2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS) | 979-8-3503-1894-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/FOCS57990.2023.00047

Fourier Growth of Communication Protocols for XOR Functions

Uma Girish[‡]
Princeton University
ugirish@cs.princeton.edu

Makrand Sinha[*]
University of Illinois at Urbana-Champaign
msinha@illinois.edu

Avishay Tal[†]
University of California at Berkeley
atal@berkeley.edu

Kewen Wu[†]
University of California at Berkeley
shlw_kevin@hotmail.com

***Abstract*—The level-k ℓ_1-Fourier weight of a Boolean function refers to the sum of absolute values of its level-k Fourier coefficients. Fourier growth refers to the growth of these weights as k grows. It has been extensively studied for various computational models, and bounds on the Fourier growth, even for the first few levels, have proven useful in learning theory, circuit lower bounds, pseudorandomness, and quantum-classical separations.**

In this work, we investigate the Fourier growth of certain functions that naturally arise from communication protocols for XOR functions (partial functions evaluated on the bitwise XOR of the inputs x and y to Alice and Bob). If a protocol $\mathcal{C}$ computes an XOR function, then $\mathcal{C}(x, y)$ is a function of the parity $x \oplus y$. This motivates us to analyze the *XOR-fiber* of the communication protocol $\mathcal{C}$, defined as $h(z) := \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}}[\mathcal{C}(\boldsymbol{x}, \boldsymbol{y}) | \boldsymbol{x} \oplus \boldsymbol{y} = z]$.

We present improved Fourier growth bounds for the XOR-fibers of randomized protocols that communicate d bits. For the first level, we show a tight $O(\sqrt{d})$ bound and obtain a new coin theorem, as well as an alternative proof for the tight randomized communication lower bound for the Gap-Hamming problem. For the second level, we show an $d^{3/2} \cdot \mathrm{polylog}(n)$ bound, which improves the previous $O(d^2)$ bound by Girish, Raz, and Tal (ITCS 2021) and implies a polynomial improvement on the randomized communication lower bound for the XOR-lift of the Forrelation problem, which extends the quantum-classical gap for this problem.

Our analysis is based on a new way of adaptively partitioning a relatively large set in Gaussian space to control its moments in all directions. We achieve this via martingale arguments and allowing protocols to transmit real values. We also show a connection between Fourier growth and lifting theorems with constant-sized gadgets as a potential approach to prove optimal bounds for the second level and beyond.

***Index Terms*—Fourier growth, communication protocol, analysis of Boolean functions, quantum-classical separation**

I. Introduction

The Fourier spectrum of Boolean functions and their various properties have played an important role in many areas of mathematics and theoretical computer science. In this work, we study a notion called ℓ_1-Fourier growth, which captures the scaling of the sum of absolute values of the level-k Fourier coefficients of a function. In a nutshell, functions with small Fourier growth cannot aggregate many weak signals in the input to obtain a considerable effect on the output. In contrast, the Majority function, which can amplify weak biases, is an example of a Boolean function with extremely *high* Fourier growth.

To formally define Fourier growth, we recall that every Boolean function $f : \{\pm 1\}^n \to [-1, 1]$ can be uniquely represented as a multilinear polynomial

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S) \cdot \prod_{i \in S} x_i$$

where the coefficients of the polynomial $\widehat{f}(S) \in \mathbb{R}$ are called the Fourier coefficients of f, and they satisfy $\widehat{f}(S) = \mathbb{E}[f(\boldsymbol{x}) \cdot \prod_{i \in S} \boldsymbol{x}_i]$ for a uniformly random $\boldsymbol{x} \in \{\pm 1\}^n$. The level-$k$ ℓ_1-Fourier growth of f is the sum of the *absolute values* of its level-k Fourier coefficients,

$$L_{1,k}(f) := \sum_{S \subseteq [n] : |S| = k} \left| \widehat{f}(S) \right| .$$

The study of Fourier growth dates back to the work of Mansour [1] who used it in the context of learning algorithms. Since then, several works have shown that upper bounds on the Fourier growth, even for the first few Fourier levels, have applications to pseudorandomness, circuit complexity, and quantum-classical separations. For example:

- A bound on the level-one Fourier growth is sufficient to control the advantage of distinguishing biased coins from unbiased ones [2].
- A bound on the level-two Fourier growth already gives pseudorandom generators [3], oracle separations between BQP and PH [4], [5], and separations between efficient quantum communication and randomized classical communication [6].

Meanwhile, Fourier growth bounds have been extensively studied and established for various computational models,

[*] Part of this work was done at the Simons Institute for the Theory of Computing and supported by a Simons-Berkeley Postdoctoral Fellowship.

[†] Supported by a Sloan Research Fellowship and NSF CAREER Award CCF-2145474.

[‡] Part of this work was done at the Simons Institute for the Theory of Computing. Supported by the Simons Collaboration on Algorithms and Geometry, a Simons Investigator Award, by the National Science Foundation grants No. CCF-1714779, CCF-2007462 and by the IBM PhD Fellowship.

979-8-3503-1894-4/23/$31.00 ©2023 IEEE
DOI 10.1109/FOCS57990.2023.00047

including small-width DNFs/CNFs [1], AC^0 circuits [7], low-sensitivity Boolean functions [8], small-width branching programs [9], [10], [11], [12], small-depth decision trees [13], [14], [15], functions related to small-cost communication protocols [16], [6], low-degree $\mathsf{GF}(2)$ polynomials [17], [3], [18], product tests [19], small-depth parity decision trees [20], [21], low-degree bounded functions [22], and more.

For any Boolean function f with outputs in $[-1, 1]$, the level-k Fourier growth $L_{1,k}(f)$ is at most $\sqrt{\binom{n}{k}}$. However, for many natural classes of Boolean functions, this bound is far from tight and not good enough for applications. Establishing better bounds require exploring structural properties of the specific class of functions in question. Even for low Fourier levels, this can be highly non-trivial and tight bounds remain elusive in many cases. For example, for degree-d $\mathsf{GF}(2)$ polynomials (which well-approximate $\mathsf{AC}^0[\oplus]$ when we set $d = \mathrm{polylog}(n)$ [23], [24]), while we know a level-one bound of $L_{1,1}(f) \le O(d)$ due to [3], the current best bound for levels $k \ge 2$ is roughly $2^{O(dk)}$ [17], whereas the conjectured bound is $d^{O(k)}$. Validating such a bound, even for the second level $k = 2$, will imply unconditional pseudorandom generators of polylogarithmic seed length for $\mathsf{AC}^0[\oplus]$ [3], a longstanding open problem in circuit complexity and pseudorandomness.

a) XOR Functions: In this work, we study the Fourier growth of certain functions that naturally arise from communication protocols for XOR-lifted functions, also referred to as XOR functions. XOR functions are an important and well-studied class of functions in communication complexity with connections to the log-rank conjecture and quantum versus classical separations [25], [26], [27], [28], [29].

In this setting, Alice gets an input $x \in \{\pm 1\}^n$ and Bob gets an input $y \in \{\pm 1\}^n$ and they wish to compute $f(x \odot y)$ where f is some partial Boolean function and $x \odot y$ is in the domain of f. Here, $x \odot y$ denotes the pointwise product of x and y. Given any communication protocol $\mathcal{C}$ that computes an XOR function exactly, the output $\mathcal{C}(x, y)$ of the protocol depends only on the parity $x \odot y$, whenever f is defined on $x \odot y$. This gives a natural motivation to analyze the XOR-fiber of a communication protocol defined below. We note that a similar notion first appeared in an earlier work of Raz [30].

Definition I.1. Let $\mathcal{C} : \{\pm 1\}^n \times \{\pm 1\}^n \to \{\pm 1\}$ be any deterministic communication protocol. The XOR-fiber of the communication protocol $\mathcal{C}$ is the function $h\colon \{\pm 1\}^n \to [-1, 1]$ defined at $z \in \{\pm 1\}^n$ as

$$h(z) = \mathop{\mathbb{E}}_{\boldsymbol{x},\boldsymbol{y}\sim\nu}[\mathcal{C}(\boldsymbol{x}, \boldsymbol{y}) \mid \boldsymbol{x} \odot \boldsymbol{y} = z],$$

where $\odot$ is the entrywise product and ν is the uniform distribution over $\{\pm 1\}^n$.

We remark that XOR-fiber is the "inverse" of XOR-lift of a function: If $\mathcal{C}$ computes the XOR function of f, then the XOR-fiber h of $\mathcal{C}$ is equal to f on the domain of f.

In this work, we investigate the Fourier growth of XOR-fibers of small-cost communication protocols and apply these bounds in several contexts. Before stating our results, we first discuss several related works.

b) Related Works: Showing optimal Fourier growth bounds for XOR-fibers is a complex undertaking in general and a first step towards this end is to obtain optimal Fourier growth bounds for parity decision trees. This is because a parity decision tree for a Boolean function f naturally gives rise to a structured communication protocol for the XOR-function corresponding to f. This protocol perfectly simulates the parity decision tree by having Alice and Bob exchange one bit each to simulate a parity query. Moreover, the XOR-fiber of this protocol exactly computes the parity decision tree. As such, parity decision trees can be seen as a special case of communication protocols, and Fourier growth bounds on XOR-fibers of communication protocols immediately imply Fourier growth bounds on parity decision trees.

