

Fourier Growth of Communication Protocols for XOR Functions

Uma Girish*

Makrand Sinha[†]Avishay Tal[‡]Kewen Wu[§]

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 $h(z) := \mathbb{E}_{\mathbf{x}, \mathbf{y}}[\mathcal{C}(\mathbf{x}, \mathbf{y}) | \mathbf{x} \oplus \mathbf{y} = z]$.

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 $d^{3/2} \cdot \text{polylog}(n)$ 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.

*Princeton University. Email: ugirish@cs.princeton.edu

[†]Simons Institute and University of California at Berkeley. Email: makrand@berkeley.edu

[‡]University of California at Berkeley. Email: atal@berkeley.edu

[§]University of California at Berkeley. Email: shlw_kevin@hotmail.com

Contents

1	Introduction	1
1.1	Main Results	3
1.2	Applications and Connections	4
1.2.1	The Coin Problem and the Gap-Hamming Problem	4
1.2.2	Quantum versus Classical Communication Separation via Lifting	5
1.2.3	General Gadgets and Fourier Growth from Lifting	7
1.2.4	Pseudorandomness for Communication Protocols	8
2	Proof Overview	8
2.1	Level-One Fourier Growth	9
2.2	Level-Two Fourier Growth	12
3	Preliminaries	14
4	Fourier Growth via Martingales in Gaussian Space	17
4.1	Communication Protocols in Gaussian Space	17
4.2	Generalized Communication Protocols	18
4.3	Fourier Growth via Martingales	18
5	Level-One Fourier Growth	20
5.1	Pairwise Clean Protocols	21
5.2	Bounding the Expected Quadratic Variation	22
5.3	Bounds on Step Sizes	24
5.4	Expected Norm of Final Center of Mass	26
5.4.1	Projection on the Subspaces H_A and H_B	26
5.4.2	Projection on the Orthogonal Subspaces H_A^\perp and H_B^\perp	31
6	Level-Two Fourier Growth	34
6.1	4-Wise Clean Protocols	35
6.2	Bounding the Expected Quadratic Variation	38
6.3	Bounds on Step Sizes	40
6.4	Conversion to Second Moment Bounds of the Depth	42
6.5	Second Moment Bounds for the Depth	44
7	Fourier Growth Reductions For General Gadgets	48
8	Directions Towards Further Improvements	52
8.1	Better Lifting Theorems Imply Better Fourier Growth	53
8.2	Sums of Squares of Quadratic Forms for Pairwise Clean Sets	54
A	Gap-Hamming Lower Bounds	59
B	Concentration for Sum of Squares of Quadratic Forms	60

1 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 $f : \{\pm 1\}^n \rightarrow [-1, 1]$ 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 $\widehat{f}(S) = \mathbb{E}[f(\mathbf{x}) \cdot \prod_{i \in S} x_i]$ for a uniformly random $\mathbf{x} \in \{\pm 1\}^n$. The level- k ℓ_1 -Fourier growth of f is the sum of the *absolute values* of its level- k Fourier coefficients,

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

The study of Fourier growth dates back to the work of Mansour [Man95] 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 [Agr20].
- A bound on the level-two Fourier growth already gives pseudorandom generators [CHLT18], oracle separations between BQP and PH [RT19, Wu22], and separations between efficient quantum communication and randomized classical communication [GRT21].

Meanwhile, Fourier growth bounds have been extensively studied and established for various computational models, including small-width DNFs/CNFs [Man95], AC^0 circuits [Tal17], low-sensitivity Boolean functions [GSTW16], small-width branching programs [RSV13, SVW17, CHRT18, LPV22], small-depth decision trees [OS07, Tal20, SSW21], functions related to small-cost communication protocols [GRZ21, GRT21], low-degree $\text{GF}(2)$ polynomials [CHHL19, CHLT18, BIJ⁺21], product tests [Lee19], small-depth parity decision trees [BTW15, GTW21], low-degree bounded functions [IRR⁺21], 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 $\text{GF}(2)$ polynomials (which well-approximate $\text{AC}^0[\oplus]$ when we set $d = \text{polylog}(n)$ [Raz87, Smo87]), while we know a level-one bound of $L_{1,1}(f) \leq O(d)$ due to [CHLT18], the current best bound for levels $k \geq 2$ is roughly $2^{O(dk)}$ [CHHL19], 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 $\text{AC}^0[\oplus]$ [CHLT18], a longstanding open problem in circuit complexity and pseudorandomness.

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 [MO10, HHL18, TWXZ13, SZ08, Zha14].

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 [Raz95].

Definition 1.1. Let $\mathcal{C} : \{\pm 1\}^n \times \{\pm 1\}^n \rightarrow \{\pm 1\}$ be any deterministic communication protocol. The XOR-fiber of the communication protocol \mathcal{C} is the function $h : \{\pm 1\}^n \rightarrow [-1, 1]$ defined at $z \in \{\pm 1\}^n$ as

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

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.

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 [OS07] proved a tight bound of $O(\sqrt{d})$ on the level-one Fourier growth. By inductive tree decompositions, Tal [Tal20] obtained bounds for the higher levels of the form $L_{1,k}(f) \leq \sqrt{d^k \cdot O(\log(n))^{k-1}}$. This was later sharpened by Sherstov, Storozhenko, and Wu [SSW21] to the asymptotically tight bound of $L_{1,k}(f) \leq \sqrt{\binom{d}{k} \cdot O(\log(n))^{k-1}}$ 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 [BTW15] proved a tight level-one bound of $O(\sqrt{d})$. For higher levels, Girish, Tal, and Wu [GTW21] showed that $L_{1,k}(f) \leq \sqrt{d^k \cdot O(k \log(n))^{2k}}$. 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 [GRT21] showed an $O(d^k)$ Fourier growth¹ 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 $d \cdot \text{polylog}(n)$ for level two already requires sophisticated ideas.

The bounds in [GRT21] 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 rectangles induced by the protocol. Such a per-rectangle-based approach cannot give better bounds than the ones in [GRT21], 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 [GTW21] 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.

1.1 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 1.2. *Let $\mathcal{C} : \{\pm 1\}^n \times \{\pm 1\}^n \rightarrow \{\pm 1\}$ be a deterministic communication protocol with at most d bits of communication. Let h be its XOR-fiber as in Definition 1.1. Then, $L_{1,1}(h) = O(\sqrt{d})$.*

Theorem 1.3. *Let $\mathcal{C} : \{\pm 1\}^n \times \{\pm 1\}^n \rightarrow \{\pm 1\}$ be a deterministic protocol communicating at most d bits. Let h be its XOR-fiber as in Definition 1.1. Then, $L_{1,2}(h) = O(d^{3/2} \log^3(n))$.*

Our bounds in Theorems 1.2 and 1.3 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 1.2.3, 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 [GRT21] 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 $x_1 \cdot y_1, \dots, x_d \cdot y_d$, which corresponds to $h(z) = \text{MAJ}(z_1, \dots, z_d)$ with $L_{1,1}(h) = \Theta(\sqrt{d})$. Furthermore, as we discuss later in Subsection 1.2, 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

¹Technically, [GRT21] 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 [GRZ21]

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 2](#) for a detailed discussion.

1.2 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 $g : \Sigma \times \Sigma \rightarrow \{\pm 1\}$ 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 $(x_1, y_1), \dots, (x_n, y_n) \in \Sigma \times \Sigma$ such that $(f \circ g)(x, y) = f(g(x_1, y_1), \dots, g(x_n, y_n))$. The function $f \circ g$ naturally defines a communication problem where Alice is given $x = (x_1, \dots, x_n)$, Bob is given $y = (y_1, \dots, y_n)$, 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.

1.2.1 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 [BV10], AC^0 and $\text{AC}^0[\oplus]$ circuits [CGR14, LSS⁺19], product tests [LV18], and more. Recently, Agrawal [Agr20] showed that the coin problem is closely related to the level-one Fourier growth of functions in \mathcal{F} .

Lemma 1.4 ([Agr20, Lemma 3.2]). *Assume that \mathcal{F} is closed under restrictions and satisfies $L_{1,1}(f) \leq t$ for all $f \in \mathcal{F}$. Then, for all $\rho \in (-1, 1)$ and $f \in \mathcal{F}$,*

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

Note that communication protocols of small cost are closed under restrictions, so are their XOR-fibers (see [GRT21, 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 1.2.3](#)), [Theorem 1.5](#) can be established for general gadgets of small size.

Theorem 1.5. *Let h be the XOR-fiber of a protocol with total communication d . Then for all ρ ,*

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

²Here 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.

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 1.5](#) 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.

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 1.5](#), 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 [Appendix A](#).

Theorem 1.6. *The randomized communication complexity of the Gap-Hamming Problem is $\Omega(n)$.*

We note that there are various different proofs [[CR12](#), [She12](#), [Vid12](#), [RY22](#)] 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.

1.2.2 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 [[GPW20](#)] 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 [[RM99](#), [GPW15](#)], randomized decision trees [[GPW20](#), [CFK⁺19](#)], and more. A beautiful line of work shows how to “lift” many lower bounds in the query model to the communication model [[RM99](#), [GPW15](#), [GLM⁺15](#), [Göös15](#), [dRNV16](#), [HHL18](#), [WYY17](#), [CKLM19](#), [KMR17](#), [SZ09](#), [She11](#), [RS10](#), [RPRC16](#), [GKPW19](#), [LRS15](#)]. 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 [[BCW98](#)]. 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 [[CFK⁺19](#), [LMM⁺22](#)] 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 $\text{polylog}(n)$. 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.

Lifting Forrelation with XOR. The Forrelation function introduced by [Aar10] is defined as follows: on input $x = (x_1, x_2) \in \{\pm 1\}^n$ where n is a power of 2,

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

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

Girish, Raz, and Tal [GRT21] 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 function³ $\text{Forr} \circ \text{XOR}: \{\pm 1\}^n \times \{\pm 1\}^n \rightarrow \{\pm 1\}$ defined as

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

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

The lower bound in [GRT21] 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 1.7. *The randomized communication complexity of $\text{Forr} \circ \text{XOR}$ is $\tilde{\Omega}(n^{1/3})$.*

Theorem 1.7 above gives an $\text{polylog}(n)$ versus $\tilde{\Omega}(n^{1/3})$ separation between the above quantum communication model and the randomized two-party communication model, improving upon the $\text{polylog}(n)$ versus $\tilde{\Omega}(n^{1/4})$ separation from [GRT21]. We emphasize that our separations are for players with *efficient quantum* running time, where the only prior separation was shown by the aforementioned work [GRT21]. 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 [ABK23]. Without the efficiency assumption, a better $\text{polylog}(n)$ versus $\tilde{\Omega}(\sqrt{n})$ separation is known [Gav20] (see [GRT21, Section 1.1] for a more detailed comparison). Optimal Fourier growth bounds of $d \cdot \text{polylog}(n)$ for level two, which we state later in [Conjecture 1.8](#), would also imply such a separation with XOR-lift of Forrelation.

