed under complementation of Boolean functions.

Theorem 1.11.

$$\mathsf{rank}^+_{1/66}(M_{\overline{\mathsf{SINK} \circ \mathsf{XOR}}}) \geq 2^{\Omega(m)}.$$

Before our work, the largest known gap between log of the approximate rank and the log of the approximate nonnegative rank for any total Boolean function was quadratic, witnessed by Set-Disjointness, as shown by Kol et al. [KMSY14]. Theorem 1.11 yields an exponential improvement in the gap via $\mathsf{SINK} \circ \mathsf{XOR}$.

Implications for quantum communication complexity

Our refutation of the LARC entails an interesting consequence for quantum communication. The Quantum-Log-Rank Conjecture (QLRC) is a weakening of the LARC that states that the quantum communication complexity of a function is upper bounded by a polynomial in the logarithm of its approximate rank. If the QLRC were true, then $\mathsf{SINK} \circ \mathsf{XOR}$ would yield the first exponential separation between the quantum and randomized communication complexities of a total function, resolving a major open problem. Otherwise, the QLRC would be falsified by $\mathsf{SINK} \circ \mathsf{XOR}$. It is worth remarking that Lee and Shraibman [LS09] considered the QLRC to be the most plausible one in the Log-Rank family of conjectures. Previous works [ZS09, Zha14] established the QLRC for special classes of functions. Much more recent work of Anshu et al. [ABG+17] gave the first (quadratic) separation of quantum communication complexity from the logarithm of the approximate rank, via an involved function, using the recently developed cheat-sheet framework [ABK16]. On the other hand, $\mathsf{SINK} \circ \mathsf{XOR}$ could plausibly provide much stronger separations.

Remark 1.12. *Note that the communication lower bound for* $\mathsf{SINK} \circ \mathsf{XOR}$ *stated in Theorem 1.9 already implies the RPDT hardness of* SINK *stated in Theorem 1.8. We still choose to state them separately in order to convey the natural intuition that led us to our results. Further, the proof of Theorem 1.8 uses just linear algebraic notions, without requiring the use of Shearer's Lemma.*

1.2 Organization

In Section 2, we review the necessary preliminaries. In Section 3, we formally define the SINK function and look at some interesting properties of it. In Section 4, we show that SINK is simple under some measures. Next, we show the hardness of SINK under various parity measures in Section 5, refuting Conjecture 1.4. In Section 6, we show a randomized communication lower bound for $\mathsf{SINK} \circ \mathsf{XOR}$, yielding refutations of Conjecture 1.1, Conjecture 1.2 and Conjecture 1.3. In the same section, we prove Theorem 1.11. Finally we state some conclusions in Section 7.

2 Preliminaries and Definitions

Lemma 2.1 (Hoeffding's Inequality, [Hoe63])**.** *Let $X_i \in [a_i, b_i]$ for $i = 1, 2, \ldots, n$ be independent random variables and $X = \sum_i X_i$. Then*

$$\Pr[|X - \mathbb{E}[X]| \geq t] < 2\exp\left(-\frac{2t^2}{\sum_i (b_i - a_i)^2}\right).$$

Consider the vector space of functions $V = \{f : \{0,1\}^n \to \mathbb{R}\}$ equipped with the inner product defined by

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

The set of *parity* functions $\{\chi_S\}_{S \subseteq [n]}$, defined by $\chi_S(x) = (-1)^{\sum_{i \in S} x_i}$, forms an orthonormal basis for this vector space under the inner product defined above. Thus, every function $f : \{0,1\}^n \to \mathbb{R}$ has a unique representation $f = \sum_{S \subseteq [n]} \hat{f}(S)\chi_S$. The coefficients $\{\hat{f}(S)\}_{S \subseteq [n]}$ are called the *Fourier coefficients* of f.

Definition 2.2 (Spectral norm)**.** *Define the spectral norm of a function $f : \{0,1\}^n \to \mathbb{R}$ as follows.*

$$\left|\left|\widehat{f}\right|\right|_1 := \sum_{S \subseteq [n]} \left|\hat{f}(S)\right|.$$

Definition 2.3 (Sparsity)**.** *Define the sparsity of a function $f : \{0,1\}^n \to \mathbb{R}$ as follows.*

$$\left|\left|\widehat{f}\right|\right|_0 := \left|\{S \subseteq [n] | \hat{f}(S) \neq 0\}\right|.$$

For any function $F : \{0,1\}^n \times \{0,1\}^n \to \mathbb{R}$, define a $2^n \times 2^n$ matrix M_F as $M_F[(x,y)] = F(x,y)$. M_F is called the *communication matrix* of F.

We now define the rank and nonnegative rank of a matrix.

Definition 2.4 (Rank and Nonnegative rank)**.** *The rank (nonnegative rank, respectively) of a matrix M, denoted $\mathsf{rank}(M)$ ($\mathsf{rank}^+(M)$, respectively), is the minimum k for which there exist k rank 1 matrices (nonnegative-valued rank 1 matrices, respectively) such that $M = \sum_{i=1}^k M_i$.*

The "approximate" versions of spectral norm, sparsity, rank and nonnegative rank are defined as follows.

Definition 2.5. *For any function $f : \{0,1\}^n \to \mathbb{R}$ and matrix M, define the following measures.*

- *ϵ-approximate spectral norm:* $\left|\left|\widehat{f}\right|\right|_{1,\epsilon} := \min\{\left|\left|\widehat{g}\right|\right|_1 : \forall x,\ |g(x) - f(x)| \leq \epsilon\}$.
- *ϵ-approximate sparsity:* $\left|\left|\widehat{f}\right|\right|_{0,\epsilon} := \min\{\left|\left|\widehat{g}\right|\right|_0 : \forall x,\ |g(x) - f(x)| \leq \epsilon\}$.
- *ϵ-approximate rank:* $\mathsf{rank}_\epsilon(M) := \min\{\mathsf{rank}(M') : \forall x, y,\ |M'(x,y) - M(x,y)| \leq \epsilon\}$.
- *ϵ-approximate nonnegative rank:* $\mathsf{rank}^+_\epsilon(M) := \min\{\mathsf{rank}^+(M') : \forall x, y,\ |M'(x,y) - M(x,y)| \leq \epsilon\}$.

Definition 2.6 (XOR functions)**.** *A function $F : \{0,1\}^n \times \{0,1\}^n \to \mathbb{R}$ is called an XOR function if there exists a function $f : \{0,1\}^n \to \mathbb{R}$ 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 $x, y \in \{0,1\}^n$. Denote $F = f \circ \mathsf{XOR}$.*

Lemma 2.7 (Folklore)**.** *For any function $f : \{0,1\}^n \to \mathbb{R}$,*

$$\mathsf{rank}(M_{f \circ \mathsf{XOR}}) = \left|\left|\widehat{f}\right|\right|_0.$$

Zhang [Zha14] showed that if a function $f : \{0,1\}^n \to \mathbb{R}$ has small spectral norm (in fact small approximate spectral norm), then f has small approximate sparsity. While not explicitly stated as such, it is a straightforward corollary of Lemma 3.1 in [Zha14].[3]

Lemma 2.8. *For any $g : \{0,1\}^n \to \mathbb{R}$ and $\delta > \epsilon \geq 0$,*

$$||\widehat{g}||_{0,\delta} \leq O\left(||\widehat{g}||_{1,\epsilon}^2 n/(\delta-\epsilon)^2\right).$$

Definition 2.9 (Subcube)**.** *A set $T \subseteq \{0,1\}^n$ is said to be a subcube if there exist coordinates $i_1, \ldots i_k$ and integers $a_1, \ldots a_k \in \{0,1\}$ such that $T = \{x \in \{0,1\}^n | x_{i_1} = a_1, x_{i_2} = a_2, \ldots, x_{i_k} = a_k\}$. We call $\mathsf{fixed}(T) := \{i_1, \ldots, i_k\}$ the set of fixed coordinates in T.*

Definition 2.10 (Affine subspace)**.** *A set $T \subseteq \{0,1\}^n$ is said to be an affine subspace if there exist independent linear forms $L_1, \ldots, L_k$ and integers $a_1, \ldots, a_k \in \{0,1\}$ such that $T = \{x \in \{0,1\}^n | L_i(x) = a_i$ for all $i \in [k]\}$. k is called the co-dimension of T, denoted* co-dim(T).

Given a system of linear equations $\{L_1 = a_1, \ldots, L_r = a_r\}$, we define its span as $\{L' = a' : L' \in \text{span}\{L_1, \ldots, L_r\}\}$ (where a' is inferred from $a_1, \ldots, a_r$).

Claim 2.11. *Let W be an affine subspace of $\{0,1\}^n$ defined by a system of equations with span $\mathcal{L}$. Let $S \subseteq [d]$. Let $\mathcal{L}_S \subseteq \mathcal{L}$ be the subset of equations that are supported completely by variables indexed within S. For any $y \in \{0,1\}^S$, the number of extensions of y in W is 0 if y violates a constraint in $\mathcal{L}_S$ and $2^{\dim(W)-(|S|-\dim(\mathcal{L}_S))}$ otherwise.*

Proof. Let $T_S \subseteq \{0,1\}^S$ be the affine subspace where the equations $\mathcal{L}_S$ are satisfied.