Fourier growth bounds for decision trees and parity decision trees are well-studied. It is not too difficult to obtain a level-k bound of $O(d)^k$ for parity decision trees of depth d, however, obtaining improved bounds is significantly more challenging. For decision trees of depth d (which form a subclass of parity decision trees of depth d), O'Donnell and Servedio [13] proved a tight bound of $O(\sqrt{d})$ on the level-one Fourier growth. By inductive tree decompositions, Tal [14] obtained bounds for the higher levels of the form $L_{1,k}(f) \le \sqrt{d^k \cdot O(\log(n))^{k-1}}$. This was later sharpened by Sherstov, Storozhenko, and Wu [15] to the asymptotically tight bound of $L_{1,k}(f) \le \sqrt{\binom{d}{k} \cdot O(\log(n))^{k-1}}$ using a more sophisticated layered partitioning strategy on the tree.

When it comes to parity decision trees, despite all the similarities, the structural decomposition approach does not seem to carry over due to the correlations between the parity queries. For parity decision trees of depth d, Blais, Tan, and Wan [20] proved a tight level-one bound of $O(\sqrt{d})$. For higher levels, Girish, Tal, and Wu [21] showed that $L_{1,k}(f) \le \sqrt{d^k \cdot O(k \log(n))^{2k}}$. These works imply almost tight Fourier growth bounds on the XOR-fibers of structured protocols that arise from simulating decision trees or parity decision trees.

For the case of XOR-fibers of arbitrary deterministic/randomized communication protocols (which do not necessarily simulate parity decision trees or decision trees), Girish, Raz, and Tal [6] showed an $O(d^k)$ Fourier growth[1] for level-k. For level-one and level-two, these bounds are $O(d)$ and $O(d^2)$ respectively and are sub-optimal — as mentioned previously, such weaker bounds for parity decision trees are easy to obtain, while obtaining optimal bounds (for parity decision trees) of $O(\sqrt{d})$ for level one and $d \cdot \mathrm{polylog}(n)$ for level two already requires sophisticated ideas.

The bounds in [6] follow by analyzing the Fourier growth of XOR-fibers of communication rectangles of measure $\approx 2^{-d}$ and then adding up the contributions from all the leaf rect-

[1] Technically, [6] only proved a level-two bound (as it suffices for their analysis), but a level-k bound follows easily from their proof approach, as noted by [16]

angles induced by the protocol. Such a per-rectangle-based approach cannot give better bounds than the ones in [6], while they also conjectured that the optimal Fourier growth of XOR-fibers of arbitrary protocols should match the growth for parity decision trees.

Showing the above is a challenging task even for the first two Fourier levels. The difficulty arises primarily since in the absence of a per-rectangle-based argument, one has to crucially leverage cancellations between different rectangles induced by the communication protocol. In the simpler case of parity decision trees (or protocols that exchange parities), such cancellations are leveraged in [21] by ensuring k-wise independence at each node of the tree — this can be achieved by adding extra parity queries. In a general protocol, the parties can send arbitrary partial information about their inputs and correlate the coordinates in complicated ways that such methods break down. This is one of the key difficulties we face in this paper.

A. Main Results

We prove new and improved bounds on the Fourier growth of the XOR-fibers associated with small-cost protocols for levels $k = 1$ and $k = 2$.

Theorem I.2. *Let $\mathcal{C} : \{\pm 1\}^n \times \{\pm 1\}^n \to \{\pm 1\}$ be a deterministic communication protocol with at most d bits of communication. Let h be its XOR-fiber as in Definition I.1. Then, $L_{1,1}(h) = O\left(\sqrt{d}\right)$.*

Theorem I.3. *Let $\mathcal{C} : \{\pm 1\}^n \times \{\pm 1\}^n \to \{\pm 1\}$ be a deterministic protocol communicating at most d bits. Let h be its XOR-fiber as in Definition I.1. Then, $L_{1,2}(h) = O\left(d^{3/2} \log^3(n)\right)$.*

Our bounds in Theorems I.2 and I.3 extend directly to randomized communication protocols. This is because $L_{1,k}$ is convex and any randomized protocol is a convex combination of deterministic protocols with the same cost. Moreover, we can use Fourier growth reductions, as described in Subsection I-B3, to demonstrate that these bounds apply to general constant-sized gadgets g and the corresponding g-fiber.

Our level-one and level-two bounds improve previous bounds in [6] by polynomial factors. Additionally, our level-one bound is tight since a deterministic protocol with $d + 1$ bits of communication can compute the majority vote of $x_1 \cdot y_1, \ldots, x_d \cdot y_d$, which corresponds to $h(z) = \mathrm{MAJ}(z_1, \ldots, z_d)$ with $L_{1,1}(h) = \Theta(\sqrt{d})$. Furthermore, as we discuss later in Subsection I-B, level-one and level-two bounds are already sufficient for many interesting applications.

In terms of techniques, our analysis presents a key new idea that enables us to exploit cancellations between different rectangles induced by the protocol. This idea involves using a novel process to adaptively partition a relatively large set in Gaussian space, which enables us to control its k-wise moments in all directions — this can be thought of as a spectral notion of almost k-wise independence. We achieve this by utilizing martingale arguments and allowing protocols to transmit *real values* rather than just discrete bits. This notion and procedure may be of independent interest. See Section II for a detailed discussion.

B. Applications and Connections

Our main theorem has applications to XOR functions, and in more generality to functions lifted with constant-sized gadgets. In this setting, there is a simple gadget $g : \Sigma \times \Sigma \to \{\pm 1\}$ and a Boolean function f defined on inputs $z \in \{\pm 1\}^n$. The lifted function $f \circ g$ is defined on n pairs of symbols $(x_1, y_1), \ldots, (x_n, y_n) \in \Sigma \times \Sigma$ such that $(f \circ g)(x, y) = f(g(x_1, y_1), \ldots, g(x_n, y_n))$. The function $f \circ g$ naturally defines a communication problem where Alice is given $x = (x_1, \ldots, x_n)$, Bob is given $y = (y_1, \ldots, y_n)$, and they are asked to compute $(f \circ g)(x, y)$.

Since XOR functions are functions lifted with the XOR gadget, our main theorem implies lower bounds on the communication complexity of specific XOR functions. Additionally, we also show connections between XOR-lifting and lifting with any constant-sized gadget. Next, we describe these lower bounds and connections, with further context.

1) The Coin Problem and the Gap-Hamming Problem: The coin problem studies the advantage that a class of Boolean functions has in distinguishing biased coins from unbiased ones. More formally, let $\mathcal{F}$ be a class of n-variate Boolean functions. Let $\rho \in [-1, 1]$ and $\pi_\rho^{\otimes n}$ denote the product distribution over $\{\pm 1\}^n$ where each coordinate has expectation ρ. The Coin Problem asks what is the maximum advantage that functions in $\mathcal{F}$ have in distinguishing $\pi_\rho^{\otimes n}$ from the uniform distribution $\pi_0^{\otimes n}$.

This quantity essentially captures how well $\mathcal{F}$ can approximate threshold functions, and in particular, the majority function. The coin problem has been studied for various models of computation including branching programs [31], AC^0 and AC$^0[\oplus]$ circuits [32], [33], product tests [34], and more. Recently, Agrawal [2] showed that the coin problem is closely related to the level-one Fourier growth of functions in $\mathcal{F}$.

Lemma I.4 ([2, Lemma 3.2])**.** *Assume that $\mathcal{F}$ is closed under restrictions and satisfies $L_{1,1}(f) \le t$ for all $f \in \mathcal{F}$. Then, for all $\rho \in (-1, 1)$ and $f \in \mathcal{F}$,*

$$\left| \mathop{\mathbb{E}}_{z \sim \pi_\rho^{\otimes n}}[f(z)] - \mathop{\mathbb{E}}_{z \sim \pi_0^{\otimes n}}[f(z)] \right| \le \ln\left(\frac{1}{1-|\rho|}\right) \cdot t.$$

Note that communication protocols of small cost are closed under restrictions, so are their XOR-fibers (see [6, Lemma 5.5]). By noting that $\ln\left(\frac{1}{1-|\rho|}\right) \approx |\rho|$ for small values of ρ, we obtain the following corollary.[2] We also remark that, using the Fourier growth reductions (see Subsection I-B3), Theorem I.5 can be established for general gadgets of small size.

[2] Here we also use the fact that the upper bound $O(|\rho| \cdot \sqrt{d})$ is vacuous for large enough ρ as it is larger than 1.

Theorem I.5. *Let h be the XOR-fiber of a protocol with total communication d. Then for all ρ,*

$$\left| \mathop{\mathbb{E}}_{z \sim \pi_\rho^{\otimes n}}[h(z)] - \mathop{\mathbb{E}}_{z \sim \pi_0^{\otimes n}}[h(z)] \right| \leq O\Big(|\rho| \cdot \sqrt{d}\Big).$$

In particular, consider the following distinguishing task: Alice and Bob either receive two uniformly random strings in $\{\pm 1\}^n$ or they receive two uniformly random strings in $\{\pm 1\}^n$ conditioned on their XOR distributed according to $\pi_\rho^{\otimes n}$ for $\rho = 1/\sqrt{n}$ (the latter is often referred to as *ρ-correlated strings*). Theorem I.5 implies that any protocol communicating $o(n)$ bits cannot distinguish these two distributions with constant advantage. This is essentially a communication lower bound for the well-known Gap-Hamming Problem.

a) The Gap-Hamming Problem: In the Gap-Hamming Problem, Alice and Bob receive strings $x, y \in \{\pm 1\}^n$ respectively and they want to distinguish if $\langle x, y\rangle \leq -\sqrt{n}$ or $\langle x, y\rangle \geq \sqrt{n}$.

This is essentially the XOR-lift of the Coin Problem with $\rho = \pm 1/\sqrt{n}$ because the distribution of (x, y) conditioned on $x \odot y \sim \pi_\rho^{\otimes n}$ with $\rho = -1/\sqrt{n}$ and $\rho = 1/\sqrt{n}$ is mostly supported on the YES and NO instances of Gap-Hamming respectively. Thus immediately from Theorem I.5, we derive a new proof for the $\Omega(n)$ lower bound on the communication complexity of the Gap-Hamming Problem. The proof is deferred to the full version.