Lifting k -Fold Forrelation with XOR. k -Fold Forrelation [AA18] 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, [BS21] showed that the randomized query complexity of k -Fold Forrelation is $\tilde{\Omega}(n^{1-1/k})$, confirming this conjecture, and a similar separation was proven in [SSW21] for variants of k -Fold Forrelation. These separations, together with lifting theorems with the *inner product* gadget [CFK⁺19], imply an $O(k \log(n))$ vs $\tilde{\Omega}(n^{1-1/k})$ separation between two-party quantum and classical communication complexity, where additionally, the number of rounds⁴ 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

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

⁴We 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 [GRT21].

required by the quantum protocol to achieve the same quantitative separation is reduced by half. Bansal and Sinha [BS21] 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 1.8. *Let $\mathcal{C} : \{\pm 1\}^n \times \{\pm 1\}^n \rightarrow \{\pm 1\}$ be a deterministic communication protocol with at most d bits of communication. Let h be its XOR-fiber as in Definition 1.1. Then for all $k \in \mathbb{N}$, we have that $L_{1,k}(h) \leq (\sqrt{d} \cdot \text{poly}(k, \log(n)))^k$.*

Note that these bounds are consistent with the Fourier growth of parity decision trees (or protocols that only send parities) as shown in [GTW21].

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.

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

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

where \mathbf{x} and \mathbf{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 $\mathcal{C}_{\downarrow \text{XOR}}$, and our main results Theorems 1.2 and 1.3 can be rephrased as

$$L_{1,1}(\text{XOR}, d) \leq O(\sqrt{d}) \quad \text{and} \quad L_{1,2}(\text{XOR}, d) \leq O(d^{3/2} \log^3(n)).$$

In Section 7, we relate $L_{1,k}(g, d)$ to $L_{1,k}(\text{XOR}, d)$, and the main takeaway is, in the study of Fourier growth bounds, constant-sized gadgets are all equivalent.

Theorem 1.9 (Informal, see Theorem 7.5 and Theorem 7.6). *Let $g : \Sigma \times \Sigma \rightarrow \{\pm 1\}$ be a “balanced” gadget. Then*

$$|\Sigma|^{-k} \cdot L_{1,k}(\text{XOR}, d) \leq L_{1,k}(g, d) \leq |\Sigma|^k \cdot L_{1,k}(\text{XOR}, d).$$

Theorem 1.9 also proposes a different approach towards Conjecture 1.8: 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 [Tal20, SSW21]. See Subsection 8.1 for detail.

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., [GPW20]). 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 [HHL18]. However, it only holds for deterministic communication protocols and has a sextic blowup in the cost.

Thus, one can see [Conjecture 1.8](#) as either a further motivation for establishing lifting results for decision trees with constant-sized gadgets, or as a necessary milestone before proving such lifting results.

1.2.4 Pseudorandomness for Communication Protocols

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

$$\left| \mathbb{E}_{\mathbf{x}, \mathbf{y} \sim \nu} [\mathcal{C}(\mathbf{x}, \mathbf{y})] - \mathbb{E}_{\mathbf{r} \sim \{\pm 1\}^\ell} [\mathcal{C}(G(\mathbf{r}))] \right| \leq \varepsilon.$$

[INW94] 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 1.8](#) and the PRG construction from [CHHL19, Theorem 4.5], we would obtain a completely different explicit PRG for this class with error ε and seed length $n + d \cdot \text{polylog}(n/\varepsilon)$.

Paper Organization. An overview of our proofs is given in [Section 2](#). In [Section 3](#) we define necessary notation and recall useful inequalities. [Section 4](#) explains a way to associate the Fourier growth to a martingale process. The proof of level-one bound ([Theorem 1.2](#)) is given in [Section 5](#), and the level-two bound ([Theorem 1.3](#)) in [Section 6](#). The Fourier growth reductions between general gadgets are presented in [Section 7](#). The future directions are discussed in [Section 8](#). Missing proofs can be found in the appendix.

2 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 \mathbf{x} and \mathbf{y} uniformly from $\{\pm 1\}^n$. Let $\mathbf{X}^{(t)} \times \mathbf{Y}^{(t)}$ be the rectangle associated with the random walk at time t . The martingale process tracks the inner product $\langle \mu(\mathbf{X}^{(t)}), \mu(\mathbf{Y}^{(t)}) \rangle$ where $\mu(\mathbf{X}^{(t)}) = \mathbb{E}[\mathbf{x} \mid \mathbf{x} \in \mathbf{X}^{(t)}]$ and $\mu(\mathbf{Y}^{(t)}) = \mathbb{E}[\mathbf{y} \mid \mathbf{y} \in \mathbf{Y}^{(t)}]$ are Alice's and Bob's center of masses.
- Second, to bound the value of the martingale, it is necessary to ensure that neither $\mathbf{X}^{(t)}$ nor $\mathbf{Y}^{(t)}$ become excessively elongated in any direction during the protocol execution. To measure the length of $\mathbf{X}^{(t)}$ in a particular direction $\theta \in \mathbb{S}^{n-1}$, we calculate the variance $\text{Var}[\langle \mathbf{x}, \theta \rangle \mid \mathbf{x} \in \mathbf{X}^{(t)}]$, i.e. the variance of a uniformly random $\mathbf{x} \in \mathbf{X}^{(t)}$ in the direction θ .

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

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* $\langle \mathbf{x}, \theta \rangle$. Through these communication steps, the sets $\mathbf{X}^{(t)}$ and $\mathbf{Y}^{(t)}$ 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 $\mathbf{X}^{(t)}$ and $\mathbf{Y}^{(t)}$ 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.

2.1 Level-One Fourier Growth

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

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

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

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

$$\sum_{i=1}^n \mathbb{E}_\ell [\mathcal{C}(\ell) \cdot \mathbb{E}[\mathbf{x}_i | \mathbf{x} \in \mathbf{X}_\ell] \cdot \mathbb{E}[\mathbf{y}_i | \mathbf{y} \in \mathbf{Y}_\ell]] = \mathbb{E}_\ell [\mathcal{C}(\ell) \cdot \langle \mu(\mathbf{X}_\ell), \mu(\mathbf{Y}_\ell) \rangle] \leq \mathbb{E}_\ell [|\langle \mu(\mathbf{X}_\ell), \mu(\mathbf{Y}_\ell) \rangle|],$$

where $\mu(\mathbf{X}_\ell) = \mathbb{E}[\mathbf{x} | \mathbf{x} \in \mathbf{X}_\ell]$ and $\mu(\mathbf{Y}_\ell) = \mathbb{E}[\mathbf{y} | \mathbf{y} \in \mathbf{Y}_\ell]$ are the center of masses of the rectangle. Our goal is to bound the magnitude of the random variable $\mathbf{z} = \langle \mu(\mathbf{X}_\ell), \mu(\mathbf{Y}_\ell) \rangle$.

We shall show that $\mathbb{E}_\ell[|\mathbf{z}|] \lesssim \sqrt{d}$. Note that $|\mathbf{z}|$ can be as large as d in the worst case — for instance if the first d coordinates of \mathbf{X}_ℓ and \mathbf{Y}_ℓ 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 $(\mathbf{z}^{(t)})_t$ where $\mathbf{z}^{(t)} := \langle \mu(\mathbf{X}^{(t)}), \mu(\mathbf{Y}^{(t)}) \rangle$ tracks the inner product between the center of masses $\mu(\mathbf{X}^{(t)})$ and $\mu(\mathbf{Y}^{(t)})$ of the

current rectangle $\mathbf{X}^{(t)} \times \mathbf{Y}^{(t)}$ at step t . Denote the martingale differences by $\Delta \mathbf{z}^{(t+1)} = \mathbf{z}^{(t+1)} - \mathbf{z}^{(t)}$ and note that if in the t^{th} step Alice sends a message, then

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

where $\Delta \mu(\mathbf{X}^{(t+1)}) = \mu(\mathbf{X}^{(t+1)}) - \mu(\mathbf{X}^{(t)})$ 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 (see [Section 3](#)) since

$$\left(\mathbb{E} \left[\left\| \mathbf{z}^{(d)} \right\|^2 \right] \right)^2 \leq \mathbb{E} \left[\left(\mathbf{z}^{(d)} \right)^2 \right] = \mathbb{E} \left[\sum_{t=0}^{d-1} \left(\Delta \mathbf{z}^{(t+1)} \right)^2 \right], \quad (2.1)$$

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

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 $\Delta \mathbf{z}^{(t+1)}$ 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 $\mu(\mathbf{X}^{(t+1)})$ 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.

Protocols in Gaussian Space. A communication protocol in Gaussian space takes as inputs $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ where \mathbf{x}, \mathbf{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 $\text{sgn}(\mathbf{x})$ and $\text{sgn}(\mathbf{y})$ where $\text{sgn}(\cdot)$ 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 $\mu(X) = \mathbb{E}_{\mathbf{x} \sim \gamma_n} [\mathbf{x} \mid \mathbf{x} \in X]$ and $\mu(Y) = \mathbb{E}_{\mathbf{y} \sim \gamma_n} [\mathbf{y} \mid \mathbf{y} \in Y]$, one can associate the same martingale $(\mathbf{z}^{(t)})_t$ with the protocol in the Gaussian space:

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

It turns out that bounding the quadratic variation of this martingale suffices to give a bound on $L_{1,2}(h)$ (see [Section 4](#)), 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.