- If $y \in \{0,1\}^S$ is not in T_S, then y must contradict one of the equations in $\mathcal{L}_S$. Since this is also an equation satisfied in W, no extension of y is in W.
- If $y \in \{0,1\}^S$ is in T_S, we now count the number of its completions that lie in W. Consider the space $S_y \subseteq \{0,1\}^n$ obtained by fixing the coordinates in S according to y. There are still $\dim(\mathcal{L}) - \dim(\mathcal{L}_S)$ linearly independent forms that W depends on. Hence the number of solutions that extend y to an element of W is $2^{(n-|S|)-(\dim(\mathcal{L})-\dim(\mathcal{L}_S))} = 2^{\dim(W)-(|S|-\dim(\mathcal{L}_S))}$.

□

Definition 2.12 (Tournament)**.** *A tournament on n vertices is a directed graph, which is obtained by assigning directions to each edge in the undirected complete graph, K_n.*

A *parity decision tree* (PDT) computing a function $f : \{0,1\}^n \to \{0,1\}$ is a binary tree with leaf nodes labelled in $\{0,1\}$, each internal node is labelled by a linear form χ_S and has two outgoing edges, labelled 1 and -1. On input x, the tree's computation proceeds by computing $\chi_S(x)$ as indicated by the node's label and following the edge indicated by the value of the computed linear form. The output value at the leaf must equal $f(x)$. The parity decision tree complexity of f, denoted $\mathsf{PDT}(f)$ is defined as follows.

$$\mathsf{PDT}(f) := \min_{\substack{T:T \text{ is a PDT} \\ \text{computing } f}} \text{depth}(T).$$

The PDT size complexity of f, denoted $\mathsf{size}_{\mathsf{PDT}}(f)$, is the minimum number of leaves in a PDT computing f.

A randomized parity decision tree (RPDT) is a parity decision tree that is equipped with an arbitrarily long string of random bits. At every vertex, the linear form queried is now a function of the random string. The output $T(x)$ of the tree T on an input x is now a random variable. To say that the tree computes a function f to within error ϵ, we require that for each x, $\Pr[T(x) = f(x)] \geq 1 - \epsilon$. The randomized parity decision tree complexity of f, denoted $\mathsf{RPDT}(f)$ is defined as follows.

$$\mathsf{RPDT}(f) := \min_{\substack{T:T \text{ is an RPDT} \\ \text{computing } f}} \text{depth}(T).$$

[3] Zhang [Zha14] attributes this lemma to Grolmusz [Gro97], who attributes it to Bruck and Smolensky [BS90]. For completeness and clarity, we reproduce a proof in Appendix A.

2.1 Communication Complexity

We are interested in the two-party classical *bounded-error* (BPP) model of communication introduced by Yao [Yao79]. In this communication model, two parties, Alice and Bob, receive $x \in X, y \in Y$, respectively. They wish to jointly compute a function $F : X \times Y \to \{0,1\}$ on the input (x, y) via a communication protocol Π. They have unbounded computational power individually and wish to minimize the number of bits communicated. Alice and Bob are given access to an arbitrarily long string of public random bits. At any point in the protocol, the party that sends the next message is decided by the previous messages communicated. The message communicated by the party is a function of the party's input, the previous messages communicated and the random string. We say that a protocol Π computes F to within error ϵ if $\Pr[\Pi(x, y) = F(x, y)] \geq 1 - \epsilon$ for each $x \in X, y \in Y$. The cost of the protocol is the largest number of bits communicated in the worst case over all possible inputs and outcomes of the random coin tosses. The ϵ-error randomized communication complexity of F, denoted $R_\epsilon(F)$, is the minimum cost of a protocol that computes F to within error ϵ.

One can also define $R^{pri}_\epsilon(F)$ in the same way, except that the protocols are restricted so that instead of having a string of public random bits, each party has their own private string of random bits.

One way to prove lower bounds against randomized communication complexity is via the corruption bound.

Definition 2.13 (Corruption). *For any function $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ and any bit $z \in \{0,1\}$, we define*

$$\mathsf{Corr}^z_\epsilon(F) := \max_\mu \min_R \log \frac{1}{\mu(R)},$$

where μ and R range over

- *Balanced distributions: Distributions μ such that $\mu(F^{-1}(z)) \in [1/3, 2/3]$,*
- *Biased rectangles: Rectangles R such that $\mu(R \cap F^{-1}(\bar{z})) < \epsilon\mu(R)$.*

$\mathsf{Corr}_\epsilon(F)$ is defined as the maximum of $\mathsf{Corr}^0_\epsilon(F)$ and $\mathsf{Corr}^1_\epsilon(F)$.

Klauck [Kla03] showed that Corr_ϵ and Corr_δ are equivalent up to a multiplicative constant for any constants $\epsilon, \delta \in (0, 1/12]$. By $\mathsf{Corr}(F)$, we refer to $\mathsf{Corr}_{1/12}(F)$. It is well known (see, for example, [KN97]) that this quantity is a lower bound against randomized communication.

Fact 2.14. *For any function $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$,*

$$R_{1/3}(F) \geq \Omega(\mathsf{Corr}(F)).$$

2.2 Entropy

Definition 2.15 (Entropy). *Let X be a discrete random variable. The entropy $H(X)$ is defined as*

$$H(X) := \sum_{s \in \mathrm{supp(X)}} \Pr[X = s] \log \frac{1}{\Pr[X = s]}.$$

Fact 2.16 (Folklore). $\mathrm{supp}(X) = k \implies H(X) \leq \log k$*, with equality if and only if X is uniform.*

Definition 2.17 (Relative Entropy). *Let ν, μ be distributions over a finite set S of outcomes. The relative entropy (or Kullback-Liebler divergence) $d_{KL}(\nu||\mu)$ is defined as*

$$d_{KL}(\nu||\mu) := \sum_{s \in S} \nu(s) \log \frac{\nu(s)}{\mu(s)}.$$

Lemma 2.18 (Pinsker's Inequality). *For two distributions ν, μ over the same set of outcomes,*

$$d_{KL}(\nu||\mu) \geq \frac{1}{2 \ln 2} ||\nu - \mu||_1^2.$$

The following claim appears in [Gav16].

Claim 2.19 (Two faraway distributions cannot both have near-maximum entropy, [Gav16])**.** *If ν_1 and ν_2 are two distributions in $\{0,1\}^n$, then $||\nu_1 - \nu_2||_1^2 \leq 8\ln 2 \cdot (n - \min\{H(\nu_1), H(\nu_2)\})$.*

We reproduce a proof for completeness.

Proof. Let u be the uniform distribution over $\{0,1\}^n$ and d be $||\nu_1 - \nu_2||_1$. Without loss of generality, assume $||\nu_1 - u||_1 \geq d/2$. From Pinsker's inequality (Lemma 2.18) we know that $d^2/4 \leq 2\ln 2\, d_{KL}(\nu_1||u)$. But

$$d_{KL}(\nu_1||u) = \sum_x \nu_1(x)\log\frac{\nu_1(x)}{2^{-n}} = \sum_x \nu_1(x)\left(\log\nu_1(x) + \log\frac{1}{2^{-n}}\right) = -H(\nu_1) + n.$$

We thus see that $d^2 \leq 8\ln 2(n - H(\nu_1))$, or that $H(\nu_1) \leq n - ||\nu_1 - \nu_2||^2/(8\ln 2)$. $\square$

Lemma 2.20 (Shearer's Lemma)**.** *Let $X = (X_1, \ldots, X_n)$ be a random variable. If S is a random variable distributed on subsets of the coordinates $[n]$, such that for every $i \in [n]$, $\Pr[i \in S] \geq t$, then $\mathbb{E}[H(X_S)] \geq tH(X)$ where X_S is the random variable $(X_i : i \in S)$.*

3 The Disjoint Subcube Function

Definition 3.1 ($\mathsf{DISJ}-\mathsf{SUBCUBE}_{\mathcal{S}}$)**.** *Consider a set $\mathcal{S} = \{S_1, S_2, \ldots, S_m\}$ of m disjoint subcubes in $\{0,1\}^k$. Define*

$$\mathsf{DISJ}-\mathsf{SUBCUBE}_{\mathcal{S}}(x) := \begin{cases} 1 & x \in \bigcup_{i\in[m]} S_i \\ 0 & \textit{otherwise.} \end{cases}$$

Claim 3.2. *Let $\mathcal{S} = \{S_1, \ldots, S_m\}$ be any set of m disjoint subcubes in $\{0,1\}^k$. Then,*

$$\left|\left|(\mathsf{DISJ}-\widehat{\mathsf{SUBCUBE}}_{\mathcal{S}})\right|\right|_1 \leq m.$$

Proof. Since the subcubes are disjoint, the exact polynomial representation for $\mathsf{DISJ}-\mathsf{SUBCUBE}_{\mathcal{S}}$ is

$$\mathsf{DISJ}-\mathsf{SUBCUBE}_{\mathcal{S}} = \sum_{S_i\in\mathcal{S}} p_{S_i},$$

where p_{S_i} is the polynomial that evaluates to 1 for all $x \in S_i$ and 0 otherwise. That is,

$$p_{S_i}(x) = \prod_{j\in\mathsf{fixed}(S_i)}\left(\frac{1+(-1)^{b_j}\chi_{\{j\}}(x)}{2}\right),$$

where S_i fixes x_j to b_j. Expanding the above gives a sum of $2^{|\mathsf{fixed}(S_i)|}$ monomials, each with a coefficient of absolute value $2^{-|\mathsf{fixed}(S_i)|}$. Thus the spectral norm of p_{S_i} is 1, and hence the spectral norm of $\mathsf{DISJ}-\mathsf{SUBCUBE}_{\mathcal{S}}$ is at most m. $\square$

3.1 The Sink Function

We first recall the definition of SINK.