Theorem I.6. *The randomized communication complexity of the Gap-Hamming Problem is $\Omega(n)$.*

We note that there are various different proofs [35], [36], [37], [38] that obtain the above lower bound but the perspective taken here is perhaps conceptually simpler: (1) Gap-Hamming is essentially the XOR-lift of the Gap-Majority function, and (2) any function that approximates the Gap-Majority function must have large level-one Fourier growth, whereas XOR-fibers of small-cost protocols have small Fourier growth.

2) Quantum versus Classical Communication Separation via Lifting: One natural approach to proving quantum versus classical separations in communication complexity is via lifting: Consider a function f separating quantum and classical query complexity and lift it using a gadget g. Naturally, an algorithm computing f with few queries to z can be translated into a communication protocol computing $f \circ g$ where we replace each query to a bit z_i with a short conversation that allows the calculation of $z_i = g(x_i, y_i)$. Göös, Pitassi, and Watson [39] showed that for randomized query/communication complexity and for various gadgets, this is essentially the best possible. Such results are referred to as *lifting theorems*.

Lifting theorems apply to different models of computation, such as deterministic decision trees [40, 41], randomized decision trees [39, 42], and more. A beautiful line of work shows how to "lift" many lower bounds in the query model to the communication model [40], [41], [43], [44], [45], [26], [46], [47], [48], [49], [50], [51], [52], [53], [54]. For quantum query complexity, only one direction (considered the "easier" direction) is known: Any quantum query algorithm for f can be translated to a communication protocol for $f \circ g$ with a small logarithmic overhead [55]. It remains widely open whether the other direction holds as well. However, this query-to-communication direction for quantum, combined with the communication-to-query direction for classical, is already sufficient for lifting quantum versus classical separations from the query model to the communication model.

One drawback of this approach to proving communication complexity separations is that the state-of-the-art lifting results [42], [56] work for gadgets with alphabet size at least n (recall that n denotes f's input length) and it is a significant challenge to reduce the alphabet size to $O(1)$ or even $\mathrm{polylog}(n)$. These large gadgets will usually result in larger overheads in terms of communication rounds, communication bits, and computations for both parties. As demonstrated next, lifting with simpler gadgets like XOR allows for a simpler quantum protocol for the lifted problem.

a) Lifting Forrelation with XOR: The Forrelation function introduced by [57] is defined as follows: on input $x = (x_1, x_2) \in \{\pm 1\}^n$ where n is a power of 2,

$$\mathrm{Forr}(x) = \frac{2}{n}\langle Hx_1, x_2\rangle,$$

where H is the $(n/2) \times (n/2)$ (unitary) Hadamard matrix.

Girish, Raz, and Tal [6] studied the XOR-lift of the Forrelation problem and obtained new separations between quantum and randomized communication protocols. In more detail, they considered the partial function[3] $\mathrm{Forr} \circ \mathrm{XOR}\colon \{\pm 1\}^n \times \{\pm 1\}^n \to \{\pm 1\}$ defined as

$$\mathrm{Forr} \circ \mathrm{XOR}(x, y) = \begin{cases} 1 & \mathrm{Forr}(x \odot y) \geq \frac{1}{200 \ln(n/2)}, \\ -1 & \mathrm{Forr}(x \odot y) \leq \frac{1}{400 \ln(n/2)}, \end{cases}$$

and showed that if Alice and Bob use a randomized communication protocol, then they must communicate at least $\widetilde{\Omega}(n^{1/4})$ bits to compute $\mathrm{Forr} \circ \mathrm{XOR}$; while it can be solved by two entangled parties in the quantum simultaneous message passing model with a $\mathrm{polylog}(n)$-qubit communication protocol and additionally the parties can be implemented with efficient quantum circuits.

The lower bound in [6] was obtained from a second level Fourier growth bound (higher levels are not needed) on the XOR-fiber of classical communication protocols. Our level-two bound strengthens their bound and immediately gives an improved communication lower bound.

Theorem I.7. *The randomized communication complexity of $\mathrm{Forr} \circ \mathrm{XOR}$ is $\widetilde{\Omega}(n^{1/3})$.*

Theorem I.7 above gives an $\mathrm{polylog}(n)$ versus $\widetilde{\Omega}(n^{1/3})$ separation between the above quantum communication model and the randomized two-party communication model, improving upon the $\mathrm{polylog}(n)$ versus $\widetilde{\Omega}(n^{1/4})$ separation from

[3] We are overloading the notation here: technically, $\mathrm{Forr} \circ \mathrm{XOR}$ is the XOR-lift of the partial boolean function which on input x outputs 1 if $\mathrm{Forr}(x)$ is large and -1 if $\mathrm{Forr}(x)$ is small.

[6]. We emphasize that our separations are for players with *efficient quantum* running time, where the only prior separation was shown by the aforementioned work [6]. Such efficiency features can also benefit real-world implementations to demonstrate quantum advantage in experiments; for instance, one such proposal was introduced recently by Aaronson, Buhrman, and Kretschmer [58]. Without the efficiency assumption, a better $\operatorname{polylog}(n)$ versus $\widetilde{\Omega}(\sqrt{n})$ separation is known [59] (see [6, Section 1.1] for a more detailed comparison). Optimal Fourier growth bounds of $d \cdot \operatorname{polylog}(n)$ for level two, which we state later in Conjecture I.8, would also imply such a separation with XOR-lift of Forrelation.

b) Lifting k-Fold Forrelation with XOR: k-Fold Forrelation [60] is a generalization of the Forrelation problem and was originally conjectured to be a candidate that exhibits a maximal separation between quantum and classical query complexity. In a recent work, [61] showed that the randomized query complexity of k-Fold Forrelation is $\widetilde{\Omega}(n^{1-1/k})$, confirming this conjecture, and a similar separation was proven in [15] for variants of k-Fold Forrelation. These separations, together with lifting theorems with the *inner product* gadget [42], imply an $O(k\log(n))$ vs $\widetilde{\Omega}(n^{1-1/k})$ separation between two-party quantum and classical communication complexity, where additionally, the number of rounds[4] in the two-party quantum protocol is $2 \cdot \lceil k/2 \rceil$.

Replacing the inner product gadget with the XOR gadget above would yield an improved quantum-classical communication separation where the gadget is simpler and the number of rounds required by the quantum protocol to achieve the same quantitative separation is reduced by half. Bansal and Sinha [61] showed that for any computational model, small Fourier growth for the first $O(k^2)$-levels implies hardness of k-Fold Forrelation in that particular model. Thus, in conjunction with their results, to prove the above XOR lifting result for the k-Fold Forrelation problem, it suffices to prove the following Fourier growth bounds for XOR-fibers.

Conjecture I.8. *Let $\mathcal{C} : \{\pm 1\}^n \times \{\pm 1\}^n \to \{\pm 1\}$ be a deterministic communication protocol with at most d bits of communication. Let h be its XOR-fiber as in Definition I.1. Then for all $k \in \mathbb{N}$, we have that $L_{1,k}(h) \le (\sqrt{d} \cdot \operatorname{poly}(k, \log(n)))^k$.*

Note that these bounds are consistent with the Fourier growth of parity decision trees (or protocols that only send parities) as shown in [21].

We prove the above conjecture for the case $k = 1$ and make progress for the case $k = 2$. While our techniques can be extended to higher levels in a straightforward manner, the bounds obtained are farther from the conjectured ones. Thus, we decided to defer dealing with higher levels to future work as we believe one needs to first prove the *optimal* bound for level $k = 2$.

In the next subsection, we give another motivation to study the above conjecture by showing a connection to lifting theorems for constant-sized gadgets.

3) General Gadgets and Fourier Growth from Lifting: Our main results are Fourier growth bounds for XOR-fibers, which corresponds to XOR-lifts of functions. To complement this, we show that similar bounds hold for general lifted functions.

Let $g\colon \Sigma \times \Sigma \to \{\pm 1\}$ be a gadget and $\mathcal{C}\colon \Sigma^n \times \Sigma^n \to \{\pm 1\}$ be a communication protocol. Define the g-fiber of $\mathcal{C}$, denoted by $\mathcal{C}_{\downarrow g}\colon \{\pm 1\}^n \to [-1, 1]$, as

$$\mathcal{C}_{\downarrow g}(z) = \mathbb{E}\left[\mathcal{C}(\boldsymbol{x}, \boldsymbol{y}) \,|\, g(\boldsymbol{x}_i, \boldsymbol{y}_i) = z_i,\ \forall i\right],$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are uniform over Σ. We use $L_{1,k}(g, d)$ to denote the upper bound of the level-k Fourier growth for the g-fibers of protocols with at most d bits of communication. Using this notation, the XOR-fiber of $\mathcal{C}$ is simply $\mathcal{C}_{\downarrow \mathrm{XOR}}$, and our main results Theorems I.2 and I.3 can be rephrased as

$$L_{1,1}(\mathrm{XOR}, d) \le O\left(\sqrt{d}\right)$$

and

$$L_{1,2}(\mathrm{XOR}, d) \le O\left(d^{3/2} \log^3(n)\right).$$

In the full version of our paper, $L_{1,k}(g, d)$ is related to $L_{1,k}(\mathrm{XOR}, d)$, and the main takeaway is, in the study of Fourier growth bounds, constant-sized gadgets are all equivalent.