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

$$\mathbb{E} \left[\left(\Delta \mathbf{z}^{(t+1)} \right)^2 \mid \mathbf{z}^{(1)}, \dots, \mathbf{z}^{(t)} \right] = \mathbb{E} \left[\left\langle \Delta \mu(\mathbf{X}^{(t+1)}), \mu(\mathbf{Y}^{(t+1)}) \right\rangle^2 \mid \mathbf{z}^{(1)}, \dots, \mathbf{z}^{(t)} \right]$$

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 $\mathbf{X} \times \mathbf{Y}$, before Alice sends the actual message, she sends the inner product $\langle x, \mu(\mathbf{Y}) \rangle$ between her input and Bob’s current center of mass $\mu(\mathbf{Y})$. This partitions Alice’s set \mathbf{X} into affine slices orthogonal to Bob’s current center of mass $\mu(\mathbf{Y})$. Thus the change in Alice’s center of mass in later rounds is orthogonal to $\mu(\mathbf{Y})$ since it only takes place inside the affine slice.

Recall that the martingale $\mathbf{z}^{(t)}$ 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 $\mathbf{X}^{(t)} \times \mathbf{Y}^{(t)}$ fixed by the messages until time t , we have, by Jensen’s inequality,

$$\begin{aligned} \mathbb{E} \left[(\Delta \mathbf{z}^{(t+1)})^2 \mid \mathbf{X}^{(t)}, \mathbf{Y}^{(t)} \right] &= \mathbb{E} \left[\left\langle \Delta \mu(\mathbf{X}^{(t+1)}), \mu(\mathbf{Y}^{(t)}) - \mu(\mathbf{Y}^{(t_{\text{prev}})}) \right\rangle^2 \mid \mathbf{X}^{(t)}, \mathbf{Y}^{(t)} \right] \\ &\leq \mathbb{E} \left[\left\langle \mathbf{x} - \mu(\mathbf{X}^{(t)}), \mu(\mathbf{Y}^{(t)}) - \mu(\mathbf{Y}^{(t_{\text{prev}})}) \right\rangle^2 \mid \mathbf{X}^{(t)}, \mathbf{Y}^{(t)} \right]. \end{aligned} \quad (2.2)$$

Note that the quantity on the right-hand side above is of the form $\langle \mathbf{x} - \mathbb{E}[\mathbf{x}], v \rangle$. In other words, it is the variance of the random vector \mathbf{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 $\mathbf{x}' = \mathbf{x} - \mathbb{E}[\mathbf{x}]$ in \mathbb{R}^n is λ -pairwise clean, if for every $v \in \mathbb{R}^n$,

$$\mathbb{E} \left[\langle \mathbf{x}', v \rangle^2 \right] \leq \lambda \cdot \|v\|^2, \quad (2.3)$$

or equivalently, the operator norm of the covariance matrix $\mathbb{E}[\mathbf{x}'\mathbf{x}'^\top]$ 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 $\mathbf{X}^{(t)}$ 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 (2.3), 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 step⁶, Alice’s current set is pairwise clean. Similar arguments work for Bob.

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

$$\mathbb{E} \left[(\mathbf{z}^{(d)})^2 \right] \lesssim \mathbb{E} \left[\sum_m \left\| \mu(\mathbf{X}^{(\tau_m)}) - \mu(\mathbf{X}^{(\tau_{m-1})}) \right\|^2 + \left\| \mu(\mathbf{Y}^{(\tau'_m)}) - \mu(\mathbf{Y}^{(\tau'_{m-1})}) \right\|^2 \right]$$

⁶We 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.

$$= \mathbb{E} \left[\left\| \mu(\mathbf{X}^{(d)}) \right\|^2 + \left\| \mu(\mathbf{Y}^{(d)}) \right\|^2 \right], \quad (2.4)$$

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 $\mathbb{E}[\mathbf{d}]$ using the level-one inequality (see [Theorem 3.1](#)) — this inequality bounds the Euclidean norm of the center of mass of a set in terms of its Gaussian measure.

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 [[CR12](#), [Vid12](#)] 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 \mathbf{x} sampled from X , there are at most $m \lesssim d$ orthogonal directions $\theta_1, \dots, \theta_m$ such that $\mathbb{E}[\langle \mathbf{x}', \theta_i \rangle^2] \gtrsim 1$ where $\mathbf{x}' = \mathbf{x} - \mathbb{E}[\mathbf{x}]$. This is a consequence of the fact that the expectation of $\mathbf{q} = \sum_{i=1}^m \langle \mathbf{x}', \theta_i \rangle^2$ 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 $\langle \mathbf{x}, \theta \rangle$ 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 $\sqrt{d} \cdot \text{polylog}(n)$ for the level one while directly working with the uniform distribution on the hypercube.

2.2 Level-Two Fourier Growth

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

$$L_{1,2}(h) = \sum_{i \neq j} \left| \widehat{h}(\{i, j\}) \right| = \sum_{i \neq j} \left| \mathbb{E}_{\mathbf{z} \sim \nu} [h(\mathbf{z}) z_i z_j] \right| = \sum_{i \neq j} \left| \mathbb{E}_{\mathbf{x}, \mathbf{y} \sim \nu} [\mathcal{C}(\mathbf{x}, \mathbf{y}) x_i x_j y_i y_j] \right|.$$

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

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

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

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

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

$$\mathbb{E} \left[\left\langle M, \mathbf{x} \dot{\otimes} \mathbf{x} - \mathbb{E} \left[\mathbf{x} \dot{\otimes} \mathbf{x} \right] \right\rangle^2 \right] \leq \lambda \|M\|^2, \quad (2.5)$$

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 $\mathbf{x}^{\dot{\otimes} k}$.

We say a set is *4-wise clean* with parameter λ if (2.5) is preserved for all M with zero diagonal⁸. 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 (2.5), Alice sends the quadratic form $\langle x \dot{\otimes} x, M \rangle$ to Bob until her set is 4-wise clean with parameter λ .

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 $\lambda \leq d \cdot \text{polylog}(n)$. Analogous to (2.1) and (2.4), this gives a bound of $d^3 \cdot \text{polylog}(n)$ on the expected quadratic variation and implies $L_{1,2}(h) \leq d^{3/2} \cdot \text{polylog}(n)$.

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

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

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 [ALM20]. This inequality gives a tail bound for sum of squares of quadratic forms: In particular if M_1, \dots, M_m are matrices with zero diagonal which form an orthonormal set when viewed as n^2 dimensional vectors, then the random variable $\mathbf{q} = \sum_{i=1}^m \langle \mathbf{x} \dot{\otimes} \mathbf{x}, M_i \rangle^2$ satisfies $\Pr_{\mathbf{x} \sim \gamma_n}[\mathbf{q} \geq t] \leq e^{-\Omega(\sqrt{t})}$ for any $t \gtrsim m^2$ (see [Theorem 3.3](#) for a precise statement). 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 $\langle \mathbf{x} \dot{\otimes} \mathbf{x}, M \rangle$ 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 $\pm 1/\text{poly}(n)$. This means that whenever Alice wants to perform cleanup $\langle \mathbf{x} \otimes \mathbf{x}, M \rangle$ 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, \dots, M_m for the cleanups are also fixed along the transcript, and one can apply the aforementioned Hanson-Wright inequality on $\mathbf{q} = \sum_{i=1}^m \langle \mathbf{x} \dot{\otimes} \mathbf{x}, M_i \rangle^2$. Combining the two facts together, we can apply the non-adaptive tail bound above and then condition on obtaining such typical transcript. This shows $\mathbb{E}[\mathbf{q}] \leq d^2 \cdot \text{polylog}(n)$. However, each quadratic form comes from a violation of (2.5) and contributes at least λ to \mathbf{q} in expectation. This implies that $\mathbb{E}[\mathbf{q}] \geq \lambda \cdot m$ and by taking $\lambda = d \cdot \text{polylog}(n)$, we derive that the number of cleanup steps $m \lesssim d$. This shows that the level-two Fourier growth is $O((m + d) \cdot \sqrt{\lambda}) = d^{3/2} \cdot \text{polylog}(n)$ completing the proof.

Note that if we could take $\lambda = \text{polylog}(n)$ while having the same number of cleanup steps $m = d \cdot \text{polylog}(n)$, then we would obtain an optimal level-two bound of $d \cdot \text{polylog}(n)$. However, it is not clear how to use current approach to show this. In [Subsection 8.2](#), 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.

3 Preliminaries

Notation. Throughout, $\log(\cdot)$ and $\ln(\cdot)$ denote logarithms with base 2 and e respectively. We use $\mathbb{N} = \{0, 1, 2, \dots\}$ to denote the set of natural numbers including 0. For $n \in \mathbb{N}$, we write $[n]$ to denote the set $\{1, 2, \dots, n\}$. We use the standard $O(\cdot), \Omega(\cdot), \Theta(\cdot)$ notation, and emphasize that in this paper they only hide universal constants that do not depend on any parameter.

We write \odot to denote the entrywise product for vectors and matrices: in particular, for any $x, y \in \mathbb{R}^n$, we define $x \odot y \in \mathbb{R}^n$ to be a vector where $(x \odot y)_i = x_i y_i$ for $i \in [n]$ and similarly

for any $X, Y \in \mathbb{R}^{n \times m}$, we define $X \odot Y \in \mathbb{R}^{n \times m}$ to be a matrix where $(X \odot Y)_{ij} = X_{ij}Y_{ij}$ for $i \in [n], j \in [m]$. We use $\dot{\otimes}$ to denote a tensor with zeros on the diagonal, i.e., for any $x \in \mathbb{R}^n$, $x \dot{\otimes} x$ is a $n \times n$ matrix where $(x \dot{\otimes} x)_{ij} = x_i x_j$ if $i \neq j$ and zero if $i = j$.

For a vector $x \in \mathbb{R}^n$, we use $\|x\|$ to denote its Euclidean norm. Similarly, for a matrix $X \in \mathbb{R}^{n \times n}$, we use $\|X\|$ to denote its Euclidean norm viewing the matrix X as an n^2 -dimensional vector. For nonzero $x \in \mathbb{R}^n$ or $X \in \mathbb{R}^{n \times n}$, we define $\text{unit}(x) \in \mathbb{R}^n$ or $\text{unit}(X) \in \mathbb{R}^{n \times n}$ as the unit vector along direction x and X respectively: $\text{unit}(x) = x/\|x\|$ and $\text{unit}(X) = X/\|X\|$. We write \mathbb{S}^{n-1} for the unit sphere in \mathbb{R}^n , and write $\mathbb{S}^{n \times n-1}$ for the unit sphere in $\mathbb{R}^{n \times n}$ where additionally the diagonal entries of the $n \times n$ matrices are zero. We use $\langle x, y \rangle$ to denote the inner product between vectors $x, y \in \mathbb{R}^n$ and $\langle X, Y \rangle$ to denote the inner product between matrices $X, Y \in \mathbb{R}^{n \times n}$ viewing them as n^2 -dimensional vectors.