Definition 1.5 (SINK)**.** *Consider a tournament (see Definition 2.12) on m vertices defined by the $\binom{m}{2}$ variables $x_{i,j}$ for $i < j \in [m]$ in the following way: $x_{i,j} = 1 \implies v_i \to v_j$ is the direction of the (v_i, v_j) edge, and $x_{i,j} = 0 \implies v_i \leftarrow v_j$ is the direction. The function SINK computes whether or not there is a sink in the graph. In other words,*

$$\mathsf{SINK}(x) = 1 \iff \exists\ i \in [m] \textit{ such that all edges adjacent to } v_i \textit{ are incoming.}$$

We now note that SINK is a specific instance of $\mathsf{DISJ} - \mathsf{SUBCUBE}$.

Lemma 3.3. *Consider the function* SINK *on* $\binom{m}{2}$ *variables. There exists a set of disjoint subcubes* $\mathcal{S} = \{S_1, \ldots, S_m\}$ *in* $\{0,1\}^{\binom{m}{2}}$ *such that* $\mathsf{SINK} = \mathsf{DISJ} - \mathsf{SUBCUBE}_{\mathcal{S}}$.

Proof. Label each of the $\binom{m}{2}$ coordinates by a unique (i,j) pair for $i < j \in [m]$. The description of $\mathsf{fixed}(S_i)$ is given below.

- $x_{i,j} = 0$ for all $j > i$.
- $x_{j,i} = 1$ for all $j < i$.

For each $i < j$, $S_i \cap S_j = \emptyset$ since $x_{i,j} = 0$ in S_i and $x_{i,j} = 1$ in S_j. Thus, the subcubes $S_1, \ldots, S_m$ are disjoint. Note that the function $\mathsf{DISJ} - \mathsf{SUBCUBE}_{\mathcal{S}}$ is exactly the same as SINK. $\square$

Unless mentioned otherwise, we consider the SINK function to be on $\binom{m}{2}$ variables.

3.1.1 Projections

While analyzing the complexity of SINK, we will often use projections of the inputs. Let $X \in \{-1,1\}^{\binom{m}{2}}$. To see how X orients the edges incident to a vertex v_i, let E_{v_i} be the set of $m-1$ input coordinates that correspond to the edges incident to v_i. We use the notation X_{v_i} to denote the input projected to the coordinates in E_{v_i}. Note that X_{v_i} decides whether or not v_i is a sink. By z_i, we refer to the $m-1$ bit string such that v_i is a sink if and only if $X_{v_i} = z_i$.

4 Simplicity of SINK

In this section, we prove Theorem 1.10, showing that SINK is simple in the sense that the spectral norm of SINK is small, the approximate rank of $\mathsf{SINK} \circ \mathsf{XOR}$ is small, and so is the approximate nonnegative rank of $\mathsf{SINK} \circ \mathsf{XOR}$.

Theorem 1.10.

1. $\left|\left|\widehat{\mathsf{SINK} \circ \mathsf{XOR}}\right|\right|_1 = \left|\left|\widehat{\mathsf{SINK}}\right|\right|_1 \leq m$.
2. $\mathsf{rank}_{1/3}(M_{\mathsf{SINK} \circ \mathsf{XOR}}) \leq \left|\left|\widehat{\mathsf{SINK}}\right|\right|_{0,1/3} = O(m^4)$.
3. $\mathsf{rank}^{+}_{1/3}(M_{\mathsf{SINK} \circ \mathsf{XOR}}) = m^{O(\log m)}$.

Proof. **Proof of Part 1:** It follows from Claim 3.2, Lemma 3.3 and the observation that composing with XOR does not change the spectral norm: if $p_{\mathsf{SINK}}(x)$ is the polynomial computing SINK, then $p_{\mathsf{SINK} \circ \mathsf{XOR}}(x,y)$ is obtained by replacing every monomial $\chi_S(x)$ in $p_{\mathsf{SINK}}(x)$ with $\chi_S(x)\chi_S(y) = \chi_S(x \oplus y)$.

Proof of Part 2: Recall that Part 1 of Theorem 1.10 implies that $\left|\left|\widehat{\mathsf{SINK} \circ \mathsf{XOR}}\right|\right|_1 \leq m$. Lemma 2.8 implies existence of a function $f : \{0,1\}^{\binom{m}{2}} \to \mathbb{R}$ with sparsity $O(m^4)$ such that $|\mathsf{SINK}(x) - f(x)| \leq 1/3$ for all x. Hence, by Lemma 2.7, $f \circ \mathsf{XOR}$ has rank $O(m^4)$ and $|\mathsf{SINK} \circ \mathsf{XOR}(x,y) - f \circ \mathsf{XOR}(x,y)| \leq 1/3$ for all x, y.

Proof of Part 3: We show that the communication matrix of $\mathsf{SINK} \circ \mathsf{XOR}$ is pointwise close to a matrix M such that M can be written as the nonnegative sum of at most $m^{O(\log m)}$ nonnegative rank 1 matrices.

Note that $\mathsf{SINK} \circ \mathsf{XOR}(X,Y)$ can be written as an OR of Equalities: $\bigvee_{i=1}^{m} (X_{v_i} = Y_{v_i} \oplus z_i)$. Since at most one of these Equalities can fire for any input, it is in fact the sum of the m Equalities. It is well known that any $2(m-1)$ size Equality can be solved to error at most $1/3m$ with a randomized communication protocol (with private randomness) of cost $O(\log^2 m)$. [Kra96] showed that $R^{pri}_{\epsilon}(F) \geq \log \mathsf{rank}^{+}_{\epsilon}(M_F)$. So the matrix for any of the Equalities has $(1/3m)$-approximate nonnegative rank at most $2^{O(\log^2 m)}$. By adding up these matrices, we get a matrix that pointwise approximates the matrix of $\mathsf{SINK} \circ \mathsf{XOR}$ as a sum of $m2^{O(\log^2 m)}$ nonnegative rank 1 matrices. $\square$

5 Parity Hardness of SINK

Recall that Conjecture 1.4 states that for any Boolean function f, the smallest co-dimension of an affine subspace on which f is a constant (i.e. $\mathsf{C}_{\oplus,\min}(f)$) is bounded from above by $O\left(\log^c\left(\left|\left|\widehat{f}\right|\right|_1\right)\right)$. Tsang et al. expressed that no counterexample to Conjecture 1.4 was known even for $c = 1$. O'Donnell et al. [OWZ+14] showed that $c \geq \log_2 3$ is necessary. We show that $\mathsf{C}_{\oplus,\min}(\mathsf{SINK}) = \Omega(\left|\left|\widehat{\mathsf{SINK}}\right|\right|_1)$, strongly refuting Conjecture 1.4.

We first show that any affine subspace in which every input has no sink must be small. We will in fact prove a stronger statement.

Lemma 5.1 (Sink Avoidance is Costly)**.** *Fix any $k \leq m$. Let W be an affine subspace such that the vertices $v_1, v_2, \ldots, v_k$ are not sinks in any input in W. Then co-dim$(W) \geq 2k/3$.*

Proof. Let W be an affine subspace of $\{0,1\}^{\binom{m}{2}}$, defined by a system of linear equations with span $\mathcal{L}$, such that for every input in W, none of $V = \{v_1, \ldots, v_k\}$ is a sink. This is equivalent to saying that for every $1 \leq i \leq k$, no extension of z_i appears in W. Then, by Claim 2.11, we know that for every $1 \leq i \leq k$, there exists a linear equation in $\mathcal{L}$ of the form $l_i = a_i$, where l_i is supported completely by the variables indexed within E_{v_i}, that is violated by z_i. Let us call such a linear constraint a *v_i-constraint.* It is not hard to see that no constraint can simultaneously be a v_i-constraint and a v_j-constraint where $i, j \in [k],\ i \neq j$. We form a set L of size k by picking a v_i-constraint for each $v_i \in V$. We argue that the dimension of span(L) is at least $2k/3$, to conclude that W has co-dimension at least $2k/3$.

Let $B \subseteq L$ be a basis of span(L), $|B| = b$. Suitably relabelling vertices, let L_B denote $L \setminus B = \{l_1, l_2, \ldots, l_{k-b}\}$ where l_i is a v_i-constraint. We make the two following simple claims which together easily imply our lemma.

Claim 5.2. *Let $l_r \in L_B$, such that $l_r = l'_1 + \cdots + l'_s$, where each $l'_i \in B$. Then, the set $B \setminus \{l'_1, \ldots, l'_s\}$ spans every element in $L_B \setminus \{l_r\}$.*

Claim 5.3. *Let $l_r \in L_B$ and $B_0 \subseteq B$, such that $l_r \in$ span(B_0). Then, $|B_0| \geq 2$.*

Before proving the above two claims, let us use them to establish our lemma. Pick any $l \in L_B$. Then, find the minimal $B_0 \subseteq B$ such that $l \in$ span(B_0). By Claim 5.3, $|B_0| \geq 2$. Now, shrink B and L_B by deleting B_0 and l from them respectively. Then, by Claim 5.2, the shrunk B still spans the new L_B. Hence, we can repeat the above step to shrink B this way at least $k - b$ times before it becomes empty. At each step B shrinks in size by at least 2. Thus, $b \geq 2(k - b)$, yielding $b \geq \lceil 2k/3 \rceil$.