Theorem I.9 (Informal)**.** *Let $g\colon \Sigma \times \Sigma \to \{\pm 1\}$ be a "balanced" gadget. Then*

$$\frac{L_{1,k}(\mathrm{XOR}, d)}{|\Sigma|^k} \le L_{1,k}(g, d) \le |\Sigma|^k \cdot L_{1,k}(\mathrm{XOR}, d).$$

Theorem I.9 also proposes a different approach towards Conjecture I.8: it suffices to establish tight Fourier growth bound for g-fibers for some constant-sized (actually, polylogarithmic size suffices) gadget g, and then apply the reduction. The benefit of switching to a different gadget is that we can perhaps first prove a lifting theorem, and then appeal to the known Fourier growth bounds of (randomized) decision trees [14], [15].

As mentioned earlier, lifting theorems show how to simulate communication protocols of cost d for lifted functions with decision trees of depth at most $O(d)$ (see e.g., [39]). A problem at the frontier of this fruitful line of work has been establishing lifting theorems for decision trees with constant-sized gadgets. Note that the XOR gadget itself cannot have such a generic lifting result: Indeed, the parity function serves as a counterexample. Nevertheless, it is speculative that some larger gadget works, which suffices for our purposes.[5] On the other hand, for lifting from *parity* decision trees, we do know an XOR-lifting theorem [26]. However, it only holds for deterministic communication protocols and has a sextic blowup in the cost.

Thus, one can see Conjecture I.8 as either a further motivation for establishing lifting results for decision trees with

[4] We remark that for $k = 2$, this is exactly the XOR-lift of the Forrelation problem and can even be computed in the quantum simultaneous model, as shown in [6].

[5] In terms of the separations between quantum and classical communication, even restricted lifting results for the specific outer function being the Forrelation function would suffice.

constant-sized gadgets, or as a necessary milestone before proving such lifting results.

4) Pseudorandomness for Communication Protocols: We say $G\colon \{\pm1\}^\ell \to \{\pm1\}^n \times \{\pm1\}^n$ is a pseudorandom generator (PRG) for a (randomized) communication protocol $\mathcal{C}\colon \{\pm1\}^n \times \{\pm1\}^n \to [-1,1]$ with error ε and seed length ℓ if

$$\left| \mathop{\mathbb{E}}_{\boldsymbol{x},\boldsymbol{y}\sim\nu}[\mathcal{C}(\boldsymbol{x},\boldsymbol{y})] - \mathop{\mathbb{E}}_{\boldsymbol{r}\sim\{\pm1\}^\ell}[\mathcal{C}(G(\boldsymbol{r}))] \right| \le \varepsilon.$$

[62] showed that for the class of protocols sending at most d communication bits, there exists an explicit PRG of error 2^{-d} and seed length $n+O(d)$ from expander graphs. Note that the overhead n is inevitable even if the protocol is only sending one bit, since it can depend arbitrarily on Alice/Bob's input.

Combining Conjecture I.8 and the PRG construction from [17, Theorem 4.5], we would obtain a completely different explicit PRG for this class with error ε and seed length $n + d\cdot \mathrm{polylog}(n/\varepsilon)$.

a) Paper Organization: An overview of our proofs is given in Section II. The full version of our paper can be found at https://arxiv.org/abs/2307.13926.

II. Proof Overview

We first briefly outline the proof strategy, which consists of three main components:

- First, we show that the level-one bound can be characterized as the expected absolute value of a martingale defined as follows: Consider the random walk induced on the protocol tree when Alice and Bob are given inputs $\boldsymbol{x}$ and $\boldsymbol{y}$ uniformly from $\{\pm1\}^n$. Let $\boldsymbol{X}^{(t)} \times \boldsymbol{Y}^{(t)}$ be the rectangle associated with the random walk at time t. The martingale process tracks the inner product $\left\langle \mu(\boldsymbol{X}^{(t)}), \mu(\boldsymbol{Y}^{(t)}) \right\rangle$ where $\mu(\boldsymbol{X}^{(t)}) = \mathbb{E}\left[\boldsymbol{x} \,\middle|\, \boldsymbol{x} \in \boldsymbol{X}^{(t)}\right]$ and $\mu(\boldsymbol{Y}^{(t)}) = \mathbb{E}\left[\boldsymbol{y} \,\middle|\, \boldsymbol{y} \in \boldsymbol{Y}^{(t)}\right]$ are Alice's and Bob's center of masses.
- Second, to bound the value of the martingale, it is necessary to ensure that neither $\boldsymbol{X}^{(t)}$ nor $\boldsymbol{Y}^{(t)}$ become excessively elongated in any direction during the protocol execution. To measure the length of $\boldsymbol{X}^{(t)}$ in a particular direction $\theta \in \mathbb{S}^{n-1}$, we calculate the variance $\mathbb{V}\mathrm{ar}\left[\langle \boldsymbol{x}, \theta\rangle \,\middle|\, \boldsymbol{x} \in \boldsymbol{X}^{(t)}\right]$, i.e. the variance of a uniformly random $\boldsymbol{x} \in \boldsymbol{X}^{(t)}$ in the direction θ. If the set is not elongated in any direction, this can be thought of as a spectral notion of almost pairwise independence. Such a notion also generalizes to almost k-wise independence by considering higher moments.

  To achieve the property that the sets are not elongated, one of the main novel ideas in our paper is to modify the original protocol to a new one that incorporates additional cleanup steps where the parties communicate *real values* $\langle \boldsymbol{x}, \theta\rangle$. Through these communication steps, the sets $\boldsymbol{X}^{(t)}$ and $\boldsymbol{Y}^{(t)}$ are recursively divided into affine slices along problematic directions.
- Last, one needs to show that the number of cleanup steps are small in order to bound the value of the martingale for the new protocol. This is the most involved part of our proof and requires considerable effort because the cleanup steps are real-valued and adaptively depend on the entire history, including the previous real values communicated.

The strategy outlined above also generalizes to level-two Fourier growth by considering higher moments and sending values of quadratic forms in the inputs. We also remark that since we view the sets $\boldsymbol{X}^{(t)}$ and $\boldsymbol{Y}^{(t)}$ above as embedded in $\mathbb{R}^n$ and allow the protocol to send real values, it is more natural for us to work in Gaussian space by doing a standard transformation. The rotational invariance of the Gaussian space also seems to be essential for us to obtain optimal level-one bound without losing additional polylogarithmic factors.

We now elaborate on the above components in detail and also highlight the differences between the level-one and level-two settings. For conciseness, in the following overview we use $f \lesssim g$ to denote $f = O(g)$ and $f \gtrsim g$ to denote $f = \Omega(g)$ where O and Ω only hide absolute constants.

A. Level-One Fourier Growth

The level-one Fourier growth of the XOR-fiber h is given by

$$\begin{aligned} L_{1,1}(h) &= \sum_{i=1}^n \left|\widehat{h}(\{i\})\right| = \sum_{i=1}^n \left| \mathop{\mathbb{E}}_{\boldsymbol{z}\sim\nu}[h(\boldsymbol{z})\boldsymbol{z}_i] \right| \\ &= \sum_{i=1}^n \left| \mathop{\mathbb{E}}_{\boldsymbol{x},\boldsymbol{y}\sim\nu}[\mathcal{C}(\boldsymbol{x},\boldsymbol{y})\boldsymbol{x}_i\boldsymbol{y}_i] \right|. \end{aligned}$$

To bound the above, it suffices to bound $\sum_{i=1}^n \eta_i \cdot \mathbb{E}[\mathcal{C}(\boldsymbol{x},\boldsymbol{y})\boldsymbol{x}_i\boldsymbol{y}_i]$ for any sign vector $\eta \in \{\pm1\}^n$. Here for simplicity we assume $\eta_i \equiv 1$ and the probability of reaching every leaf is $\approx 2^{-d}$.

a) A Martingale Perspective: To evaluate the quantity $\sum_{i=1}^n \mathbb{E}[\mathcal{C}(\boldsymbol{x},\boldsymbol{y})\boldsymbol{x}_i\boldsymbol{y}_i]$, consider a random leaf $\boldsymbol{\ell}$ of the protocol and let $\boldsymbol{X_\ell} \times \boldsymbol{Y_\ell}$ be the corresponding rectangle. Since the leaf determines the answer of the protocol, denoted by $\mathcal{C}(\boldsymbol{\ell})$, the quantity above equals

$$\begin{aligned} &\sum_{i=1}^n \mathop{\mathbb{E}}_{\boldsymbol{\ell}}\left[\mathcal{C}(\boldsymbol{\ell}) \cdot \mathbb{E}[\boldsymbol{x}_i \,|\, \boldsymbol{x} \in \boldsymbol{X_\ell}] \cdot \mathbb{E}[\boldsymbol{y}_i \,|\, \boldsymbol{y} \in \boldsymbol{Y_\ell}]\right] \\ &= \mathop{\mathbb{E}}_{\boldsymbol{\ell}}[\mathcal{C}(\boldsymbol{\ell}) \cdot \langle \mu(\boldsymbol{X_\ell}), \mu(\boldsymbol{Y_\ell})\rangle] \\ &\le \mathop{\mathbb{E}}_{\boldsymbol{\ell}}[| \langle \mu(\boldsymbol{X_\ell}), \mu(\boldsymbol{Y_\ell})\rangle |], \end{aligned}$$

where we define $\mu(\boldsymbol{X_\ell}) = \mathbb{E}\left[\boldsymbol{x} \,|\, \boldsymbol{x} \in \boldsymbol{X_\ell}\right]$ and $\mu(\boldsymbol{Y_\ell}) = \mathbb{E}\left[\boldsymbol{y} \,|\, \boldsymbol{y} \in \boldsymbol{Y_\ell}\right]$ to be the center of masses of the rectangle. Our goal is to bound the magnitude of the random variable $\boldsymbol{z} = \langle \mu(\boldsymbol{X_\ell}), \mu(\boldsymbol{Y_\ell})\rangle$.