Probability. A probability space is a triple $(\Omega, \mathcal{F}, \xi)$ where Ω is the sample space, \mathcal{F} is a σ -algebra which describes the measurable sets (or events) in the probability space, and ξ is a probability measure. We use $\mathbf{x} \sim \xi$ to denote a random sample distributed according to ξ and $\mathbb{E}_{\mathbf{x} \sim \xi}[f(\mathbf{x})]$ to denote the expectation of a function f under the measure ξ . For any event $S \in \mathcal{F}$, we use $\xi(S)$ to denote the measure of S under ξ . We say an event S holds *almost surely* if $\xi(S) = 1$, i.e., the exceptions to the event have measure zero. For a measurable event $\mathcal{E} \in \mathcal{F}$, we write $\mathcal{F} \cap \{\mathcal{E}\}$ to denote the intersection of the sigma-algebra \mathcal{F} and the sigma-algebra generated by \mathcal{E} .

We use ν_n to denote the uniform probability measure over $\{\pm 1\}^n$ and γ_n to denote the n -dimensional standard Gaussian measure in \mathbb{R}^n . We say a random variable $\mathbf{x} \in \mathbb{R}^n$ is a standard Gaussian in \mathbb{R}^n if its probability distribution is γ_n . We will drop the subscript if the dimension is clear from context. We will also need lower dimensional Gaussian measures: given a linear subspace V of dimension k , there is a k -dimensional standard Gaussian measure on it, which we denote by γ_V . For any measurable subset $S \subseteq \mathbb{R}^n$, we define its ambient space to be the smallest affine subspace $V + t$ that contains it where V is a linear subspace of \mathbb{R}^n and $t \in \mathbb{R}^n$. The relative Gaussian measure of S denoted by $\gamma_{\text{rel}}(S)$ is then defined to be the Gaussian measure of the set $S - t$ under γ_V .

Martingales. Given a sequence of real-valued random variables $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ in a probability space $(\Omega, \mathcal{F}, \xi)$ and a function $f(\mathbf{x}_1, \dots, \mathbf{x}_n)$ satisfying $\mathbb{E}[|f(\mathbf{x}_1, \dots, \mathbf{x}_n)|] < \infty$, the sequence of random variables $\mathbf{z}^{(t)} = \mathbb{E}[f(\mathbf{x}_1, \dots, \mathbf{x}_n) \mid \mathcal{F}^{(t-1)}]$ is called the *Doob martingale* where $\mathcal{F}^{(t-1)}$ is the σ -algebra generated by $\mathbf{x}_1, \dots, \mathbf{x}_{t-1}$ which should be viewed as a record of the randomness of the process until time $t - 1$. The sequence $(\mathcal{F}^{(t)})_t$ is called a *filtration*. A sequence of random variables $(\mathbf{z}^{(t)})_t$ is called *predictable* (or *adapted*) with respect to $\mathcal{F}^{(t)}$ if $\mathbf{z}^{(t)}$ is $\mathcal{F}^{(t)}$ -measurable for every t , meaning that it is determined by the randomness in $\mathcal{F}^{(t)}$.

A discrete random variable $\tau \in \mathbb{N}$ is called a *stopping time* with respect to the filtration $(\mathcal{F}^{(t)})_t$ if the event $\{\tau = t\} \in \mathcal{F}^{(t)}$ for all $t \in \mathbb{N}$, or in words, whether the event $\tau = t$ occurs is determined by the history of the process until time t . All stopping times considered in this paper will be finite. The sigma-algebra $\mathcal{F}^{(\tau)}$ which contains all events that imply the stopping condition is defined as the set of all events \mathcal{E} such that $\mathcal{E} \cap \{\tau = t\} \in \mathcal{F}^{(t)}$ for all $t \in \mathbb{N}$. We also note if one takes an increasing sequence of stopping times $(\tau_m)_m$ then the process defined by $(\mathbf{z}^{(\tau_m)})_m$ is also a martingale.

Let $\Delta \mathbf{z}^{(t)} := \mathbf{z}^{(t)} - \mathbf{z}^{(t-1)}$ be the martingale differences. Note that $\mathbb{E}[\Delta \mathbf{z}^{(t)} \mid \mathcal{F}^{(t-1)}] = 0$ and thus

$$\mathbb{E} \left[\left(\mathbf{z}^{(t)} \right)^2 \right] = \mathbb{E} \left[\left(\sum_{t=1}^n \Delta \mathbf{z}^{(t)} \right)^2 \right] = \mathbb{E} \left[\sum_{t=1}^n \left(\Delta \mathbf{z}^{(t)} \right)^2 \right], \quad (3.1)$$

where the cross terms disappear upon taking expectation. In other words, the martingale differences are orthogonal under taking expectations. The right hand side above is the *expected quadratic variation* of the martingale $(\mathbf{z}^{(t)})_t$. If the sequence $(\mathbf{z}^{(t)})_t$ is vector-valued (resp., matrix-valued) and satisfies $\mathbb{E}[\Delta \mathbf{z}^{(t)} | \mathcal{F}^{(t-1)}] = 0$ where 0 is zero vector (resp., matrix), then we say it is a vector-valued (resp., matrix-valued) martingale with respect to $(\mathcal{F}^{(t)})_t$. Since each coordinate of a vector or matrix-valued martingale is itself a real-valued martingale, vector-valued or matrix-valued martingale differences are also orthogonal under Euclidean norms:

$$\mathbb{E} \left[\left\| \mathbf{z}^{(t)} \right\|^2 \right] = \mathbb{E} \left[\left\| \sum_{t=1}^n \Delta \mathbf{z}^{(t)} \right\|^2 \right] = \mathbb{E} \left[\sum_{t=1}^n \left\| \Delta \mathbf{z}^{(t)} \right\|^2 \right]. \quad (3.2)$$

Useful Inequalities. We will use the well-known level- k inequality [Tal96, KKL88] (see e.g., [O’D14, Level- k Inequalities]). A statement in the Gaussian setting can be found in, e.g., [EM22, Lemma 2.2]. We remark that we will only use the case for $k = 1$ and $k = 2$ here which we state below.⁹

Below we write $\mathbf{1}_A$ for the indicator function of a set and $x_S = \prod_{i \in S} x_i$ for a monomial.

Theorem 3.1 (Level- k Inequality). *Let $k \in \{1, 2\}$. Assume $A \subseteq \mathbb{R}^n$ is measurable and $\mu := \mathbb{E}_{\mathbf{x} \sim \gamma}[\mathbf{1}_A(\mathbf{x})]$. Then, we have*

$$\sum_{|S|=k} \left(\mathbb{E}_{\mathbf{x} \sim \gamma} [\mathbf{1}_A(\mathbf{x}) x_S] \right)^2 \leq 2e^2 \mu^2 \cdot \ln^k(e/\mu).$$

In particular, if μ is non-zero, dividing both sides by μ^2 , we get the following more convenient form for $k \in \{1, 2\}$:

$$\sum_{|S|=k} \left(\mathbb{E}_{\mathbf{x} \sim \gamma} [x_S | \mathbf{x} \in A] \right)^2 \leq 2e^2 \cdot \ln^k(e/\mu).$$

We also make use of the following standard concentration inequality for sums of squares of independent standard Gaussians (see [Ver18]).

Fact 3.2. *Let $m \in \mathbb{N}$ be arbitrary. For any $r \geq 2m$, we have $\Pr_{\mathbf{x} \sim \gamma_m} \left[\sum_{i=1}^m x_i^2 \geq r \right] \leq e^{-r/4}$.*

We also need a concentration inequality for sums of squares of orthogonal quadratic forms over Gaussian random variables. In particular, we prove the following inequality which follows from a generalization of the Hanson-Wright inequality to a Banach space-valued setting [ALM20, Theorem 6]. Since, we only need a special case that is easier to prove, we include a self-contained proof using the Gaussian isoperimetric inequality in Appendix B following [ALM20, Proposition 23].

Theorem 3.3. *Let $m \in \mathbb{N}$ be arbitrary. Let M_1, \dots, M_m be $n \times n$ real matrices where each M_i has zero diagonal, $\langle M_i, M_i \rangle = 1$ and $\langle M_i, M_j \rangle = 0$ for $i \neq j$. Then for any $r \geq 98m$, we have*

$$\Pr_{\mathbf{x} \sim \gamma_m} \left[\sum_{i=1}^m \langle \mathbf{x} \otimes \mathbf{x}, M_i \rangle^2 \geq r \right] \leq \exp \left\{ -\Omega \left(\frac{r}{m + \sqrt{r}} \right) \right\}.$$

We remark that the tail bound above holds more generally for sub-Gaussian random variables \mathbf{x} (see [ALM20]).

⁹Our Theorem 3.1 is slightly different from the references, where they additionally require $\mu \leq 1/e$. By Parseval’s identity, the left hand side is always at most one. Therefore we use a slightly worse bound for the right hand side to allow for the whole range of μ .

4 Fourier Growth via Martingales in Gaussian Space

In this section, we reduce the question of bounding the level-one and level-two Fourier growth to bounding the expected quadratic variation of certain martingales. To analyze these martingales and to prove the optimal bound for the level-one setting, it seems to be crucial to work in the Gaussian setting, so first we give a generic transformation from Boolean to Gaussian. We shall also additionally allow protocols that communicate real numbers to make the analysis easier.

4.1 Communication Protocols in Gaussian Space

Let $\mathcal{C} : \{\pm 1\}^n \times \{\pm 1\}^n \rightarrow \{\pm 1\}$ be a communication protocol with total communication d and h be its XOR-fiber defined in [Definition 1.1](#).