All that is left is to establish the two claims. Let us begin by proving Claim 5.2. First, consider the vertices $v'_1, \ldots, v'_s, v$, where l'_i is a v_i constraint and l_r is a v_r constraint. It is simple to observe that $x_{i,j}$ is supported by l'_i iff it is supported by l'_j. Hence, consider the undirected graph G_r with vertex-set $\{v_1, \ldots, v_s\} \cup \{v_r\}$, where edge (v_i, v_j) is present iff $x_{i,j}$ is supported by l'_i. Further, it is also simple to observe, using the fact that B is a basis, that G_r is connected. This also means that in any non-trivial linear sum of linear forms in L that is identically zero, l'_i participates for any $1 \leq i \leq s$ iff each of $l'_1, \ldots, l'_s$ participate. Otherwise, there will be some $j \neq k$ and $1 \leq j, k \leq s$ such that $x_{j,k}$ will appear exactly once in the sum, contradicting the fact that the sum is zero.Now take any $l \in L_B \setminus l_r$. Because B is a basis, there is a linear sum of elements just from B that equals l. If any element from B_0 participates in this, then by the above argument l_r will also participate, yielding a contradiction as $l_r \notin B$.

Finally, we prove Claim 5.3. For the sake of contradiction, let $B_0 = \{l_j\}$, for some $j \leq s$. Then, $l_r = l_j = x_{j,r}$, leading to a contradiction.

□

We now observe that the lower bound on the co-dimension obtained in Lemma 5.1 is tight.

Claim 5.4. *Fix any $k \leq m$. There is an affine subspace W of co-dimension $\lceil 2k/3 \rceil$ such that the vertices $V = \{v_1, v_2, \ldots, v_k\}$ are not sinks in any input in W.*

Proof. Let k' be the largest multiple of 3 less than or equal to k. For vertex $v_{k'+1}$, should it exist, we set $x_{(1,k'+1)}$ so that the $(v_1, v_{k'+1})$ edge is directed out of $v_{k'+1}$. We do the same for $v_{k'+2}$. Now we only need to ensure that none of $v_1, \ldots, v_{k'}$ are sinks.

Group these k' vertices into triples $(v_1, v_2, v_3), \ldots, (v_{k'-2}, v_{k'-1}, v_{k'})$. Consider the space W obtained as the solution space to the constraints defined below. For each triple (v_i, v_{i+1}, v_{i+2}), add the following two constraints to the constraint list of W.

- $x_{(i,i+1)} + x_{(i,i+2)}$ is set so that exactly one of the two edges (v_i, v_{i+1}) and (v_i, v_{i+2}) is directed out of v_i.
- $x_{(i,i+1)} + x_{(i+1,i+2)}$ is set so that exactly one of the two edges (v_i, v_{i+1}) and (v_{i+1}, v_{i+2}) is directed out of v_{i+1}.

The two constraints above are simultaneously satisfied if and only if v_i, v_{i+1}, v_{i+2} forms a cycle. Hence the above constraints ensure that none of $v_1, \ldots, v_{k'}$ are sinks. The total number of constraints is $2\lfloor k/3 \rfloor + (k \mod 3) = \lceil 2k/3 \rceil$. ☐

Proof of Theorem 1.7. **Proof of Part 1:** Let W be an affine subspace of $\{0,1\}^{\binom{m}{2}}$ such that every input in W has a sink. Since the number of such inputs is at most $2^{\binom{m}{2}} \cdot m/2^{m-1}$, this means W must must have co-dimension at least $m - 1 - \log m$.

Now let W be an affine subspace of $\{0,1\}^{\binom{m}{2}}$ such that every input in W has no sink. By Lemma 5.1 (set $k = m$), W must have co-dimension at least $2m/3$.

By Lemma 5.4, we see that there in fact is a monochromatic affine subspace of co-dimension $\lceil 2m/3 \rceil$.

Proof of Part 2: Every leaf of the PDT is a monochromatic affine subspace of co-dimension at most the depth of the leaf. From Part 1, we know that every leaf in a PDT computing SINK has to be at depth at least $2m/3$. Hence the number of leaves in any PDT computing SINK is at least $2^{2m/3}$. ☐

5.1 RPDT Hardness

Throughout this section, we refer to the function SINK on $\binom{m}{2}$ inputs as f to avoid clutter. Our RPDT lower bound will mimic a corruption bound, but for affine subspaces.

Fact 5.5. *Given a randomized parity decision tree Π_1 of cost c and error ϵ for a function $g : \{0,1\}^k \to \{0,1\}$ and any distribution μ on $\{0,1\}^k$, there exists a deterministic parity decision tree Π_2 of cost at most c such that $\Pr_{x\sim\mu}[\Pi_2(x) \neq g(x)] \leq \epsilon$.*

We fix μ to be the distribution that is a half-and-half combination of "uniform over the universe" and "uniform over the 1-inputs". More precisely, let μ_0 be uniform over $\{0,1\}^{\binom{m}{2}}$ and μ_1 be uniform over $f^{-1}(1)$. Define $\mu := (\mu_0 + \mu_1)/2$.

Claim 5.6. *For μ as defined above, $\mu(f^{-1}(1)) = 1/2 + o(1)$.*

Proof. By the definition of μ, this is equivalent to proving that $\mu_0(f^{-1}(1)) = o(1)$. Recall that μ_0 is the uniform distribution on all inputs. In order for a vertex v to be a sink, all the edges adjacent to it must be incoming. Thus,

$$\Pr_{\mu_0}\left[v \text{ is a sink } = 2^{-(m-1)}\right].$$

A union bound yields the claim. ☐

Lemma 5.7. *Let $\epsilon \leq 1/8$ be any constant. Any deterministic parity decision tree Π of cost c for f with error probability ϵ under the input distribution μ induces an affine subspace W such that $\mu(W \cap f^{-1}(1)) \leq 4\epsilon\mu(W)$ and* co-dim$(W) \leq c$.

Proof. A deterministic parity decision tree gives a partition of the universe into at most 2^c labelled affine subspaces, each of co-dimension at most c, where the label is 1 iff the affine subspace has a larger mass over its 1-inputs than over its 0-inputs. We note the following about the affine subspaces in the partition.

- By Claim 5.6, only $1/2 + o(1)$ mass of the distribution is on 1-inputs. Since the error probability of Π is ϵ, it cannot have 1-affine subspaces covering more than $1/2 + o(1) + \epsilon$ mass under the distribution μ.
- Affine subspaces W which have error $\geq 4\epsilon\mu(W)$ can only make up $1/4$ mass, to keep the total error below ϵ.

Hence, if $\epsilon \leq 1/8$ is a constant, Π must induce a 0-affine subspace W such that $\mu(W \cap f^{-1}(1)) \leq 4\epsilon\mu(W)$ and $\text{co-dim}(W) \leq c$. $\square$

Lemma 5.8 (Smallness of Biased Affine Subspaces)**.** *Let $\epsilon < 1/16$. Any affine subspace W that has $\mu(W \cap f^{-1}(1)) \leq 4\epsilon\mu(W)$ satisfies*

$$\text{co-dim}(W) \geq m/3.$$

The proof of the Lemma 5.8 is given in Section 5.1.1. We first prove Theorem 1.8 assuming the lemma. Recall that Theorem 1.8 states that f is hard for randomized parity decision trees.

Theorem 1.8. $\mathsf{RPDT}_{1/3}(\mathsf{SINK}) = \Theta(m)$.

Proof. Set $\epsilon = 1/32$. By Fact 5.5, a randomized parity decision tree of cost c for f implies existence of a deterministic parity decision tree of cost at most c, which errs on at most ϵ mass of the inputs under the distribution μ. Lemma 5.7 then implies the existence of an affine subspace W satisfying $\mu(W \cap f^{-1}(1)) \leq 4\epsilon\mu(W)$ and $\text{co-dim}(W) \leq c$. By Lemma 5.8, any such W has $\text{co-dim}(W) \geq m/3$. Thus, $\mathsf{RPDT}_{1/32}(f) \geq m/3$. A standard error reduction argument shows that any randomized parity decision tree for f with $1/3$ error would also require $\Omega(m)$ cost. $\square$

5.1.1 Smallness of Biased Affine Subspaces

Lemma 5.8 states that any 0-biased affine subspace under μ must have large co-dimension. We prove this in two steps.

- We show that any 0-biased affine subspace must have a very small fraction of 1 inputs.
- We then show that any affine subspace with a very small fraction of 1 inputs must have large co-dimension.

We formally state these two steps below and then prove them.

Claim 5.9. *Let $\epsilon \leq 1/8$. An affine subspace W such that $\mu(W \cap f^{-1}(1)) \leq 4\epsilon\mu(W)$ must satisfy*

$$\frac{|W \cap f^{-1}(1)|}{|W|} \leq 8\epsilon \frac{|f^{-1}(1)|}{2^{\binom{m}{2}}}.$$

Claim 5.10. *Let $\epsilon < 1/16$. If an affine subspace W satisfies*

$$\frac{|W \cap f^{-1}(1)|}{|W|} \leq 8\epsilon \frac{|f^{-1}(1)|}{2^{\binom{m}{2}}},$$

then $\text{co-dim}(W) \geq m/3$.