We shall show that $\mathbb{E}_{\boldsymbol{\ell}}[|\boldsymbol{z}|] \lesssim \sqrt{d}$. Note that $|\boldsymbol{z}|$ can be as large as d in the worst case — for instance if the first d coordinates of $\boldsymbol{X_\ell}$ and $\boldsymbol{Y_\ell}$ are fixed to the same value — thus we cannot argue for each leaf separately.

To analyze it for a random leaf, we first characterize the above as a martingale process using the tree structure of the protocol. The martingale process is defined as $\left(\boldsymbol{z}^{(t)}\right)_t$ where $\boldsymbol{z}^{(t)} := \left\langle \mu(\boldsymbol{X}^{(t)}), \mu(\boldsymbol{Y}^{(t)}) \right\rangle$ tracks the inner product between the center of masses $\mu(\boldsymbol{X}^{(t)})$ and $\mu(\boldsymbol{Y}^{(t)})$ of the current rectangle $\boldsymbol{X}^{(t)} \times \boldsymbol{Y}^{(t)}$ at step t. Denote the martingale differences by $\Delta \boldsymbol{z}^{(t+1)} = \boldsymbol{z}^{(t+1)} - \boldsymbol{z}^{(t)}$ and note that if in the t^{th} step Alice sends a message, then

$$\Delta \boldsymbol{z}^{(t+1)} = \left\langle \Delta\mu(\boldsymbol{X}^{(t+1)}), \mu(\boldsymbol{Y}^{(t+1)}) \right\rangle,$$

where $\Delta\mu(\boldsymbol{X}^{(t+1)}) = \mu(\boldsymbol{X}^{(t+1)}) - \mu(\boldsymbol{X}^{(t)})$ is the change in Alice's center of mass. A similar expression holds if Bob sends a message. Then it suffices to bound the expected quadratic variation since

$$\begin{aligned}\left(\mathbb{E}\left[\left|\boldsymbol{z}^{(d)}\right|\right]\right)^2 &\le \mathbb{E}\left[\left(\boldsymbol{z}^{(d)}\right)^2\right] \\ &= \mathbb{E}\left[\sum_{t=0}^{d-1}\left(\Delta \boldsymbol{z}^{(t+1)}\right)^2\right], \qquad \text{(II.1)}\end{aligned}$$

where the equality holds due to the martingale property: $\mathbb{E}\left[\Delta \boldsymbol{z}^{(t+1)} \,\middle|\, \boldsymbol{z}^{(1)}, \dots \boldsymbol{z}^{(t)}\right] = 0$.

To obtain the desired bound, we need to bound the expected quadratic variation by $O(d)$. Note that it could be the case that a single $\Delta \boldsymbol{z}^{(t+1)}$ scales like $\sqrt{d}$. For instance, if Bob first announces his first d coordinates, $y_1, \dots, y_d$, and then Alice sends a majority of $x_1 \cdot y_1, \dots, x_d \cdot y_d$, then in the last step Alice's center of mass $\mu(\boldsymbol{X}^{(t+1)})$ changes by $\approx 1/\sqrt{d}$ in each of the first d coordinates, and the inner product with Bob's center of mass changes by $\approx \sqrt{d}$ in a single step.

Such cases make it difficult to directly control the individual step sizes of the martingale and we will only be able to obtain an amortized bound. It turns out, as we explain later, that such an amortized bound on the martingale can be obtained if Alice and Bob's sets are not elongated in any direction. Therefore, we will transform the original protocol into a *clean* protocol by introducing real communication steps that slice the elongated directions. For this, it will be convenient to work in Gaussian space which also turns out to be essential in proving the optimal $O(\sqrt{d})$ bound.

b) Protocols in Gaussian Space: A communication protocol in Gaussian space takes as inputs $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ where $\boldsymbol{x}, \boldsymbol{y}$ are independently sampled from the Gaussian distribution γ_n. One can embed the original Boolean protocol in the Gaussian space by running the protocol on the uniformly distributed Boolean inputs $\operatorname{sgn}(\boldsymbol{x})$ and $\operatorname{sgn}(\boldsymbol{y})$ where $\operatorname{sgn}(\cdot)$ takes the sign of each coordinate. Note that any node of the protocol tree in the Gaussian space corresponds to a rectangle $X \times Y$ where $X, Y \subseteq \mathbb{R}^n$. Abusing the notation and defining their *Gaussian* centers of masses as $\mu(X) = \mathbb{E}_{\boldsymbol{x}\sim\gamma_n}[\boldsymbol{x} \mid \boldsymbol{x} \in X]$ and $\mu(Y) = \mathbb{E}_{\boldsymbol{y}\sim\gamma_n}[\boldsymbol{y} \mid \boldsymbol{y} \in Y]$, one can associate the same martingale $(\boldsymbol{z}^{(t)})_t$ with the protocol in the Gaussian space:

$$\boldsymbol{z}^{(t)} = \left\langle \mu(\boldsymbol{X}^{(t)}), \mu(\boldsymbol{Y}^{(t)}) \right\rangle.$$

It turns out that bounding the quadratic variation of this martingale suffices to give a bound on $L_{1,2}(h)$, so we will stick to the Gaussian setting. We now describe the ideas behind the cleanup process so that the step sizes can be controlled more easily.

c) Cleanup with Real Communication: The cleanup protocol runs the original protocol interspersed with some cleanup steps where Alice and Bob send real values. As outlined before, one of the goals of these cleanup steps is to ensure that the sets are not elongated in any direction, in order to control the martingale steps. In more detail, recall that we want to control

$$\begin{aligned}&\mathbb{E}\left[(\Delta \boldsymbol{z}^{(t+1)})^2 \,\middle|\, \boldsymbol{z}^{(1)}, \dots, \boldsymbol{z}^{(t)}\right] \\ &= \mathbb{E}\left[\left\langle \Delta\mu(\boldsymbol{X}^{(t+1)}), \mu(\boldsymbol{Y}^{(t+1)}) \right\rangle^2 \,\middle|\, \boldsymbol{z}^{(1)}, \dots, \boldsymbol{z}^{(t)}\right]\end{aligned}$$

in the t^{th} step where Alice speaks. There are two key underlying ideas for the cleanup steps:

- **Gram-Schmidt Orthogonalization:**
At each round, if the current rectangle is $\boldsymbol{X} \times \boldsymbol{Y}$, before Alice sends the actual message, she sends the inner product $\langle x, \mu(\boldsymbol{Y})\rangle$ between her input and Bob's current center of mass $\mu(\boldsymbol{Y})$. This partitions Alice's set $\boldsymbol{X}$ into affine slices orthogonal to Bob's current center of mass $\mu(\boldsymbol{Y})$. Thus the change in Alice's center of mass in later rounds is orthogonal to $\mu(\boldsymbol{Y})$ since it only takes place inside the affine slice.
Recall that the martingale $\boldsymbol{z}^{(t)}$ is the inner product of Alice and Bob's center of masses, and Bob's center of mass does not change when Alice speaks. The original communication steps now do not contribute to the martingale and only the steps where the inner products are revealed do. In particular, if $t_{\text{prev}} < t$ are two consecutive times where Alice revealed the inner product, then the change in Alice's center of mass is orthogonal to change in Bob's center of mass between time t_{prev} and t. Thus, conditioned on the rectangle $\boldsymbol{X}^{(t)} \times \boldsymbol{Y}^{(t)}$ fixed by the messages until time t, we have, by Jensen's inequality,
$$\begin{aligned}&\mathbb{E}\left[(\Delta \boldsymbol{z}^{(t+1)})^2\right] \\ &= \mathbb{E}\left[\left\langle \Delta\mu(\boldsymbol{X}^{(t+1)}), \mu(\boldsymbol{Y}^{(t)}) - \mu(\boldsymbol{Y}^{(t_{\text{prev}})}) \right\rangle^2\right] \\ &\le \mathbb{E}\left[\left\langle \boldsymbol{x} - \mu(\boldsymbol{X}^{(t)}), \mu(\boldsymbol{Y}^{(t)}) - \mu(\boldsymbol{Y}^{(t_{\text{prev}})}) \right\rangle^2\right].\end{aligned}$$
Note that the quantity on the right-hand side above is of the form $\langle \boldsymbol{x} - \mathbb{E}[\boldsymbol{x}], v\rangle$. In other words, it is the variance of the random vector $\boldsymbol{x}$ along direction v. To maintain a bound on this quantity, we introduce the notion of "not being elongated in any direction".
- **Not elongated in any direction:** We define the following notion to capture the fact that the random vector is not elongated in any direction: we say that a mean-zero random vector $\boldsymbol{x}' = \boldsymbol{x} - \mathbb{E}[\boldsymbol{x}]$ in $\mathbb{R}^n$ is λ*-pairwise clean*, if for every $v \in \mathbb{R}^n$,
$$\mathbb{E}\left[\langle \boldsymbol{x}', v\rangle^2\right] \le \lambda \cdot \|v\|^2, \qquad \text{(II.2)}$$
or equivalently, the operator norm of the covariance matrix $\mathbb{E}[\boldsymbol{x}'\boldsymbol{x}'^\top]$ is at most λ. This can be considered a

spectral notion of almost pairwise independence, since the pairwise moments are well-behaved in every direction.