We embed the protocol in the Gaussian space by allowing Alice's and Bob's inputs, x and y respectively, to be real vectors in \mathbb{R}^n — the new protocol $\tilde{\mathcal{C}}$ runs the original protocol \mathcal{C} with Boolean inputs $\text{sgn}(x)$ and $\text{sgn}(y)$ where $\text{sgn}(v) = (\text{sgn}(v_1), \dots, \text{sgn}(v_n))$ denotes the sign function applied pointwise to each coordinate for a vector $v \in \mathbb{R}^n$. The behavior of the communication protocol $\tilde{\mathcal{C}}$ can be defined arbitrarily if any coordinate of $\text{sgn}(x)$ or $\text{sgn}(y)$ is zero since such points have zero measure under the standard n -dimensional Gaussian measure γ_n .

This translation from the Boolean hypercube to the Gaussian space preserves the measure of sets: for any subset $S \subseteq \{\pm 1\}^n$, we have $\nu_n(S) = \gamma_n(\{x \in \mathbb{R}^n \mid \text{sgn}(x) \in S\})$ where ν_n is the uniform measure over $\{\pm 1\}^n$. Moreover, up to some normalizing factor, the Fourier coefficients of h can also be computed by looking at Gaussian inputs. In particular, denoting by $x_S = \prod_{i \in S} x_i$ for a subset $S \subseteq [n]$, we have the following fact.

Fact 4.1. *For all $S \subseteq [n]$, we have $\mathbb{E}_{z \sim \nu_n} [h(z)z_S] = (\pi/2)^{|S|} \mathbb{E}_{\mathbf{x}, \mathbf{y} \sim \gamma_n} [\tilde{\mathcal{C}}(\mathbf{x}, \mathbf{y})x_S y_S]$.*

Proof. Note that for $\mathbf{x} \sim \gamma_n$, the random variable $\text{sgn}(\mathbf{x})$ is distributed as ν_n . Thus, by the definition of the XOR-fiber h and the protocol $\tilde{\mathcal{C}}$, we have

$$\begin{aligned} \mathbb{E}_{z \sim \nu_n} [h(z)z_S] &= \mathbb{E}_{\mathbf{x}, \mathbf{y} \sim \gamma_n} \left[\mathcal{C}(\text{sgn}(\mathbf{x}), \text{sgn}(\mathbf{y})) \cdot \prod_{i \in S} \text{sgn}(x_i) \cdot \text{sgn}(y_i) \right] \\ &= (\pi/2)^{|S|} \mathbb{E}_{\mathbf{x}, \mathbf{y} \sim \gamma_n} \left[\mathcal{C}(\text{sgn}(\mathbf{x}), \text{sgn}(\mathbf{y})) \cdot \prod_{i \in S} x_i \cdot y_i \right] \\ &= (\pi/2)^{|S|} \mathbb{E}_{\mathbf{x}, \mathbf{y} \sim \gamma_n} [\tilde{\mathcal{C}}(\mathbf{x}, \mathbf{y})x_S y_S], \end{aligned}$$

where the second line follows since the expected value of a standard Gaussian in \mathbb{R} conditioned on its sign being fixed to η is $\sqrt{\frac{2}{\pi}} \cdot \eta$ by the following calculation:

$$\mathbb{E}_{x_i \sim \gamma} [x_i \mid \text{sgn}(x_i) = \eta] = \eta \cdot \int_0^\infty \sqrt{\frac{2}{\pi}} \cdot r \cdot e^{-r^2/2} dr = \sqrt{\frac{2}{\pi}} \cdot \eta. \quad \square$$

Remark 4.2. We remark that instead of the Gaussian distribution above, one can work with any distribution where the coordinates are i.i.d. and symmetric around zero. In particular, if ξ is a symmetric probability measure on the real line, and \mathbf{x}, \mathbf{y} are independently drawn vectors in \mathbb{R}^n where each coordinate is i.i.d. sampled from ξ , then $\mathbb{E}_{z \sim \nu_n} [h(z)z_S] = c_\xi^{|S|} \mathbb{E}_{\mathbf{x}, \mathbf{y} \sim \xi^{\otimes n}} [\tilde{\mathcal{C}}(\mathbf{x}, \mathbf{y})x_S y_S]$ where $c_\xi = (\mathbb{E}_{x_i \sim \xi} [|x_i|])^{-2}$. In the case of level-two we will need to work with the truncated Gaussian distribution where each coordinate is sampled independently from the one dimensional standard Gaussian conditioned on being in some interval $[-T, T]$ for $T = \Omega(1)$ in which case c_ξ is upper bounded by a universal constant.

4.2 Generalized Communication Protocols

In the protocol $\tilde{\mathcal{C}}$ defined above, Alice and Bob's inputs x and y are real vectors in \mathbb{R}^n , but in each round they still exchange a single bit based on $\text{sgn}(x)$ and $\text{sgn}(y)$. In order to bound the Fourier growth, it will be more convenient for us to define a notion of generalized communication protocols where parties are also allowed to send real numbers with arbitrary precision in each round. To define this formally, we place certain restrictions on the real communication in the protocol. More formally, in a generalized communication protocol, in each round a player with input $z \in \mathbb{R}^n$ can either send:

- (i) a bit in $\{0, 1\}$ which is purely a function of the Boolean input $\text{sgn}(z)$ and the previous *Boolean* messages, or
- (ii) a real number that is a measurable function of z and the previous (real or Boolean) messages.

The *depth* of a generalized communication protocol is defined to be the maximum number of rounds of communication.

Note that a generalized protocol also generates a “protocol tree” where if in a round a real number is sent, the “children” of that particular “node” are indexed by all possible values in \mathbb{R} . A “transcript” of the protocol can be defined in an analogous way. The set of inputs that reach a particular node of this generalized protocol tree still form a rectangle $X \times Y$ where $X, Y \subseteq \mathbb{R}^n$. We say that a generalized protocol $\bar{\mathcal{C}}$ is equivalent to the protocol $\tilde{\mathcal{C}}$ if $\bar{\mathcal{C}}(x, y) = \tilde{\mathcal{C}}(x, y)$ for every $x, y \in \mathbb{R}^n$ except on a measure zero set.

We will be interested in random walks on such generalized protocol trees when the inputs \mathbf{x} and \mathbf{y} are sampled from a product measure $\xi_x \times \xi_y$ on $\mathbb{R}^n \times \mathbb{R}^n$ and the parties send messages according to the protocol to reach a “leaf”. The random variables corresponding to the messages until any time t generate a filtration $(\mathcal{F}^{(t)})_t$ — this filtration can be thought of as specifying a particular node of the generalized protocol at depth t (equivalently, a partial transcript of the protocol till time t) that was sampled by the process. In this case, conditioned on any event in $\mathcal{F}^{(t)}$, (e.g., any realization of the transcript till time t), almost surely the conditional probability measure on the inputs \mathbf{x}, \mathbf{y} is some product measure on $\xi_x^{(t)} \times \xi_y^{(t)}$ supported on a rectangle $\mathbf{X}^{(t)} \times \mathbf{Y}^{(t)}$ where $\mathbf{X}^{(t)}, \mathbf{Y}^{(t)} \subseteq \mathbb{R}^n$. We shall refer to the random variable $\mathbf{X}^{(t)} \times \mathbf{Y}^{(t)}$ as the current rectangle determined by $\mathcal{F}^{(t)}$. Since we will be working with product measures on inputs \mathbf{x}, \mathbf{y} , the reader can think of conditioning on the filtration $\mathcal{F}^{(t)}$ as essentially conditioning on the inputs being in the rectangle $\mathbf{X}^{(t)} \times \mathbf{Y}^{(t)}$ or equivalently a partial transcript till time t .

4.3 Fourier Growth via Martingales

We will now relate Fourier growth to the quadratic variation of a martingale. Towards this end, we first note that in light of [Fact 4.1](#), the level- k Fourier growth of the XOR-fiber h of the original communication protocol is given by

$$\begin{aligned} L_{1,k}(h) &= \sum_{\substack{S \subseteq [n] \\ |S|=k}} \left| \mathbb{E}_{z \sim \nu_n} [h(z) \mathbf{z}_S] \right| = (\pi/2)^k \sum_{\substack{S \subseteq [n] \\ |S|=k}} \left| \mathbb{E}_{\mathbf{x}, \mathbf{y} \sim \gamma_n} [\bar{\mathcal{C}}(\mathbf{x}, \mathbf{y}) \mathbf{x}_S \mathbf{y}_S] \right| \\ &= (\pi/2)^k \max_{(\eta_S)_{|S|=k}} \sum_{\substack{S \subseteq [n] \\ |S|=k}} \eta_S \mathbb{E}_{\mathbf{x}, \mathbf{y} \sim \gamma_n} [\bar{\mathcal{C}}(\mathbf{x}, \mathbf{y}) \mathbf{x}_S \mathbf{y}_S], \end{aligned} \quad (4.1)$$

where $\bar{\mathcal{C}}$ is any generalized protocol that is equivalent to $\tilde{\mathcal{C}}$ and $\eta_S \in \{\pm 1\}$.

We now express the right hand side above as an inner product. Let ℓ be a random leaf of the generalized protocol tree $\bar{\mathcal{C}}$ induced by taking $\mathbf{x}, \mathbf{y} \sim \gamma_n$ and let $\mathbf{X}_\ell \times \mathbf{Y}_\ell$ be the corresponding rectangle in the generalized protocol tree. Then,

$$\begin{aligned} \sum_{S \subseteq [n], |S|=k} \eta_S \mathbb{E}_{\mathbf{x}, \mathbf{y} \sim \gamma_n} [\bar{\mathcal{C}}(\mathbf{x}, \mathbf{y}) \mathbf{x}_S \mathbf{y}_S] &= \mathbb{E}_\ell \left[\mathbb{E}_{\mathbf{x}, \mathbf{y} \sim \gamma} \left[\sum_{S \subseteq [n], |S|=k} \eta_S \cdot \bar{\mathcal{C}}(\mathbf{x}, \mathbf{y}) \mathbf{x}_S \mathbf{y}_S \mid (\mathbf{x}, \mathbf{y}) \in \mathbf{X}_\ell \times \mathbf{Y}_\ell \right] \right] \\ &= \mathbb{E}_\ell \left[\bar{\mathcal{C}}(\ell) \mathbb{E}_{\mathbf{x}, \mathbf{y} \sim \gamma} \left[\sum_{S \subseteq [n], |S|=k} \eta_S \cdot \mathbf{x}_S \mathbf{y}_S \mid (\mathbf{x}, \mathbf{y}) \in \mathbf{X}_\ell \times \mathbf{Y}_\ell \right] \right] \\ &\leq \mathbb{E}_\ell \left[\left| \sum_{S \subseteq [n], |S|=k} \eta_S \mathbb{E}[\mathbf{x}_S \mid \mathbf{x} \in \mathbf{X}_\ell] \cdot \mathbb{E}[\mathbf{y}_S \mid \mathbf{y} \in \mathbf{Y}_\ell] \right| \right], \quad (4.2) \end{aligned}$$

where the second line follows since ℓ is a leaf and determines the answer and the third line follows since \mathbf{x} and \mathbf{y} are independent conditioned on being in the rectangle $\mathbf{X}_\ell \times \mathbf{Y}_\ell$.