Proof of Claim 5.9. Since $\mu(W) = \mu(W \cap f^{-1}(1)) + \mu(W \cap f^{-1}(0))$,

$$\mu(W \cap f^{-1}(1)) \leq 4\epsilon\mu(W) \implies \mu(W \cap f^{-1}(1)) \leq \frac{4\epsilon}{1-4\epsilon}\mu(W \cap f^{-1}(0)).$$

Note that

$$\mu(W \cap f^{-1}(1)) \geq \frac{\mu_1(W \cap f^{-1}(1))}{2} = \frac{1}{2}\frac{|W \cap f^{-1}(1)|}{|f^{-1}(1)|} \text{ and}$$
$$\mu(W \cap f^{-1}(0)) \leq \frac{\mu_0(W)}{2} = \frac{1}{2}\frac{|W|}{2^{\binom{m}{2}}}.$$

$$\text{Therefore,}\quad \frac{|W \cap f^{-1}(1)|}{|f^{-1}(1)|} \leq \frac{4\epsilon}{1-4\epsilon}\frac{|W|}{2^{\binom{m}{2}}}.$$

Cross multiplying and using the assumption that $\epsilon \leq 1/8$,

$$\frac{|W \cap f^{-1}(1)|}{|W|} \leq 8\epsilon \frac{|f^{-1}(1)|}{2^{\binom{m}{2}}}.$$

□

Proof of Claim 5.10. Let S be the set of inputs X that represent a graph with a sink, i.e. $S = f^{-1}(1)$. Let $S_i \subset S$ be the set of inputs in which the graph represented has vertex v_i as a sink.

Consider the set I of all $i \in [m]$ such that

$$\frac{|W \cap S_i|}{|W|} \leq 16\epsilon \frac{|S_i|}{2^{\binom{m}{2}}}.$$

Then, by the condition assumed in the claim

$$8\epsilon \frac{|S|}{2^{\binom{m}{2}}} \geq \frac{|W \cap S|}{|W|} = \sum_{i=1}^{m} \frac{|W \cap S_i|}{|W|} \geq \sum_{i \in \bar{I}} \frac{|W \cap S_i|}{|W|} > |\bar{I}| 16\epsilon \frac{|S|/m}{2^{\binom{m}{2}}} = |\bar{I}| \cdot \frac{2}{m} \cdot 8\epsilon \frac{|S|}{2^{\binom{m}{2}}},$$

where the second inequality follows from the definition of I and the fact that $|S_i| = |S|/m$. Hence $|I| \geq m/2$, and for all $i \in I$,

$$\frac{|W \cap S_i|}{|W|} \leq 16\epsilon \frac{|S_i|}{2^{\binom{m}{2}}} = 16\epsilon 2^{-(m-1)}.$$

Fix any $i \in I$. Define the distribution W_{v_i} by the following sampling procedure: Sample an input uniformly at random from W and project it to E_{v_i}.

We know from the last inequality that,

$$\Pr_{X \sim W_{v_i}} [X = z_i] \leq 16\epsilon 2^{-(m-1)}.$$

But the distribution W_{v_i} is, by Claim 2.11, the uniform distribution over some affine subspace W' of size $\leq 2^{m-1}$. Hence every element in the support of W_{v_i} must have probability at least $2^{-(m-1)}$. So if $\epsilon < 1/16$, W_{v_i} cannot have z_i in its support.

So for all $i \in I$, v_i is never a sink in W. Since $|I| \geq m/2$, Lemma 5.1 implies that the co-dimension of W is at least $m/3$. □

Proof of Lemma 5.8. By chaining together Claim 5.9 and Claim 5.10, we get

$$\mu(W \cap f^{-1}(1)) < \frac{1}{4}\mu(W) \implies \text{co-dim}(W) \geq m/3.$$

□

6 Randomized Communication Lower Bound

In this section, we prove that the randomized communication complexity of $\mathsf{SINK} \circ \mathsf{XOR}$ is large. While it is similar to the RPDT lower bound, we will phrase this lower bound as a lower bound on corruption.

Let $F := \mathsf{SINK} \circ \mathsf{XOR}$. Again, we fix ν to be the distribution that is a half-and-half combination of "uniform over the universe" and "uniform over the 1-inputs". More precisely, let ν_0 be uniform over $\{0,1\}^{\binom{m}{2}+\binom{m}{2}}$ and ν_1 be uniform over $F^{-1}(1)$. Define $\nu := (\nu_0 + \nu_1)/2$.

Claim 6.1 (Balanced Distribution, analogous to Claim 5.6)**.** *For ν as defined above, $\nu(F^{-1}(1)) = 1/2 + o(1)$.*

The proof of the above is omitted due to syntactic equivalence with that of Claim 5.6.

Lemma 6.2 (Corruption Bound, analogous to Lemma 5.8)**.** *Let $\epsilon \leq 1/8$. Any rectangle R that has $\nu(R \cap F^{-1}(1)) \leq 4\epsilon\nu(R)$ satisfies*

$$\nu(R) \leq 2^{-m(1/2-32\epsilon)^2/(64\ln 2)}.$$

We will prove the above lemma in Subsection 6.1 after noting that it shows that $\mathsf{Corr}^0(F) \geq \Omega(m)$, and hence gives an $\Omega(m)$ randomized communication lower bound.

Setting $\epsilon = 1/128$, Lemma 6.2 says that any rectangle R with $\nu(R \cap F^{-1}(1)) \leq \nu(R)/32$ must have $\nu(R) < 2^{-m/800}$. Since we know from Claim 6.1 that ν is balanced, this shows that $\mathsf{Corr}^0_{1/32}(F) \geq m/800$. Using Fact 2.14, we conclude Theorem 1.9.

Theorem 1.9. $R_{1/3}(\mathsf{SINK} \circ \mathsf{XOR}) = \Theta(m)$.

6.1 The Corruption Bound

Lemma 6.2 states that any 0-biased rectangle under ν must have small ν-mass. We prove this in three steps. We show that any 0-biased rectangle must have a very small fraction (under the uniform distribution) of 1-inputs. We then show that any rectangle with a very small fraction of 1-inputs must be small. We finish the proof by showing that any 0-biased small rectangle must have small ν mass.

Claim 6.3 (Analogous to Claim 5.9)**.** *Let $\epsilon \leq 1/8$. A rectangle R such that $\nu(R \cap F^{-1}(1)) \leq 4\epsilon\nu(R)$ must satisfy*

$$\frac{|R \cap F^{-1}(1)|}{|R|} \leq 8\epsilon \frac{|F^{-1}(1)|}{2^{2\binom{m}{2}}}.$$

Claim 6.4 (Analogous to Claim 5.10)**.** *If a rectangle $R = A \times B$ satisfies*

$$\frac{|R \cap F^{-1}(1)|}{|F^{-1}(1)|} \leq 8\epsilon \frac{|R|}{2^{2\binom{m}{2}}},$$

then $\min\{|A|, |B|\} \leq 2^{\binom{m}{2} - m(1/2-32\epsilon)^2/(64\ln 2)}$.

Claim 6.5. *If a rectangle R satisfies $\nu(R \cap g^{-1}(1)) \leq \nu(R)/2$, then $\nu(R) \leq |R|/2^{2\binom{m}{2}}$.*

The proof of Claim 6.3 is omitted due to syntactic equivalence with the proof of Claim 5.9.

Proof idea of Claim 6.4: This proof goes via the following intuition.

1. The rectangle R has a very small fraction of sinks relative to its size.
2. Hence for many vertices, R has a very small fraction of those vertices as sinks.
3. Any vertex v that is a sink very rarely in R must have its E_v projections on Alice's side and Bob's side quite "different" from each other.
4. Hence, either Alice's or Bob's projections must be small.
5. All these projections being small for Alice, say, shows that A must be really small, thus completing the proof via Shearer's lemma.

We now formalize this intuition.

Proof of Claim 6.4. We define S to be $F^{-1}(1)$. For $i \in [m]$, S_i is defined as the subset of S in which the vertex v_i is the sink. We note, by applying the same argument that we did at the beginning of the proof of Claim 5.10, that there is a set $I \subseteq [m]$ of size at least $m/2$ such that for all $i \in I$,

$$\frac{|R \cap S_i|}{|R|} \leq 16\epsilon \frac{|S_i|}{2^{2\binom{m}{2}}} = 16\epsilon 2^{-(m-1)}.$$

Here, the proof departs from the parallel RPDT lower bound. Fix any $i \in I$. Define the distribution A_{v_i} by the following sampling procedure. Sample an input uniformly at random from A and project it to E_{v_i}. Similarly define B_{v_i}. Define $B'_{v_i} = B_{v_i} \oplus z_i$.