If the input distribution conditioned on Alice's set $\boldsymbol{X}^{(t)}$ is $O(1)$-pairwise clean, we say that her set is *pairwise clean*. Based on the above ideas, after Alice sends the initial message, if her set is not yet clean, she partitions it recursively by taking affine slices and transmitting real values. More precisely, while there is direction $\theta \in \mathbb{S}^{n-1}$ violating (II.2), Alice does a cleanup of her set by sending the inner product $\langle x, \theta \rangle$. This direction is known to Bob as it only depends on Alice's current space. In addition, this cleanup does not contribute to the martingale *in the future* because the inner product along this direction is fixed now.

The resulting protocol is pairwise clean in the sense that at each step[6], Alice's current set is pairwise clean. Similar arguments work for Bob.

Let $\boldsymbol{d}$ be the total number of communication rounds including all the cleanup steps. Then, by the above argument, and denoting by $(\boldsymbol{\tau}_m)_m$ and $(\boldsymbol{\tau}'_m)_m$ the indices of the inner product steps for Alice and Bob, we can ultimately bound

$$\begin{aligned}\mathbb{E}\left[(\boldsymbol{z}^{(\boldsymbol{d})})^2\right] \lesssim \mathbb{E}\Bigg[\sum_m \left\|\mu(\boldsymbol{X}^{(\boldsymbol{\tau}_m)}) - \mu(\boldsymbol{X}^{(\boldsymbol{\tau}_{m-1})})\right\|^2 \\ + \left\|\mu(\boldsymbol{Y}^{(\boldsymbol{\tau}'_m)}) - \mu(\boldsymbol{Y}^{(\boldsymbol{\tau}'_m - 1)})\right\|^2\Bigg] \\ = \mathbb{E}\left[\left\|\mu(\boldsymbol{X}^{(\boldsymbol{d})})\right\|^2 + \left\|\mu(\boldsymbol{Y}^{(\boldsymbol{d})})\right\|^2\right], \qquad \text{(II.3)}\end{aligned}$$

where again, the last equality follows from the martingale property. The right hand side above can be bounded by the expected number of communication rounds $\mathbb{E}[\boldsymbol{d}]$ using the level-one inequality — this inequality bounds the Euclidean norm of the center of mass of a set in terms of its Gaussian measure.

d) Expected Number of Cleanup steps: Since the original communication only consists of d rounds, the analysis essentially reduces to bounding the expected number of cleanup steps by $O(d)$, which is technically the most involved part of the proof.

It is implicit in the previous works on the Gap-Hamming Problem [35], [37] that large sets are not elongated in many directions: if a set $X \subseteq \mathbb{R}^n$ has Gaussian measure $\approx 2^{-d}$, then for a random vector $\boldsymbol{x}$ sampled from X, there are at most $m \lesssim d$ orthogonal directions $\theta_1, \ldots, \theta_m$ such that $\mathbb{E}[\langle \boldsymbol{x}', \theta_i \rangle^2] \gtrsim 1$ where $\boldsymbol{x}' = \boldsymbol{x} - \mathbb{E}[\boldsymbol{x}]$. This is a consequence of the fact that the expectation of $\boldsymbol{q} = \sum_{i=1}^{m} \langle \boldsymbol{x}', \theta_i \rangle^2$ can be bounded by $O(d)$ provided that X has measure $\approx 2^{-d}$.

The above argument suggests that maybe we can clean up the set X along these $O(d)$ bad orthogonal directions. However this is not enough for our purposes: after taking an affine slice, the set may not be clean in a direction where it was clean before. Moreover, since the parties take turns to send messages and clean up, the bad directions will also depend on the entire history of the protocol, including the previous real and Boolean communication. This adaptivity makes the analysis more delicate and to prove the optimal bound we crucially utilize the rotational symmetry of the Gaussian distribution. Indeed, the fact that a large set is not elongated in many directions also holds even when we replace the Gaussian distribution with the uniform distribution on $\{\pm 1\}^n$, but it is unclear how to obtain an optimal level-one bound using the latter.

In the final protocol, since the parties only send Boolean bits and linear forms of their inputs, conditioned on the history of the martingale, one can still say what the distribution of the next cleanup $\langle \boldsymbol{x}, \theta \rangle$ looks like, as the Gaussian distribution is well-behaved under linear projections. We then use martingale concentration and stopping time arguments to show that the expected number of cleanup steps is indeed bounded by $O(d)$ even if the cleanup is adaptive.

We make two remarks in passing: First, we can also prove the optimal level-one bound using information-theoretic ideas but they do not seem to generalize to the level-two setting, so we adopt the alternative concentration-based approach here and they are similar in spirit. Second, it is possible from our proof approach (in particular, the approach for level two described next) to derive a weaker upper bound of $\sqrt{d} \cdot \mathrm{polylog}(n)$ for the level one while directly working with the uniform distribution on the hypercube.

B. Level-Two Fourier Growth

We start by noting that the level-two Fourier growth of the XOR-fiber h is given by

$$\begin{aligned}L_{1,2}(h) = \sum_{i \neq j} \left|\widehat{h}(\{i,j\})\right| = \sum_{i \neq j} \left|\mathop{\mathbb{E}}_{\boldsymbol{z} \sim \nu}[h(\boldsymbol{z})\boldsymbol{z}_i \boldsymbol{z}_j]\right| \\ = \sum_{i \neq j} \left|\mathop{\mathbb{E}}_{\boldsymbol{x},\boldsymbol{y} \sim \nu}[\mathcal{C}(\boldsymbol{x},\boldsymbol{y})\boldsymbol{x}_i \boldsymbol{x}_j \boldsymbol{y}_i \boldsymbol{y}_j]\right|.\end{aligned}$$

To bound the above, it suffices to bound $\sum_{i \neq j} \eta_{ij} \cdot \mathbb{E}[\mathcal{C}(\boldsymbol{x},\boldsymbol{y})\boldsymbol{x}_i \boldsymbol{x}_j \boldsymbol{y}_i \boldsymbol{y}_j]$ for any symmetric sign matrix (η_{ij}). For this proof overview, we assume for simplicity that $\eta_{ij} \equiv 1$.

a) Martingales and Gram-Schmidt Orthogonalization: Similar to the case of level one, the level-two Fourier growth also has a martingale formulation. In particular, let $\boldsymbol{X}^{(t)}$ and $\boldsymbol{Y}^{(t)}$ be Alice and Bob's sets at time t as before and define $\sigma(\boldsymbol{X}^{(t)}) = \mathbb{E}\left[\boldsymbol{x} \dot{\otimes} \boldsymbol{x} \,\middle|\, \boldsymbol{x} \in \boldsymbol{X}^{(t)}\right], \sigma(\boldsymbol{Y}^{(t)}) = \mathbb{E}\left[\boldsymbol{y} \dot{\otimes} \boldsymbol{y} \,\middle|\, \boldsymbol{y} \in \boldsymbol{Y}^{(t)}\right]$ to be the $n \times n$ matrices that represent the *level-two center of masses* of the two sets. Here $\boldsymbol{x} \dot{\otimes} \boldsymbol{y}$ denotes the tensor product $\boldsymbol{x} \otimes \boldsymbol{y}$ with the diagonal zeroed out.[7] To bound the level-two Fourier growth, it suffices to bound the expected quadratic variation of the martingale $\left(\boldsymbol{z}^{(t)}\right)_t$ defined by taking the inner product of the level-two center of masses $\boldsymbol{z}^{(t)} := \left\langle \sigma(\boldsymbol{X}^{(t)}), \sigma(\boldsymbol{Y}^{(t)}) \right\rangle$ where $\langle \cdot, \cdot \rangle$ is the inner product of two matrices viewed as vectors.

[6] We remark that the sets are only clean at intermediate steps where a cleanup phase ends, but we show that because of the orthogonalization step, the other steps do not contribute to the value of the martingale.

[7] Here $x \dot{\otimes} y$ is an $n \times n$ matrix. We will also interchangeably view $n \times n$ matrices as n^2-length vectors.

To this end, we again move to Gaussian space where the inputs $x, y \in \mathbb{R}^n$ and transform the protocol to a clean protocol. First, we need an analog of the *Gram-Schmidt orthogonalization* step — this is achieved in a natural way by Alice sending inner product $\left\langle x \dot{\otimes} x, \sigma(\boldsymbol{Y}^{(t)}) \right\rangle$ with Bob's level-two center of mass, and Bob does the same. Note that Alice and Bob are now exchanging values of quadratic polynomials in their inputs. Thus, to control the step sizes, we now need to control the second moment of quadratic forms which naturally motivates the following spectral analogue of 4-wise independence.

b) 4-wise Cleanup with Quadratic Forms: We say a random vector $\boldsymbol{x}$ is 4-wise clean with parameter λ if the operator norm of the $n^2 \times n^2$ covariance matrix

$$\mathbb{E}\left[\left(\boldsymbol{x} \dot{\otimes} \boldsymbol{x} - \mathbb{E}\left[\boldsymbol{x} \dot{\otimes} \boldsymbol{x}\right]\right)\left(\boldsymbol{x} \dot{\otimes} \boldsymbol{x} - \mathbb{E}\left[\boldsymbol{x} \dot{\otimes} \boldsymbol{x}\right]\right)^{\top}\right]$$

is at most λ where we view $\boldsymbol{x} \dot{\otimes} \boldsymbol{x} - \mathbb{E}[\boldsymbol{x} \dot{\otimes} \boldsymbol{x}]$ as an n^2-dimensional vector. This is equivalent to saying that for any quadratic form $\left\langle M, \boldsymbol{x} \dot{\otimes} \boldsymbol{x} \right\rangle$,

$$\mathbb{E}\left[\left\langle M, \boldsymbol{x} \dot{\otimes} \boldsymbol{x} - \mathbb{E}\left[\boldsymbol{x} \dot{\otimes} \boldsymbol{x}\right]\right\rangle^2\right] \leq \lambda \left\|M\right\|^2, \tag{II.4}$$

where $\|M\|$ denotes the Euclidean norm of M when viewed as a vector. Thus, this allows us to control the second moment of any quadratic polynomial (and in particular, fourth moments of linear functions). We note that one can generalize the above spectral notion to k-wise independence in the natural way by looking at the covariance matrix of the tensor $\boldsymbol{x}^{\dot{\otimes} k}$.