Thus, specializing (4.2) to the level-one ($k = 1$) and level-two cases ($k = 2$), from (4.1) we get that

$$\begin{aligned} L_{1,1}(h) &\leq \frac{\pi}{2} \cdot \max_\eta \mathbb{E}_\ell \left[\left| \sum_{i=1}^n \eta_i \cdot \mathbb{E}[\mathbf{x}_i \mid \mathbf{x} \in \mathbf{X}_\ell] \cdot \mathbb{E}[\mathbf{y}_i \mid \mathbf{y} \in \mathbf{Y}_\ell] \right| \right], \\ L_{1,2}(h) &\leq \frac{\pi^2}{4} \cdot \max_\eta \mathbb{E}_\ell \left[\left| \sum_{i,j=1}^n \eta_{ij} \cdot \mathbb{E}[\mathbf{x}_{ij} \mid \mathbf{x} \in \mathbf{X}_\ell] \cdot \mathbb{E}[\mathbf{y}_{ij} \mid \mathbf{y} \in \mathbf{Y}_\ell] \right| \right], \end{aligned}$$

where for $L_{1,1}$ we optimize over $\eta \in \{\pm 1\}^n$ and for $L_{1,2}$ we optimize over η being an $n \times n$ symmetric matrix with zeros on the diagonals and ± 1 entries otherwise.

To make the above more compact, we respectively define $\mu(X) \in \mathbb{R}^n$ and $\sigma(X) \in \mathbb{R}^{n \times n}$ to be the level-one and level-two centers of mass of a set $X \subseteq \mathbb{R}^n$:

$$\mu(X) = \mathbb{E}_{\mathbf{x} \sim \gamma_n} [\mathbf{x} \mid \mathbf{x} \in X] \quad \text{and} \quad \sigma(X) = \mathbb{E}_{\mathbf{x} \sim \gamma_n} [\mathbf{x} \dot{\otimes} \mathbf{x} \mid \mathbf{x} \in X]. \quad (4.3)$$

Then, upper bounding the constants in the above inequality ($\pi/2$ and $\pi^2/4$) by 4, we get

$$\begin{aligned} L_{1,1}(h) &\leq 4 \cdot \max_\eta \mathbb{E}_\ell [|\langle \mu(\mathbf{X}_\ell), \eta \odot \mu(\mathbf{Y}_\ell) \rangle|], \\ L_{1,2}(h) &\leq 4 \cdot \max_\eta \mathbb{E}_\ell [|\langle \sigma(\mathbf{X}_\ell), \eta \odot \sigma(\mathbf{Y}_\ell) \rangle|], \end{aligned} \quad (4.4)$$

where η is understood to be the same as before.

Moving forward, we fix an arbitrary η for both cases $k \in \{1, 2\}$ and define a martingale process $(\mathbf{z}_k^{(t)})_t$ that captures the right hand side above. For this we note that a generalized communication protocol, where Alice's and Bob's inputs are sampled from the Gaussian distribution, naturally induces a discrete-time random walk on the corresponding (generalized) protocol tree where at time t we are at a node at depth t with the corresponding rectangle $\mathbf{X}^{(t)} \times \mathbf{Y}^{(t)}$. Then, we have the following proposition.

Proposition 4.3. $\mu(\mathbf{X}^{(t)})$ and $\mu(\mathbf{Y}^{(t)})$ are vector-valued martingales taking values in \mathbb{R}^n and $\sigma(\mathbf{X}^{(t)})$ and $\sigma(\mathbf{Y}^{(t)})$ are matrix-valued martingales taking values in $\mathbb{R}^{n \times n}$.

Note that if in the t^{th} round Alice speaks, then $\mu(\mathbf{Y}^{(t)})$ and $\sigma(\mathbf{Y}^{(t)})$ do not change and similarly if Bob speaks, then $\mu(\mathbf{X}^{(t)})$ and $\sigma(\mathbf{X}^{(t)})$ do not change. The above proposition implies that the real-valued processes

$$\mathbf{z}_1^{(t)} = \left\langle \mu(\mathbf{X}^{(t)}), \eta \odot \mu(\mathbf{Y}^{(t)}) \right\rangle \text{ and } \mathbf{z}_2^{(t)} = \left\langle \sigma(\mathbf{X}^{(t)}), \eta \odot \sigma(\mathbf{Y}^{(t)}) \right\rangle, \quad (4.5)$$

each form a Doob martingale with respect to the natural filtration induced by the random walk on the protocol tree. Note that taking a random walk on the tree until we hit a leaf generates the marginal distribution on ℓ given in (4.4). Let \mathbf{d} be the stopping time when this martingale hits a leaf and stops (i.e., the depth of the random leaf). Thus, by the orthogonality of martingale differences $\Delta \mathbf{z}_k^{(t)} = \mathbf{z}_k^{(t)} - \mathbf{z}_k^{(t-1)}$ from (3.1), we get that for $k \in \{1, 2\}$, one can upper bound the Fourier growth in terms of expected quadratic variation of the above martingales:

Proposition 4.4. For $k \in \{1, 2\}$, $\frac{1}{4} \cdot L_{1,k}(h) \leq \max_{\eta} \sqrt{\mathbb{E} \left[\left(\mathbf{z}_k^{(\mathbf{d})} \right)^2 \right]} = \max_{\eta} \sqrt{\mathbb{E} \left[\sum_{t=1}^{\mathbf{d}} \left(\Delta \mathbf{z}_k^{(t)} \right)^2 \right]}$.

The martingale implicitly depends on η as used in (4.4) and hence the maximum. Moreover, the martingale also depends on the underlying generalized communication protocol $\bar{\mathcal{C}}$. In the next two sections, we will show that after transforming the original communication protocol into “clean” protocols, the expected quadratic variations of $(\mathbf{z}_1^{(t)})_t$ and $(\mathbf{z}_2^{(t)})_t$ are $O(d)$ and $O(d^3) \cdot \text{polylog}(n)$ respectively. This will then imply our main theorems.

Remark 4.5. Note that Proposition 4.3 still holds even if the input distribution is not the Gaussian distribution, but some other product probability measure on the inputs \mathbf{x}, \mathbf{y} . This also implies that $\mathbf{z}_k^{(t)}$ for $k \in \{1, 2\}$ is a martingale. In particular, for the level-two case, we will need to use a truncated Gaussian distribution. In light of Remark 4.2, Proposition 4.4 still suffices for us with a different constant instead of 1/4. We also remark that we shall also need to truncate the real messages being used in the protocol for the level-two case to a finite precision, so the generalized protocols for the level-two case only have Boolean communication. However, to obtain the optimal level-one bound allowing generalized protocols that communicate real values seems to be crucial.

5 Level-One Fourier Growth

In this section, we will give a proof of Theorem 1.2 that $L_{1,1}(h) = O(\sqrt{d})$. We start with a d -round communication protocol $\tilde{\mathcal{C}}$ over the Gaussian space as defined in Subsection 4.1. Given the discussion in the previous section and Proposition 4.4, our task ultimately reduces to bounding the expected quadratic variation of the martingale that results from the protocol $\bar{\mathcal{C}}$. For example, one can simply take $\bar{\mathcal{C}} = \tilde{\mathcal{C}}$, but, as discussed in Section 2, the individual step sizes of this martingale can be quite large in the worst-case and it is not so easy to leverage cancellations here to bound the quadratic variation by $O(d)$.

So, we first define a *generalized* communication protocol $\bar{\mathcal{C}}$ that is equivalent to the original protocol $\tilde{\mathcal{C}}$ but has additional “cleanup” rounds where Alice and Bob reveal certain linear forms of their inputs so that their sets are pairwise clean in the sense described in the overview. These cleanup steps allow us to keep track of the quadratic variation more easily.

5.1 Pairwise Clean Protocols

To define a clean protocol, we first define the notion of a pairwise clean set. Let $X \subseteq \mathbb{R}^n$. We say that the set X is *pairwise clean in a direction* $a \in \mathbb{S}^{n-1}$ with parameter λ if

$$\mathbb{E}_{\mathbf{x} \sim \gamma} \left[\langle \mathbf{x} - \mu(X), a \rangle^2 \mid \mathbf{x} \in X \right] \leq \lambda, \quad (5.1)$$

where we recall that $\mu(X) = \mathbb{E}_{\mathbf{x} \sim \gamma} [\mathbf{x} \mid \mathbf{x} \in X]$ is the level-one center of mass of X .

The above condition implies that for a random vector \mathbf{x} sampled from γ conditioned on X , its variance along the direction a is bounded by λ . We say that the set X is *pairwise clean* (with parameter λ) if it is clean in *every direction* $a \in \mathbb{S}^{n-1}$. Equivalently, the operator norm of the covariance matrix of the random vector \mathbf{x} is bounded by λ .

We call a generalized communication protocol pairwise clean with parameter λ if at the start of a new “phase” of the protocol, the corresponding rectangle $X \times Y$ satisfies that both X and Y are pairwise clean. Starting from a communication protocol $\tilde{\mathcal{C}}$ in the Gaussian space, we will transform it into a pairwise clean protocol $\bar{\mathcal{C}}$ by proceeding from top to bottom and adding certain Gram-Schmidt orthogonalization and cleanup steps.