We now show that the distributions A_{v_i} and B'_{v_i} are far apart, and hence one of them has a loss in entropy by Claim 2.19. We use the notation $\alpha \in_U S$ to denote that α is drawn uniformly at random from S.

$$\begin{aligned}
16\epsilon 2^{-(m-1)} &\geq \frac{|R \cap S_i|}{|R|} \\
&= \Pr_{X,Y \in_U R}[X_{v_i} \oplus Y_{v_i} = z_i] \\
&= \mathop{\mathbb{E}}_{X \in_U A}\left[\mathop{\mathbb{E}}_{Y \in_U B}\left[\mathbb{1}_{X_{v_i} \oplus Y_{v_i} = z_i}\right]\right] \\
&= \mathop{\mathbb{E}}_{X \in_U A}\left[\Pr_{Y \in_U B}[X_{v_i} \oplus Y_{v_i} = z_i]\right] \\
&= \mathop{\mathbb{E}}_{X \sim A_{v_i}}\left[\Pr_{Y \sim B_{v_i}}[X \oplus Y = z_i]\right] \\
&= \mathop{\mathbb{E}}_{X \sim A_{v_i}}\left[\Pr_{Y \sim B'_{v_i}}[X = Y]\right].
\end{aligned}$$

Let

$$T = \left\{ x \in \mathrm{supp}(A_{v_i}) \,\middle|\, \Pr_{Y \sim B'_{v_i}}[Y = x] \leq 32\epsilon 2^{-(m-1)} \right\}.$$

By Markov's inequality, $A_{v_i}(T) \geq 1/2$. But T is defined such that $B'_{v_i}(T) \leq 32\epsilon 2^{-(m-1)} \cdot |\mathrm{supp}(A_{v_i})| \leq 32\epsilon$. Hence,

$$\begin{aligned}
&||A_{v_i} - B'_{v_i}||_1 \geq 1/2 - 32\epsilon \\
\implies &(1/2 - 32\epsilon)^2 \leq 8 \ln 2 \cdot (m - 1 - \min\{H(A_{v_i}), H(B'_{v_i})\}) && \text{(by Claim 2.19)} \\
\implies &\min\{H(A_{v_i}), H(B'_{v_i})\} \leq m - 1 - (1/2 - 32\epsilon)^2/(8 \ln 2).
\end{aligned}$$

Note that the distributions B_{v_i} and B'_{v_i} are the same distribution but for a relabelling of the elements in its support. Hence $H(B_{v_i}) = H(B'_{v_i})$.

Either Alice's side or Bob's side hence experiences a loss in entropy for at least half the projections in I. Without loss of generality, we assume it is Alice's side. Since $|I| \geq m/2$, the expected entropy (when uniformly sampling a projection) for Alice is at most $m - 1 - 1/4 \cdot (1/2 - 32\epsilon)^2/(8 \ln 2)$.

Note that each coordinate in Alice's input appears in exactly 2 out of the m projections. We now apply Shearer's lemma (Lemma 2.20) with $X \in_U A$ and S uniform over $\{E_{v_i}\}_{i \in [m]}$. We have $t = 2/m$ and $\mathbb{E}[H(X_S)] \leq m - 1 - (1/2 - 32\epsilon)^2/(32 \ln 2)$. Hence we can conclude that

$$H(X) \leq \frac{m}{2} \cdot \left(m - 1 - (1/2 - 32\epsilon)^2/(32 \ln 2)\right).$$

Since X is uniform over A, we also have

$$|A| \leq 2^{(m/2)\left(m-1-(1/2-32\epsilon)^2/(32 \ln 2)\right)} = 2^{\binom{m}{2} - m(1/2 - 32\epsilon)^2/(64 \ln 2)}.$$

$\square$

Proof of Claim 6.5.

$$\begin{aligned}
\frac{|R|}{2^{2\binom{m}{2}}} &\geq \frac{|R \cap F^{-1}(0)|}{2^{2\binom{m}{2}}} \\
&= \nu_0(R \cap F^{-1}(0)) \\
&= 2\nu(R \cap F^{-1}(0)) && (\nu_1 \text{ places no mass on 0-inputs}) \\
&\geq \nu(R). && \text{(by assumption)}
\end{aligned}$$

□

Proof of Lemma 6.2. By chaining together Claim 6.3, Claim 6.4 and Claim 6.5, we get that if $\epsilon \leq 1/8$ and $\nu(R \cap F^{-1}(1)) \leq 4\epsilon\nu(R)$, then

$$\nu(R) \leq \frac{|R|}{2^{2\binom{m}{2}}} \leq 2^{-m(1/2-32\epsilon)^2/(64\ln 2)}.$$

□

6.2 Large Approximate Nonnegative Rank

Proof of Theorem 1.11. To prove this theorem, we will need to introduce two measures: the rectangle bound, $\mathsf{rec}^z_\epsilon(F)$, and the smooth rectangle bound, $\mathsf{srec}^z_\epsilon(F)$, where $z \in \{0,1\}$. (We provide their definitions in Appendix B). Jain and Klauck [JK10] introduce the latter measure and show the following.

$$\mathsf{srec}^z_\epsilon(F) \geq \mathsf{rec}^z_\epsilon(F) \geq \frac{1}{2} \cdot \left(\frac{1}{2} - \epsilon\right) \cdot \widetilde{\mathsf{rec}}^z_{2\epsilon}(F)$$

where $\widetilde{\mathsf{rec}}^z_{2\epsilon}$ is a measure they define (definition provided in Appendix B) that is a variant of our definition of corruption. Upon translating it to our definition of corruption, it can be seen that the latter inequality becomes

$$\log\left(\mathsf{rec}^z_\epsilon(F)\right) \geq \log\left(\frac{1}{3} \cdot \left(\frac{1}{2} - \epsilon\right) \cdot (1+2\epsilon)\right) + \mathsf{Corr}^z_{2\epsilon/(1+2\epsilon)}(F).$$

Kol et al. [KMSY14] show that the approximate nonnegative rank is equivalent to the smooth rectangle bound. In particular they show that

$$\mathsf{rank}^+_\epsilon(M_{\overline{F}}) \geq \mathsf{srec}^0_{3\epsilon}(F).$$

Put together, we have

$$\log \mathsf{rank}^+_\epsilon(M_{\overline{F}}) \geq \log\left(\frac{1}{3} \cdot \left(\frac{1}{2} - 3\epsilon\right) \cdot (1+6\epsilon)\right) + \mathsf{Corr}^0_{6\epsilon/(1+6\epsilon)}(F)$$

Since F has an $\Omega(m)$ corruption bound for 0 rectangles (Lemma 6.2) with $\epsilon = 1/12$ (via error reduction for corruption, see Section 2), it follows that $\log \mathsf{rank}^+_{1/66}(M_{\overline{F}}) \geq \Omega(m)$. □

6.3 A Variant of SINK

Recall from Part 3 of Theorem 1.10 that the approximate nonnegative rank of $\mathsf{SINK} \circ \mathsf{XOR}$ is bounded above by $m^{O(\log m)}$. In this subsection, we define a variant of $\mathsf{SINK} \circ \mathsf{XOR}$ which still has small approximate rank, but the approximate nonnegative rank of both this function and its complement are large.

Define the function $\mathsf{VARSINK} : \{0,1\}^{1+\binom{m}{2}} \to \{0,1\}$ as follows. We interpret the last $\binom{m}{2}$ variables exactly the way we did for SINK. The output of the function is given below, where b is the first bit and x is the remaining $\binom{m}{2}$ bits.

$$\mathsf{VARSINK}(b, x) = b \oplus \mathsf{SINK}(x).$$

Let $M_{0,i}(b,x)$ be the $0-1$ indicator that is 1 if and only if $b=0$ and v_i is a sink in x. Similarly, let $M_{1,i}(b,x)$ be 1 if and only if $b=1$ and v_i is a sink in x. Let $M_1(b,x)$ be 1 if and only if $b=1$. Note that each of $M_{0,i}$ and $M_{1,i}$ and M_1 is a subcube. Also

$$\mathsf{VARSINK}(b,x) = \sum_{i=1}^{m} M_{0,i}(b,x) + M_1(b,x) - \sum_{i=1}^{m} M_{1,i}(b,x).$$

Hence $\left|\left|\widehat{\mathsf{VARSINK}}\right|\right|_1 \leq 2m+1$ and exactly as we analyzed for SINK, we have $\mathsf{rank}_{1/66}(M_{\mathsf{VARSINK}\circ\mathsf{XOR}}) \leq O(m^4)$.

Now, note that if we set $b=0$, $\overline{\mathsf{VARSINK}}(0,x) = \overline{\mathsf{SINK}}(x)$, so

$$\mathsf{rank}^+_{1/66}(M_{\overline{\mathsf{VARSINK}\circ\mathsf{XOR}}}) \geq 2^{\Omega(m)}.$$

If we set $b=1$, $\mathsf{VARSINK}(1,x) = \overline{\mathsf{SINK}}(x)$. Hence

$$\mathsf{rank}^+_{1/66}(M_{\mathsf{VARSINK}\circ\mathsf{XOR}}) \geq 2^{\Omega(m)}.$$

7 Conclusions

To recall, we construct a simple total function called SINK. Quite remarkably, this function and its natural lift, $\mathsf{SINK}\circ\mathsf{XOR}$, refute several conjectures in communication complexity and parity decision tree complexity as summarized below.

1. $\mathsf{SINK}\circ\mathsf{XOR}$ refutes Grolmusz's conjecture [Gro97].