We say a set is 4*-wise clean* with parameter λ if (II.4) is preserved for all M with zero diagonal[8]. Combined with this notion, one can define the cleanup in an analogous way to the level-one cleanup: While there exists some $M \in \mathbb{R}^{n \times n}$ violating (II.4), Alice sends the quadratic form $\left\langle x \dot{\otimes} x, M \right\rangle$ to Bob until her set is 4-wise clean with parameter λ.

c) Cleanup Analysis via Hanson-Wright Inequalities: The crux of the proof is to bound the number of cleanup steps which, together with a similar analysis as in the level-one case, gives us the desired bound. We show that $m \lesssim d$ cleanup steps suffice in expectation to make the sets 4-wise clean for $\lambda \leq d \cdot \mathrm{polylog}(n)$. Analogous to (II.1) and (II.3), this gives a bound of $d^3 \cdot \mathrm{polylog}(n)$ on the expected quadratic variation and implies $L_{1,2}(h) \leq d^{3/2} \cdot \mathrm{polylog}(n)$.

Since the parties send values of quadratic forms now, the analysis here is significantly more involved compared to the level-one case, even after moving to the Gaussian setting, where one could previously use the fact that the Gaussian distribution behaves nicely under linear projections. We rely on a powerful generalization of the Hanson-Wright inequality to a Banach-space-valued setting due to Adamczak, Latała, and Meller [63]. This inequality gives a tail bound for sum of squares of quadratic forms: In particular if $M_1, \ldots, M_m$ are matrices with zero diagonal which form an orthonormal set when viewed as n^2 dimensional vectors, then the random variable $\boldsymbol{q} = \sum_{i=1}^m \left\langle \boldsymbol{x} \dot{\otimes} \boldsymbol{x}, M_i \right\rangle^2$ satisfies $\mathbf{Pr}_{\boldsymbol{x} \sim \gamma_n}[\boldsymbol{q} \geq t] \leq e^{-\Omega(\sqrt{t})}$ for any $t \gtrsim m^2$. We remark that this tail bound relies on the orthogonality of the quadratic forms and is much sharper than, for example, the bound obtained from hypercontractivity or other standard polynomial concentration inequalities.

In our setting, the matrices are being chosen adaptively. In addition, the parties are sending quadratic forms in their inputs, and the distribution of the next $\left\langle \boldsymbol{x} \dot{\otimes} \boldsymbol{x}, M \right\rangle$ conditioned on the history is hard to determine, unlike the level-one case. To handle this, we replace the real communication with Boolean communication of finite precision $\pm 1/\mathrm{poly}(n)$. This means that whenever Alice wants to perform cleanup $\langle \boldsymbol{x} \otimes \boldsymbol{x}, M \rangle$ for some M known to both parties, she sends only $O(\log(n))$ bits. On the one hand, this modification is similar enough to the cleanup protocol with real messages so that most of the argument carries through. On the other hand, now the protocol is completely discrete, which allows us to condition on any particular transcript.

For intuition, assume we fix a transcript of $L = d + O(m \log(n))$ bits which has gone through m cleanups. Typically, this transcript should capture $\approx 2^{-L}$ of the probability mass. More crucially, the matrices $M_1, \ldots, M_m$ for the cleanups are also fixed along the transcript, and one can apply the aforementioned Hanson-Wright inequality on $\boldsymbol{q} = \sum_{i=1}^m \left\langle \boldsymbol{x} \dot{\otimes} \boldsymbol{x}, M_i \right\rangle^2$. Combining the two facts together, we can apply the non-adaptive tail bound above and then condition on obtaining such typical transcript. This shows $\mathbb{E}[\boldsymbol{q}] \leq d^2 \cdot \mathrm{polylog}(n)$. However, each quadratic form comes from a violation of (II.4) and contributes at least λ to $\boldsymbol{q}$ in expectation. This implies that $\mathbb{E}[\boldsymbol{q}] \geq \lambda \cdot m$ and by taking $\lambda = d \cdot \mathrm{polylog}(n)$, we derive that the number of cleanup steps $m \lesssim d$. This shows that the level-two Fourier growth is $O((m + d) \cdot \sqrt{\lambda}) = d^{3/2} \cdot \mathrm{polylog}(n)$ completing the proof.

Note that if we could take $\lambda = \mathrm{polylog}(n)$ while having the same number of cleanup steps $m = d \cdot \mathrm{polylog}(n)$, then we would obtain an optimal level-two bound of $d \cdot \mathrm{polylog}(n)$. However, it is not clear how to use current approach to show this. In the full version, we identify examples showing the tightness of our current analysis and also discuss potential ways to circumvent the obstacles within.

We remark that by replacing the Hanson-Wright inequality with its higher-degree variants and performing level-k cleanups, we can analyze level-k Fourier growth in the similar way. However, since the first two levels already suffice for our applications and we believe that our level-two bound can be further improved, we do not make the effort of generalizing it to higher levels here.

Acknowledgement

Part of this work was done while the authors were visiting the Simons Institute for the Theory of Computing.

[8] The requirement of zero diagonal is for analysis purposes only and can be assumed without loss of generality since $\boldsymbol{x} \dot{\otimes} \boldsymbol{x}$ is zero diagonal anyway.

References

[1] Y. Mansour, "An $O(n^{\log \log n})$ learning algorithm for DNF under the uniform distribution," *J. Comput. Syst. Sci.*, vol. 50, no. 3, pp. 543–550, 1995, appeared in COLT, 1992. 1, 2

[2] R. Agrawal, "Coin theorems and the fourier expansion," *Chic. J. Theor. Comput. Sci.*, vol. 2020, 2020. [Online]. Available: http://cjtcs.cs.uchicago.edu/articles/2020/4/contents.html 1, 3

[3] E. Chattopadhyay, P. Hatami, S. Lovett, and A. Tal, "Pseudorandom generators from the second fourier level and applications to ac0 with parity gates," in *10th Innovations in Theoretical Computer Science Conference (ITCS 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018. 1, 2

[4] R. Raz and A. Tal, "Oracle separation of BQP and PH," in *STOC*. ACM, 2019, pp. 13–23, presented in QIP, 2019 as a plenary talk. Accepted to the Journal of the ACM. 1

[5] X. Wu, "A stochastic calculus approach to the oracle separation of BQP and PH," *Theory Comput.*, vol. 18, pp. 1–11, 2022. [Online]. Available: https://theoryofcomputing.org/articles/v018a017/ 1

[6] U. Girish, R. Raz, and A. Tal, "Quantum versus randomized communication complexity, with efficient players," in *ITCS*, ser. LIPIcs, vol. 185, 2021, pp. 54:1–54:20, presented in QIP, 2020 as a contributed talk. 1, 2, 3, 4, 5

[7] A. Tal, "Tight bounds on the Fourier spectrum of AC0," in *Computational Complexity Conference*, ser. LIPIcs, vol. 79. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2017, pp. 15:1–15:31. 2

[8] P. Gopalan, R. A. Servedio, A. Tal, and A. Wigderson, "Degree and sensitivity: tails of two distributions," *CoRR*, vol. abs/1604.07432, 2016. [Online]. Available: http://arxiv.org/abs/1604.07432 2

[9] O. Reingold, T. Steinke, and S. P. Vadhan, "Pseudorandomness for regular branching programs via Fourier analysis," in *APPROX-RANDOM*. Springer, 2013, pp. 655–670. 2

[10] T. Steinke, S. P. Vadhan, and A. Wan, "Pseudorandomness and Fourier-growth bounds for width-3 branching programs," *Theory of Computing*, vol. 13, no. 1, pp. 1–50, 2017, appeared in APPROX-RANDOM, 2014. 2

[11] E. Chattopadhyay, P. Hatami, O. Reingold, and A. Tal, "Improved pseudorandomness for unordered branching programs through local monotonicity," in *STOC*. ACM, 2018, pp. 363–375. 2

[12] C. H. Lee, E. Pyne, and S. P. Vadhan, "Fourier growth of regular branching programs," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms DBLP:conf/approx/LeePV22and Techniques, APPROX/RANDOM 2022, September 19-21, 2022, University of Illinois, Urbana-Champaign, USA (Virtual Conference)*, ser. LIPIcs, A. Chakrabarti and C. Swamy, Eds., vol. 245. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022, pp. 2:1–2:21. 2

[13] R. O'Donnell and R. A. Servedio, "Learning monotone decision trees in polynomial time," *SIAM Journal on Computing*, vol. 37, no. 3, pp. 827–844, 2007. 2

[14] A. Tal, "Towards optimal separations between quantum and randomized query complexities," in *FOCS*. IEEE, 2020, pp. 228–239. 2, 5

[15] A. A. Sherstov, A. A. Storozhenko, and P. Wu, "An optimal separation of randomized and quantum query complexity," in *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, 2021, pp. 1289–1302. 2, 5

[16] U. Girish, R. Raz, and W. Zhan, "Lower bounds for xor of forrelations," *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, 2021. 2

[17] E. Chattopadhyay, P. Hatami, K. Hosseini, and S. Lovett, "Pseudorandom generators from polarizing random walks," *Theory Comput.*, vol. 15, pp. 1–26, 2019. 2, 6

[18] J. Błasiok, P. Ivanov, Y. Jin, C. H. Lee, R. A. Servedio, and E. Viola, "Fourier growth of structured $\mathbb{F}_2$-polynomials and applications," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2021)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2021. 2