In particular, consider an intermediate node in the protocol tree of $\tilde{\mathcal{C}}$. Before Alice sends her bit as in the original protocol $\tilde{\mathcal{C}}$, she first performs an orthogonalization step by revealing the inner-product between her input and Bob’s current level-one center of mass. After this, she sends her bit according to the original protocol and afterwards she repeatedly cleans her current set X by revealing $\langle x, a \rangle \in \mathbb{R}$ while X is not clean along the direction a orthogonal to previous directions. Once X becomes clean, they proceed to the next round. We now describe this formally.

Construction of pairwise clean protocol $\bar{\mathcal{C}}$ from $\tilde{\mathcal{C}}$. We set $\lambda = 100$. The construction of the new protocol is recursive and we first define some notation. Consider an intermediate node of the new protocol $\bar{\mathcal{C}}$ at depth t . We use the random variable $\mathbf{X}^{(t)} \subseteq \mathbb{R}^n$ (resp., $\mathbf{Y}^{(t)} \subseteq \mathbb{R}^n$) to denote the set of inputs of Alice (resp., Bob) reaching the node. If Alice reveals a linear form in this step, we use $\mathbf{a}^{(t)} \in \mathbb{R}^n$ to denote the vector of the linear form; otherwise, we set $\mathbf{a}^{(t)}$ to be the all-zeroes vector. We define $\mathbf{b}^{(t)}$ similarly for Bob. Throughout the protocol, we will abbreviate $\mathbf{u}^{(t)} = \mu(\mathbf{X}^{(t)})$ and $\mathbf{v}^{(t)} = \mu(\mathbf{Y}^{(t)})$ for Alice’s and Bob’s current center of mass respectively.

1. At the beginning, Alice receives an input $x \in \mathbb{R}^n$ and Bob receives an input $y \in \mathbb{R}^n$.
2. We initialize $t \leftarrow 0$, $\mathbf{X}^{(0)}, \mathbf{Y}^{(0)} \leftarrow \mathbb{R}^n$, and $\mathbf{a}^{(0)}, \mathbf{b}^{(0)} \leftarrow 0^n$.
3. For each phase $i = 1, 2, \dots, d$: suppose we are starting the cleanup for a node at depth i in the original protocol $\tilde{\mathcal{C}}$ and suppose we are at a node of depth t in the new protocol $\bar{\mathcal{C}}$. If it is Alice’s turn to speak in $\tilde{\mathcal{C}}$:

(a) **Orthogonalization by revealing the correlation with Bob’s center of mass.**

Alice begins by revealing the inner product of her input x with Bob’s current (signed) center of mass $\eta \odot \mathbf{v}^{(t)}$. Since in the previous steps, she has already revealed the inner product with Bob’s previous centers of mass, for technical reasons, we will only have Alice announce the inner product with the component of $\eta \odot \mathbf{v}^{(t)}$ that is orthogonal to the previous directions along which Alice announced the inner product. More formally, let $\mathbf{a}^{(t+1)}$ be the component of $\eta \odot \mathbf{v}^{(t)}$ that is orthonormal to all previous directions $\mathbf{a}^{(1)}, \dots, \mathbf{a}^{(t)}$, i.e.,

$$\mathbf{a}^{(t+1)} = \text{unit} \left(\eta \odot \mathbf{v}^{(t)} - \sum_{\tau=1}^t \langle \eta \odot \mathbf{v}^{(t)}, \mathbf{a}^{(\tau)} \rangle \cdot \mathbf{a}^{(\tau)} \right).$$

Alice computes $\bar{\mathbf{c}}^{(t+1)} \leftarrow \langle x, \mathbf{a}^{(t+1)} \rangle$ and sends $\bar{\mathbf{c}}^{(t+1)}$ to Bob. Set $\mathbf{b}^{(t+1)} \leftarrow 0^n$. Increment t by 1 and go to step (b).

- (b) **Original communication.** Alice sends the bit $\bar{\mathbf{c}}^{(t+1)}$ that she was supposed to send in $\tilde{\mathcal{C}}$ based on previous messages and the input x . Set $\mathbf{a}^{(t+1)}, \mathbf{b}^{(t+1)} \leftarrow 0^n$. Increment t by 1 and go to step (c).
- (c) **Cleanup steps.** While there exists some direction $a \in \mathbb{S}^{n-1}$ orthogonal to the previous directions (i.e., satisfying $\langle a, \mathbf{a}^{(\tau)} \rangle = 0$ for all $\tau \in [t]$) such that $\mathbf{X}^{(t)}$ is *not pairwise clean* in direction a , Alice computes $\bar{\mathbf{c}}^{(t+1)} \leftarrow \langle x, a \rangle$ and sends this to Bob. Set $\mathbf{a}^{(t+1)} \leftarrow a$ and $\mathbf{b}^{(t+1)} \leftarrow 0^n$. Increment t by 1. Repeat step (c) as long as $\mathbf{X}^{(t)}$ is not pairwise clean; otherwise increment i by 1 and go back to the for-loop in step 3 which starts the new phase.

If it is Bob's turn to speak, we define everything similarly with the role of $x, \mathbf{a}, \mathbf{X}, \mathbf{v}$ switched with $y, \mathbf{b}, \mathbf{Y}, \mathbf{u}$.

4. Finally at the end of the protocol, the value $\bar{\mathcal{C}}(x, y)$ is determined based on all the previous communication and the corresponding output it defines in $\tilde{\mathcal{C}}$.

We note some basic properties that directly follow from the description. First we note that the steps 3(a), 3(b), and 3(c) always occur in sequence for each party and we refer to such a sequence of steps as a *phase* for that party. Note that there are at most d phases. If a new phase starts at time t , then the current rectangle $\mathbf{X}^{(t)} \times \mathbf{Y}^{(t)}$ is pairwise clean for both parties by construction. Also, note that the non-zero vectors in the sequence $(\mathbf{a}^{(t)})_t$ (resp., $(\mathbf{b}^{(t)})_t$) form an orthonormal set. We also note that the Boolean communication in step 3(b) is solely determined by the original protocol and hence only depends on the previous Boolean messages.

Lastly, each phase has one 3(a) and 3(b) step, followed by potentially many 3(c) steps. However, the following claim shows that it is always finite.

Claim 5.1. Let ℓ be an arbitrary leaf of the protocol $\bar{\mathcal{C}}$ and $D(\ell)$ be its depth. Then $D(\ell) \leq 2n + 2d$. Moreover, along this path there are at most $2d$ many steps 3(a) and 3(b).

Proof. We count the number of communication steps separately:

- **Steps 3(a) and 3(b).** Steps 3(a) and 3(b) occur once in every phase, thus at most d times.
- **Step 3(c).** For Alice, each time she communicates at step 3(c) $a \in \mathbb{R}^n$, the direction is orthogonal to all previous $\mathbf{a}^{(t)}$'s. Since the dimension of \mathbb{R}^n is n , this happens at most n times. Similar argument works for Bob.

Thus in total we have at most $2n + 2d$ steps. □

We will eventually show that the *expected* depth of the protocol $\bar{\mathcal{C}}$ is $O(d)$ when $\mathbf{x}, \mathbf{y} \sim \gamma_n$.

5.2 Bounding the Expected Quadratic Variation

Consider a random walk on the protocol tree generated by the new protocol $\bar{\mathcal{C}}$ when the parties are given independent inputs $\mathbf{x}, \mathbf{y} \sim \gamma_n$. Consider the corresponding level-one martingale process defined in (4.5). Formally, at time t the process is defined by

$$\mathbf{z}_1^{(t)} = \left\langle \mathbf{u}^{(t)}, \eta \odot \mathbf{v}^{(t)} \right\rangle,$$

where we recall that $\mathbf{u}^{(t)} = \mu(\mathbf{X}^{(t)})$ and $\mathbf{v}^{(t)} = \mu(\mathbf{Y}^{(t)})$ and $\eta \in \{\pm 1\}^n$ is a fixed sign vector.

The martingale process stops once it hits a leaf of the protocol $\bar{\mathcal{C}}$. Let \mathbf{d} denote the (stopping) time when this happens. Note that $\mathbb{E}[\mathbf{d}]$ is exactly the expected depth of the protocol $\bar{\mathcal{C}}$. Then, in light of [Proposition 4.4](#), to prove [Theorem 1.2](#), it suffices to prove the following.

Lemma 5.2. $\mathbb{E} \left[\sum_{t=1}^{\mathbf{d}} \left(\Delta \mathbf{z}_1^{(t)} \right)^2 \right] = O(d)$.

We will prove this in two steps. We first show that the only change in the value of the martingale occurs during the orthogonalization step 3(a). This is because in each phase, Alice's change of center of mass in steps 3(b) and 3(c) is always orthogonal to $\eta \odot \mathbf{v}^{(t)}$ so they do not change the value of the martingale $\mathbf{z}_1^{(t)}$ as discussed in [Section 2](#). Moreover, recalling [\(2.2\)](#), since Alice's node was pairwise clean just before Alice sent the message in step 3(a), the expected change $\mathbb{E} \left[\left(\Delta \mathbf{z}_1^{(t+1)} \right)^2 \right]$ can be bounded in terms of the squared norm of the change that occurred in $\mathbf{u}^{(t)}$ between the current round and the last round where Alice was in step 3(a). A similar argument works for Bob.

Formally, this is encapsulated by the next lemma for which we need some additional definition. Let $(\mathcal{F}^{(t)})_t$ be the natural filtration induced by the random walk on the generalized protocol tree with respect to which $\mathbf{z}_1^{(t)}$ is a Doob martingale and also $\mathbf{u}^{(t)}, \mathbf{v}^{(t)}$ form vector-valued martingales (recall [Proposition 4.3](#)). Note that $\mathcal{F}^{(t)}$ fixes all the rectangles encountered during times $0, \dots, t$ and thus for $\tau \leq t$, the random variables $\mathbf{u}^{(\tau)}, \mathbf{v}^{(\tau)}, \mathbf{z}_1^{(\tau)}$ are determined, in particular, they are $\mathcal{F}^{(t)}$ -measurable. Recalling that $\lambda = 100$ is the cleanup parameter, we then have the following. Below we assume without any loss of generality that Alice speaks first and, in particular, we note that Alice speaks in step 3(a) for the first time at time zero.