2. It refutes the strong form of the Log-Approximate-Nonnegative rank conjecture (Conjecture 1.2) suggested relatively recently by Kol, Moran, Shpilka and Yehudayoff [KMSY14]. It is worth noting that, in contrast, the Log-Rank Conjecture for deterministic protocols is long known to be true when rank is replaced by nonnegative rank. The LANRC being true would allow for the compression of a protocol down to its information content (see Conjecture 3, [KMSY14]). This and various related notions of protocol compression are of fundamental interest and have been the topic of a lot of recent research [BBCR13, BR11, GKR16, BGKR18].

3. Both the above conjectures were implied by the LARC (formulated by Lee and Shraibman [LS09] about ten years ago), which therefore gets refuted. Approximate rank is known to dominate powerful analytic measures like generalized/smooth discrepancy. Our main result shows an exponential separation between the classical corruption bound and approximate rank. No such separation between corruption and even generalized/smooth discrepancy was known for any total function.

4. SINK disproves a conjecture on parity kill number made by Tsang et al. [TWXZ13]. This conjecture implied the Log-Rank Conjecture for XOR functions. SINK also falsifies the belief of Shpilka et al. [STV17] that the deterministic PDT size of f can be upper bounded by a quasi-polynomially growing function of its spectral norm.

5. Our result yields the following interesting situation for quantum communication complexity: either the quantum log-rank conjecture is falsified by $\mathsf{SINK}\circ\mathsf{XOR}$ or the function establishes exponential quantum supremacy over classical communication for computing even total Boolean functions. The latter would be a major breakthrough. Proving the former, on the other hand, raises an interesting challenge of developing new techniques. All known lower bounds for quantum complexity are within a polynomial of the log of the approximate rank. Anshu et al. [ABG$^+$17] gave a quadratic separation of quantum communication complexity from the logarithm of the approximate rank. Yet, $\mathsf{SINK}\circ\mathsf{XOR}$ could conceivably show an $O(\log m)$ vs $\Omega(\sqrt{m})$ separation. The $\tilde{O}(\sqrt{m})$ upper bound follows from composing an $O(\sqrt{m})$ cost Grover's search with $O(\log m)$ cost communication protocols for the equalities. (See [ABG$^+$17, BCW98] for details on such compositions).

To conclude, it is worth noting the following curious situation: $\log(\mathsf{rank}_\epsilon(M_f))$ is known to exactly characterize communication complexity when ϵ approaches $1/2$ from below, as was shown in the classical work of Paturi and Simon [PS86]. The LRC asks the corresponding question for $\epsilon = 0$ and remains wide open. Our result can be interpreted as saying that when $0 < \epsilon < 1/2$ is a constant, the logarithm of the rank measure completely fails to capture the corresponding communication complexity.

Acknowledgements

The first author thanks Shachar Lovett for stimulating discussions and Troy Lee for answering several questions about approximate rank and related notions.

References

[ABG+17] Anurag Anshu, Shalev Ben-David, Ankit Garg, Rahul Jain, Robin Kothari, and Troy Lee. Separating quantum communication and approximate rank. In *32nd Computational Complexity Conference, CCC 2017*, pages 24:1–24:33, 2017.

[ABK16] Scott Aaronson, Shalev Ben-David, and Robin Kothari. Separations in query complexity using cheat sheets. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 863–876, 2016.

[AFH12] 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.

[BBCR13] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. *SIAM J. Comput.*, 42(3):1327–1363, 2013.

[BCW98] Harry Buhrman, Richard Cleve, and Avi Wigderson. Quantum vs. classical communication and computation. In *Proceedings of the 30th Annual ACM SIGACT Symposium on Theory of Computing, STOC*, pages 63–68, 1998.

[BGKR18] Mark Braverman, Anat Ganor, Gillat Kol, and Ran Raz. A candidate for a strong separation of information and communication. In *9th Innovations in Theoretical Computer Science Conference, ITCS 2018, January 11-14, 2018, Cambridge, MA, USA*, pages 11:1–11:13, 2018.

[BR11] Mark Braverman and Anup Rao. Information equals amortized communication. In *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011, Palm Springs, CA, USA, October 22-25, 2011*, pages 748–757, 2011.

[BS90] Jehoshua Bruck and Roman Smolensky. Polynomial threshold functions, AC^0 functions and spectral norms (extended abstract). In *31st Annual Symposium on Foundations of Computer Science, St. Louis, Missouri, USA, October 22-24, 1990, Volume II*, pages 632–641, 1990.

[CKLM17] Arkadev Chattopadhyay, Michal Koucký, Bruno Loff, and Sagnik Mukhopadhyay. Simulation theorems via pseudorandom properties. *CoRR*, abs/1704.06807, 2017.

[FMP+15] Samuel Fiorini, Serge Massar, Sebastian Pokutta, Hans Raj Tiwary, and Ronald de Wolf. Exponential lower bounds for polytopes in combinatorial optimization. *J. ACM*, 62(2):17:1–17:23, 2015.

[Gav16] Dmitry Gavinsky. Entangled simultaneity versus classical interactivity in communication complexity. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 877–884, 2016.

[GJPW17] Mika Göös, T. S. Jayram, Toniann Pitassi, and Thomas Watson. Randomized communication vs. partition number. In *Automata, Languages, and Programming - 44th International Colloquium on Automata, languages and Programming, ICALP*, pages 52:1–52:15, 2017.

[GKR16] Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of communication and external information. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 977–986, 2016.

[GL14] Dmitry Gavinsky and Shachar Lovett. En route to the log-rank conjecture: New reductions and equivalent formulations. In *Automata, Languages, and Programming - 41st International Colloquium, ICALP 2014, Copenhagen, Denmark, July 8-11, 2014, Proceedings, Part I*, pages 514–524, 2014.

[GPW15] Mika Göös, Toniann Pitassi, and Thomas Watson. Deterministic communication vs. partition number. In *56th Annual Symposium on Foundations of Computer Science, FOCS*, pages 1077–1088, 2015.

[GPW17] Mika Göös, Toniann Pitassi, and Thomas Watson. Query-to-communication lifting for BPP. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS*, pages 132–143, 2017.

[Gro97] Vince Grolmusz. On the power of circuits with gates of low l_1 norms. *Theor. Comput. Sci.*, 188(1-2):117–128, 1997.

[GS08] Ben Green and Tom Sanders. Boolean functions with small spectral norm. *Geometric and Functional Analysis (GAFA)*, 18:144–162, 2008.

[HHL18] Hamed Hatami, Kaave Hosseini, and Shachar Lovett. Structure of protocols for XOR functions. *SIAM J. Comput.*, 47(1):208–217, 2018.

[Hoe63] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

[JK10] Rahul Jain and Hartmut Klauck. The partition bound for classical communication complexity and query complexity. In *Proceedings of the 25th Annual IEEE Conference on Computational Complexity, CCC 2010, Cambridge, Massachusetts, USA, June 9-12, 2010*, pages 247–258, 2010.

[Kla03] Hartmut Klauck. Rectangle size bounds and threshold covers in communication complexity. In *18th Annual IEEE Conference on Computational Complexity (Complexity 2003), 7-10 July 2003, Aarhus, Denmark*, pages 118–134, 2003.

[KMR17] Pravesh K. Kothari, Raghu Meka, and Prasad Raghavendra. Approximating rectangles by juntas and weakly-exponential lower bounds for LP relaxations of csps. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC*, pages 590–603, 2017.

[KMSY14] Gillat Kol, Shay Moran, Amir Shpilka, and Amir Yehudayoff. Approximate nonnegative rank is equivalent to the smooth rectangle bound. In *Automata, Languages, and Programming - 41st International Colloquium, ICALP 2014, Copenhagen, Denmark, July 8-11, 2014, Proceedings, Part I*, pages 701–712, 2014.

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, 1997.

[Kra96] Matthias Krause. Geometric arguments yield better bounds for threshold circuits and distributed computing. *Theor. Comput. Sci.*, 156(1&2):99–117, 1996.

[KS92] Bala Kalyanasundaram and Georg Schnitger. The probabilistic communication complexity of set intersection. *SIAM J. Discrete Math.*, 5(4):545–557, 1992.

[Lee12] Troy Lee. Some open problems about nonnegative rank. `http://research.cs.rutgers.edu/~troyjlee/open_problems.pdf`, 2012.

[Lov90] László Lovász. Communication complexity: A survey. In H. Prömel B. Korte, L. Lovász and A. Schrijver, editors, *Paths, flows, and VLSI-layout*, pages 235–265. Springer-Verlag, 1990.

[Lov14] Shachar Lovett. Recent advances on the log-rank conjecture in communication complexity. *Bulletin of the EATCS*, 112, 2014.

[Lov16] Shachar Lovett. Communication is bounded by root of rank. *J. ACM*, 63(1):1:1–1:9, 2016.

[LS88] László Lovász and Michael E. Saks. Lattices, möbius functions and communication complexity. In *29th Annual Symposium on Foundations of Computer Science, White Plains, New York, USA, 24-26 October 1988*, pages 81–90, 1988.

[LS09] Troy Lee and Adi Shraibman. Lower bounds in communication complexity. *Foundations and Trends in Theoretical Computer Science*, 3(4):263–398, 2009.

[OWZ+14] Ryan O'Donnell, John Wright, Yu Zhao, Xiaorui Sun, and Li-Yang Tan. A composition theorem for parity kill number. In *IEEE 29th Conference on Computational Complexity, CCC 2014, Vancouver, BC, Canada, June 11-13, 2014*, pages 144–154, 2014.