[19] C. H. Lee, "Fourier bounds and pseudorandom generators for product tests," in *34th Computational Complexity Conference, CCC 2019, July 18-20, 2019, New Brunswick, NJ, USA*, ser. LIPIcs, A. Shpilka, Ed., vol. 137. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019, pp. 7:1–7:25. [Online]. Available: https://doi.org/10.4230/LIPIcs.CCC.2019.7 2

[20] E. Blais, L. Tan, and A. Wan, "An inequality for the fourier spectrum of parity decision trees," *CoRR*, vol. abs/1506.01055, 2015. [Online]. Available: http://arxiv.org/abs/1506.01055 2

[21] U. Girish, A. Tal, and K. Wu, "Fourier growth of parity decision trees," in *36th Computational Complexity Conference (CCC 2021)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2021. 2, 3, 5

[22] S. Iyer, A. Rao, V. Reis, T. Rothvoss, and A. Yehudayoff, "Tight bounds on the fourier growth of bounded functions on the hypercube," *arXiv preprint arXiv:2107.06309*, 2021. 2

[23] A. A. Razborov, "Lower bounds on the size of bounded depth circuits over a complete basis with logical addition," *Mathematical Notes of the Academy of Sciences of the USSR*, vol. 41, no. 4, pp. 333–338, 1987. 2

[24] R. Smolensky, "Algebraic methods in the theory of lower bounds for boolean circuit complexity," in *Proceedings of the 19th Annual ACM Symposium on Theory of Computing, 1987, New York, New York, USA*, A. V. Aho, Ed. ACM, 1987, pp. 77–82. [Online]. Available: https://doi.org/10.1145/28395.28404 2

[25] A. Montanaro and T. Osborne, “On the communication complexity of xor functions,” 2010. 2

[26] H. Hatami, K. Hosseini, and S. Lovett, “Structure of protocols for XOR functions,” *SIAM J. Comput.*, vol. 47, no. 1, pp. 208–217, 2018. 2, 4, 5

[27] H. Y. Tsang, C. H. Wong, N. Xie, and S. Zhang, “Fourier sparsity, spectral norm, and the log-rank conjecture,” in *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, 2013, pp. 658–667. 2

[28] Y. Shi and Z. Zhang, “Communication complexities of xor functions,” *arXiv preprint arXiv:0808.1762*, 2008. 2

[29] S. Zhang, “Efficient quantum protocols for xor functions,” in *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*. SIAM, 2014, pp. 1878–1885. 2

[30] R. Raz, “Fourier analysis for probabilistic communication complexity,” *Comput. Complex.*, vol. 5, no. 3/4, pp. 205–221, 1995. [Online]. Available: https://doi.org/10.1007/BF01206318 2

[31] J. Brody and E. Verbin, “The coin problem and pseudorandomness for branching programs,” in *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, 2010, pp. 30–39. 3

[32] G. Cohen, A. Ganor, and R. Raz, “Two sides of the coin problem,” in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2014)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2014. 3

[33] N. Limaye, K. Sreenivasaiah, S. Srinivasan, U. Tripathi, and S. Venkitesh, “A fixed-depth size-hierarchy theorem for $AC^0[\oplus]$ via the coin problem,” in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, 2019, pp. 442–453. 3

[34] C. H. Lee and E. Viola, “The coin problem for product tests,” *ACM Transactions on Computation Theory (TOCT)*, vol. 10, no. 3, pp. 1–10, 2018. 3

[35] A. Chakrabarti and O. Regev, “An optimal lower bound on the communication complexity of gap-hamming-distance,” *SIAM J. Comput.*, vol. 41, no. 5, pp. 1299–1317, 2012. [Online]. Available: https://doi.org/10.1137/120861072 4, 8

[36] A. A. Sherstov, “The communication complexity of gap hamming distance,” *Theory Comput.*, vol. 8, no. 1, pp. 197–208, 2012. [Online]. Available: https://doi.org/10.4086/toc.2012.v008a008 4

[37] T. Vidick, “A concentration inequality for the overlap of a vector on a large set, with application to the communication complexity of the gap-hamming-distance problem,” *Chic. J. Theor. Comput. Sci.*, vol. 2012, 2012. [Online]. Available: http://cjtcs.cs.uchicago.edu/articles/2012/1/contents.html 4, 8

[38] A. Rao and A. Yehudayoff, “Anticoncentration and the exact gap-hamming problem,” *SIAM Journal on Discrete Mathematics*, vol. 36, no. 2, pp. 1071–1092, 2022. [Online]. Available: https://doi.org/10.1137/21M1435288 4

[39] M. Göös, T. Pitassi, and T. Watson, “Query-to-communication lifting for BPP,” *SIAM J. Comput.*, vol. 49, no. 4, 2020. 4, 5

[40] R. Raz and P. McKenzie, “Separation of the monotone NC hierarchy,” *Comb.*, vol. 19, no. 3, pp. 403–435, 1999. 4

[41] M. Göös, T. Pitassi, and T. Watson, “Deterministic communication vs. partition number,” in *FOCS*. IEEE Computer Society, 2015, pp. 1077–1088. 4

[42] A. Chattopadhyay, Y. Filmus, S. Koroth, O. Meir, and T. Pitassi, “Query-to-communication lifting for bpp using inner product,” in *ICALP*, 2019. 4, 5

[43] M. Göös, S. Lovett, R. Meka, T. Watson, and D. Zuckerman, “Rectangles are nonnegative juntas,” in *STOC*. ACM, 2015, pp. 257–266. 4

[44] M. Göös, “Lower bounds for clique vs. independent set,” in *FOCS*. IEEE Computer Society, 2015, pp. 1066–1076. 4

[45] S. F. de Rezende, J. Nordström, and M. Vinyals, “How limited interaction hinders real communication (and what it means for proof and circuit complexity),” in *FOCS*. IEEE Computer Society, 2016, pp. 295–304. 4

[46] X. Wu, P. Yao, and H. S. Yuen, “Raz-mckenzie simulation with the inner product gadget,” *Electron. Colloquium Comput. Complex.*, vol. 24, p. 10, 2017. 4

[47] A. Chattopadhyay, M. Koucký, B. Loff, and S. Mukhopadhyay, “Simulation theorems via pseudorandom properties,” *Comput. Complex.*, vol. 28, no. 4, pp. 617–659, 2019. 4

[48] P. K. Kothari, R. Meka, and P. Raghavendra, “Approximating rectangles by juntas and weakly-exponential lower bounds for LP relaxations of csps,” in *STOC*. ACM, 2017, pp. 590–603. 4

[49] Y. Shi and Y. Zhu, “Quantum communication complexity of block-composed functions,” *Quantum Inf. Comput.*, vol. 9, no. 5&6, pp. 444–460, 2009. 4

[50] A. A. Sherstov, “The pattern matrix method,” *SIAM J. Comput.*, vol. 40, no. 6, pp. 1969–2000, 2011. 4

[51] A. A. Razborov and A. A. Sherstov, “The sign-rank of ac^0,” *SIAM J. Comput.*, vol. 39, no. 5, pp. 1833–1855, 2010. 4

[52] R. Robere, T. Pitassi, B. Rossman, and S. A. Cook, “Exponential lower bounds for monotone span programs,” in *FOCS*. IEEE Computer Society, 2016, pp. 406–415. 4

[53] M. Göös, P. Kamath, T. Pitassi, and T. Watson, “Query-to-communication lifting for P NP,” *Comput. Complex.*, vol. 28, no. 1, pp. 113–144, 2019. 4

[54] J. R. Lee, P. Raghavendra, and D. Steurer, “Lower bounds on the size of semidefinite programming relaxations,” in *STOC*. ACM, 2015, pp. 567–576. 4

[55] H. Buhrman, R. Cleve, and A. Wigderson, “Quantum vs. classical communication and computation,” in *STOC*. ACM, 1998, pp. 63–68. 4

[56] S. Lovett, R. Meka, I. Mertz, T. Pitassi, and J. Zhang, “Lifting with sunflowers,” in *13th Innovations in Theoret-*

ical Computer Science Conference (ITCS 2022). Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2022. 4

[57] S. Aaronson, "BQP and the polynomial hierarchy," in *STOC*, 2010, pp. 141–150. [Online]. Available: http://doi.acm.org/10.1145/1806689.1806711 4

[58] S. Aaronson, H. Buhrman, and W. Kretschmer, "A qubit, a coin, and an advice string walk into a relational problem," *arXiv preprint arXiv:2302.10332*, 2023. 5

[59] D. Gavinsky, "Entangled simultaneity versus classical interactivity in communication complexity," *IEEE Trans. Inf. Theory*, vol. 66, no. 7, pp. 4641–4651, 2020. [Online]. Available: https://doi.org/10.1109/TIT.2020.2976074 5

[60] S. Aaronson and A. Ambainis, "Forrelation: A problem that optimally separates quantum from classical computing," *SIAM J. Comput.*, vol. 47, no. 3, pp. 982–1038, 2018. [Online]. Available: https://doi.org/10.1137/15M1050902 5

[61] N. Bansal and M. Sinha, "k-forrelation optimally separates quantum and classical query complexity," in *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, 2021, pp. 1303–1316. 5

[62] R. Impagliazzo, N. Nisan, and A. Wigderson, "Pseudorandomness for network algorithms," in *STOC*. ACM, 1994, pp. 356–364. 6

[63] R. Adamczak, R. Latała, and R. Meller, "Hanson–wright inequality in banach spaces," *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, vol. 56, no. 4, nov 2020. [Online]. Available: https://doi.org/10.1214%2F19-aihp1041 9