[PS86] Ramamohan Paturi and Janos Simon. Probabilistic communication complexity. *J. Comput. Syst. Sci.*, 33(1):106–123, 1986.

[Raz92] Alexander A. Razborov. On the distributional complexity of disjointness. *Theoretical Computer Science*, 106:385–390, 1992.

[Raz03] Alexander A. Razborov. Quantum communication complexity of symmetric predicates. *Izvestia: Mathematics*, 67(1):145–159, 2003.

[San18] Tom Sanders. Boolean functions with small spectral norm, revisited. *Arxiv*, 2018.

[STV17] Amir Shpilka, Avishay Tal, and Ben Lee Volk. On the structure of boolean functions with small spectral norm. *Computational Complexity*, 26(1):229–273, 2017.

[TWXZ13] Hing Yin Tsang, Chung Hoi Wong, Ning Xie, and Shengyu Zhang. Fourier sparsity, spectral norm, and the log-rank conjecture. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 658–667, 2013.

[TXZ16] Hing Yin Tsang, Ning Xie, and Shengyu Zhang. Fourier sparsity of GF(2) polynomials. In *11th International Computer Science Symposium in Russia, CSR*, pages 409–424, 2016.

[Yan91] Mihalis Yannakakis. Expressing combinatorial optimization problems by linear programs. *J. Comput. Syst. Sci.*, 43(3):441–466, 1991.

[Yao79] 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.

[Zha14] Shengyu Zhang. Efficient quantum protocols for XOR functions. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 1878–1885, 2014.

[ZS09] Zhiqiang Zhang and Yaoyun Shi. Communication complexities of symmetric XOR functions. *Quantum Information & Computation*, 9(3):255–263, 2009.

A LARC implies Grolmusz's Conjecture

In this section, we observe that it is implicit from prior work that the LARC would have implied Grolmusz's conjecture, up to a polylogarithmic factor in the input size.

Lemma A.1 (LARC implies Grolmusz's conjecture)**.** *For any function* $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$,

$$R_{1/3}(F) \le \log^{O(1)} \mathsf{rank}_{1/3}(M_F) \implies R_{1/3}(F) \le (\log \left|\left|\widehat{F}\right|\right|_1 + \log n)^{O(1)}.$$

The observation follows from Lemma 2.8, which we restate here for convenience and prove for completeness.

Lemma 2.8. *For any* $g : \{0,1\}^n \to \mathbb{R}$ *and* $\delta > \epsilon \ge 0$,

$$\left|\left|\widehat{g}\right|\right|_{0,\delta} \le O\left(\left|\left|\widehat{g}\right|\right|_{1,\epsilon}^2 n/(\delta-\epsilon)^2\right).$$

Proof. Let $g : \{0,1\}^N \to \mathbb{R}$ be any function. Let g' be a function witnessing $\left|\left|\widehat{g}\right|\right|_{1,\epsilon}$. Consider the distribution over subsets of $[n]$ where the probability of sampling α is $|\widehat{g'}(\alpha)|/\left|\left|\widehat{g'}\right|\right|_1$. Draw $M = O\left(\left|\left|\widehat{g'}\right|\right|_1^2 n\log(1/\lambda)/(\delta-\epsilon)^2\right)$ samples $\alpha^1, \ldots, \alpha^M$ independently from this distribution. Define the function

$$h(x) = \frac{\left|\left|\widehat{g'}\right|\right|_1}{M}\sum_{i=1}^{M} \operatorname{sign}(\widehat{g'}(\alpha^i))\chi_{\alpha^i}(x).$$

Then,

$$\Pr[\forall x \in \{0,1\}^n, |h(x) - g'(x)| \le \delta - \epsilon] \ge 1 - \lambda. \tag{1}$$

Setting λ to be any constant in $(0,1)$ yields an h that witnesses the small approximate sparsity of g.

The proof of Equation (1) follows from a union bound over all x, after using Hoeffding's inequality.

Fix an input $x \in \{0,1\}^n$. We use Hoeffding's inequality (Lemma 2.1) with $X_i = \operatorname{sign}(\widehat{g'}(\alpha^i))\chi_{\alpha^i}(x)$. Hence $X = \frac{M}{\left|\left|\widehat{g'}\right|\right|_1}h(x)$ and

$$\mathbb{E}[X] = \sum_{i=1}^{M}\mathbb{E}[\operatorname{sign}(\widehat{g'}(\alpha^i))\chi_{\alpha^i}(x)] = \sum_{i=1}^{M}\sum_{\alpha}\frac{|\widehat{g'}(\alpha)|}{\left|\left|\widehat{g'}\right|\right|_1}\operatorname{sign}(\widehat{g'}(\alpha))\chi_\alpha(x) = \frac{M}{\left|\left|\widehat{g'}\right|\right|_1}g'(x).$$

The lemma gives us that

$$\Pr\left[\left|\frac{M}{\left|\left|\widehat{g'}\right|\right|_1}h(x) - \frac{M}{\left|\left|\widehat{g'}\right|\right|_1}g'(x)\right| \ge \frac{M}{\left|\left|\widehat{g'}\right|\right|_1}(\delta-\epsilon)\right] < 2\exp\left(-\frac{2M^2(\delta-\epsilon)^2/\left|\left|\widehat{g'}\right|\right|_1^2}{4M}\right) = 2\exp\left(-\frac{M(\delta-\epsilon)^2}{2\left|\left|\widehat{g'}\right|\right|_1^2}\right).$$

Hence for $M = O(\left|\left|\widehat{g'}\right|\right|_1^2 n\log(1/\lambda)/(\delta-\epsilon)^2)$, we can make $\Pr[|h(x) - g'(x)| > \delta - \epsilon] < \lambda 2^{-n}$. By a union bound over all x, it is clear that with probability $\ge 1-\lambda$, our sampling of h is such that $\forall x \in \{0,1\}^n, |h(x) - g'(x)| \le \delta - \epsilon$.

□

Proof of Lemma A.1. Denote $w = \left|\left|\widehat{F}\right|\right|_1$. Lemma 2.8 implies existence of a function $G = \sum_{S\subseteq[n]\times[n]} c_S\chi_S$ such that $|G(x) - F(x)| \le 1/3$ for all $x \in \{0,1\}^n \times \{0,1\}^n$ and $\left|\left|\widehat{G}\right|\right|_0 = O(\left|\left|\widehat{F}\right|\right|_1^2 n)$.

Next, note that for any $S \subseteq [n]\times[n]$, the function $c_S\chi_S$ is a matrix of rank at most 4. By the sub-additivity of rank, $\log \mathsf{rank}(G)$, and thus $\log \mathsf{rank}_{1/3}(F)$, is at most $O(\log \left|\left|\widehat{F}\right|\right|_1 + \log n)$. □

B Definitions of rectangle bounds

In this section, we define a few measures as defined in [JK10] (but for total Boolean functions). Let $F : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ and $\mathcal{R}$ be the set of all rectangles in $\mathcal{X} \times \mathcal{Y}$. z takes values in $\{0,1\}$.

Definition B.1 (Rectangle Bound). $\mathsf{rec}^z_\epsilon(F)$ *is defined to be the optimal value of the following linear program.*

$$
\begin{array}{lll}
\textit{Variables} & \{w_R : R \in \mathcal{R}\} & \\
\textit{Minimize} & \displaystyle\sum_{R \in \mathcal{R}} w_R & \\
\textit{s.t.} & \forall (x,y) \in F^{-1}(z) & \displaystyle\sum_{R:(x,y)\in R} w_R \geq 1-\epsilon \\
 & \forall (x,y) \in F^{-1}(\bar{z}) & \displaystyle\sum_{R:(x,y)\in R} w_R \leq \epsilon \\
 & \forall R \in \mathcal{R} & w_R \geq 0
\end{array}
$$

Definition B.2 (Rectangle Bound: Conventional Definition). $\widetilde{\mathsf{rec}}^z_\epsilon(F)$ *is defined as follows.*

$$\widetilde{\mathsf{rec}}^z_\epsilon(F) = \max_\mu \min_R \frac{1}{\mu(R \cap F^{-1}(z))}$$

where μ and R range over

- *Heavy distributions: Distributions μ on $\mathcal{X} \times \mathcal{Y}$ with $\mu(F^{-1}(z)) \geq 0.5$.*
- *Biased rectangles: Rectangles $R \in \mathcal{R}$ with $\mu(R \cap F^{-1}(\bar{z})) < \epsilon\mu(R \cap F^{-1}(z))$.*

Definition B.3 (Smooth Rectangle Bound). $\mathsf{srec}^z_\epsilon(F)$ *is defined to be the optimal value of the following linear program.*

$$
\begin{array}{lll}
\textit{Variables} & \{w_R : R \in \mathcal{R}\} & \\
\textit{Minimize} & \displaystyle\sum_{R \in \mathcal{R}} w_R & \\
\textit{s.t.} & \forall (x,y) \in F^{-1}(z) & \displaystyle\sum_{R:(x,y)\in R} w_R \geq 1-\epsilon \\
 & \forall (x,y) \in F^{-1}(z) & \displaystyle\sum_{R:(x,y)\in R} w_R \leq 1 \\
 & \forall (x,y) \in F^{-1}(\bar{z}) & \displaystyle\sum_{R:(x,y)\in R} w_R \leq \epsilon \\
 & \forall R \in \mathcal{R} & w_R \geq 0
\end{array}
$$

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il