Lemma 5.3 (Step Size). *Let $0 = \tau_1 < \tau_2 < \dots \leq \mathbf{d}$ be a sequence of stopping times with τ_m being the index of the round where Alice speaks in step 3(a) for the m^{th} time or \mathbf{d} if there is no such round. Then, for any integer $m \geq 2$,*

$$\mathbb{E} \left[\left(\Delta \mathbf{z}_1^{(\tau_{m+1})} \right)^2 \mid \mathcal{F}^{(\tau_m)} \right] \leq \lambda \cdot \left\| \mathbf{v}^{(\tau_m)} - \mathbf{v}^{(\tau_{m-1})} \right\|^2,$$

and moreover, for any $t \in \mathbb{N}$, we have that

$$\mathbb{E} \left[\left(\Delta \mathbf{z}_1^{(t+1)} \right)^2 \mid \mathcal{F}^{(t)}, \tau_{m-1} < t < \tau_m, \text{ Alice speaks at time } t \right] = 0.$$

A similar statement also holds if Bob speaks where \mathbf{v} is replaced by \mathbf{u} and the sequence (τ_m) is replaced by (τ'_m) where τ'_m is the index of the round where Bob speaks in step 3(a) for the m^{th} time or \mathbf{d} if there is no such round.

In particular, we see that the steps 3(b) and 3(c) do not contribute to the quadratic variation and only the steps 3(a) do. Also, since the first time Alice and Bob speak, they start in step 3(a), we also note that $\mathbf{u}^{(\tau_1)}$ and $\mathbf{v}^{(\tau'_1)}$ are their initial centers of mass which are both zero.

We shall prove the above lemma in [Subsection 5.3](#) and continue with the bound on the quadratic variation here. Using [Lemma 5.3](#), we have

$$\mathbb{E} \left[\sum_{t=1}^{\mathbf{d}} \left(\Delta \mathbf{z}_1^{(t)} \right)^2 \right] \leq \lambda \cdot \mathbb{E} \left[\sum_{m \geq 2} \left\| \mathbf{v}^{(\tau_m)} - \mathbf{v}^{(\tau_{m-1})} \right\|^2 + \left\| \mathbf{u}^{(\tau'_m)} - \mathbf{u}^{(\tau'_{m-1})} \right\|^2 \right].$$

On the other hand, by the orthogonality of vector-valued martingale differences from (3.2), we have

$$\mathbb{E} \left[\sum_{m \geq 2} \left\| \mathbf{v}^{(\tau_m)} - \mathbf{v}^{(\tau_{m-1})} \right\|^2 \right] = \mathbb{E} \left[\left\| \mathbf{v}^{(d)} \right\|^2 \right].$$

A similar statement holds for $(\mathbf{u}^{(t)})_t$. Therefore,

$$\mathbb{E} \left[\sum_{t=1}^d \left(\Delta z_1^{(t)} \right)^2 \right] \leq \lambda \cdot \left(\mathbb{E} \left[\left\| \mathbf{u}^{(d)} \right\|^2 \right] + \mathbb{E} \left[\left\| \mathbf{v}^{(d)} \right\|^2 \right] \right). \quad (5.2)$$

We prove the following in Subsection 5.4 to upper bound the quantity on the right hand side above. Loosely speaking, by an application of level-one inequalities (see Theorem 3.1), the lemma below ultimately boils down to a bound on the expected number of cleanup steps.

Lemma 5.4 (Final Center of Mass). $\mathbb{E} \left[\left\| \mathbf{u}^{(d)} \right\|^2 + \left\| \mathbf{v}^{(d)} \right\|^2 \right] = O(d)$.

Since $\lambda = 100$, plugging in the bounds from the above into (5.2) readily implies Lemma 5.2. Together with Proposition 4.4, this completes the proof of Theorem 1.2.

5.3 Bounds on Step Sizes (Proof of Lemma 5.3)

Let us abbreviate $\tau = \tau_m$. Observe that

$$\begin{aligned} \mathbb{E} \left[\left(\Delta z_1^{(\tau+1)} \right)^2 \middle| \mathcal{F}^{(\tau)} \right] &= \mathbb{E} \left[\left\langle \mathbf{u}^{(\tau+1)} - \mathbf{u}^{(\tau)}, \eta \odot \mathbf{v}^{(\tau)} \right\rangle^2 \middle| \mathcal{F}^{(\tau)} \right] \\ &= \mathbb{E} \left[\left\langle \mathbf{u}^{(\tau+1)}, \eta \odot \mathbf{v}^{(\tau)} \right\rangle^2 - \left\langle \mathbf{u}^{(\tau)}, \eta \odot \mathbf{v}^{(\tau)} \right\rangle^2 \middle| \mathcal{F}^{(\tau)} \right], \end{aligned} \quad (5.3)$$

where the second line is due to $(\mathbf{u}^{(t)})_t$ being a vector-valued martingale and thus $\mathbb{E} [\mathbf{u}^{(\tau+1)} \mid \mathcal{F}^{(\tau)}] = \mathbf{u}^{(\tau)}$.

We first consider the case that at time τ a new phase starts for Alice. By construction, this means that the current rectangle $\mathbf{X}^{(\tau)} \times \mathbf{Y}^{(\tau)}$ determined by $\mathcal{F}^{(\tau)}$ is pairwise clean with parameter λ , and since Alice is in step 3(a) at the start of a new phase, $\mathbf{a}^{(\tau+1)}$ is chosen to be the (normalized) component of $\eta \odot \mathbf{v}^{(\tau)}$ that is orthogonal to previous directions $\mathbf{a}^{(0)}, \dots, \mathbf{a}^{(\tau)}$. Let $\beta^{(\tau+1)} := \langle \eta \odot \mathbf{v}^{(\tau)}, \mathbf{a}^{(\tau+1)} \rangle$ be the length of this component before normalization. Note that $\beta^{(\tau+1)}$ is $\mathcal{F}^{(\tau)}$ -measurable (i.e., it is determined by $\mathcal{F}^{(\tau)}$).

We now claim that components of $\mathbf{u}^{(\tau+1)}$ and $\mathbf{u}^{(\tau)}$ are the same along any of the previous directions $\mathbf{a}^{(0)}, \dots, \mathbf{a}^{(\tau)}$. So in (5.3), they cancel out and the only relevant quantity is the component in the direction $\mathbf{a}^{(\tau+1)}$. This follows since, in all the previous steps $t \leq \tau$, Alice has already fixed $\langle x, \mathbf{a}^{(t)} \rangle$. This implies that for any $\mathbf{X}^{(\tau)}$ and $\mathbf{X}^{(\tau+1)}$ that are determined by $\mathcal{F}^{(\tau+1)}$, the inner product with all the previous $\mathbf{a}^{(0)}, \dots, \mathbf{a}^{(\tau)}$ is fixed over the choice of x from both rectangles. Formally, we have that for any $x \in \mathbf{X}^{(\tau)}$ and $x' \in \mathbf{X}^{(\tau+1)}$, it holds that $\langle x, \mathbf{a}^{(t)} \rangle = \langle x', \mathbf{a}^{(t)} \rangle$ for any $t \leq \tau$. In particular, since $\mathbf{u}^{(\tau)} = \mu(\mathbf{X}^{(\tau)})$ and $\mathbf{u}^{(\tau+1)} = \mu(\mathbf{X}^{(\tau+1)})$ are the corresponding centers of mass, we have that

$$\left\langle \mathbf{u}^{(\tau+1)}, \mathbf{a}^{(t)} \right\rangle = \left\langle \mathbf{u}^{(\tau)}, \mathbf{a}^{(t)} \right\rangle \text{ for all } t \leq \tau. \quad (5.4)$$

This, together with (5.3) and recalling that $\beta^{(\tau+1)}$ is determined by $\mathcal{F}^{(\tau)}$, implies that

$$\mathbb{E} \left[\left(\Delta z_1^{(\tau+1)} \right)^2 \middle| \mathcal{F}^{(\tau)} \right] = \left(\beta^{(\tau+1)} \right)^2 \cdot \mathbb{E} \left[\left\langle \mathbf{u}^{(\tau+1)}, \mathbf{a}^{(\tau+1)} \right\rangle^2 - \left\langle \mathbf{u}^{(\tau)}, \mathbf{a}^{(\tau+1)} \right\rangle^2 \middle| \mathcal{F}^{(\tau)} \right]. \quad (5.5)$$

We now bound the term outside the expectation by the change in the center of mass $\mathbf{v}^{(\cdot)}$ and the term inside the expectation by the fact that the set is pairwise clean.

Term Outside the Expectation. Recall that $\mathbf{a}^{(\tau+1)}$ is chosen to be the (normalized) component of $\eta \odot \mathbf{v}^{(\tau)}$ that is orthogonal to the span of $\mathbf{a}^{(0)}, \dots, \mathbf{a}^{(\tau)}$. Since $\eta \odot \mathbf{v}^{(\tau_{m-1})}$ is in the span of $\mathbf{a}^{(1)}, \dots, \mathbf{a}^{(\tau_{m-1}+1)}$ and $\tau_{m-1} + 1 \leq \tau = \tau_m$, it is orthogonal to $\mathbf{a}^{(\tau+1)}$. Hence,

$$\boldsymbol{\beta}^{(\tau+1)} = \left\langle \eta \odot \mathbf{v}^{(\tau)}, \mathbf{a}^{(\tau+1)} \right\rangle = \left\langle \eta \odot \left(\mathbf{v}^{(\tau)} - \mathbf{v}^{(\tau_{m-1})} \right), \mathbf{a}^{(\tau+1)} \right\rangle.$$

Since $\mathbf{a}^{(\tau+1)}$ is a unit vector and each entry of η is in $\{\pm 1\}$, this implies that

$$\left(\boldsymbol{\beta}^{(\tau+1)} \right)^2 \leq \left\| \mathbf{v}^{(\tau)} - \mathbf{v}^{(\tau_{m-1})} \right\|^2. \quad (5.6)$$

Term Inside the Expectation. Since $(\mathbf{u}^{(\tau)})$ is a vector-valued martingale with respect to $\mathcal{F}^{(\tau)}$, and $\mathbf{a}^{(\tau+1)}$ is $\mathcal{F}^{(\tau)}$ -measurable (determined by $\mathcal{F}^{(\tau)}$), we have that

$$\mathbb{E} \left[\left\langle \mathbf{u}^{(\tau+1)}, \mathbf{a}^{(\tau+1)} \right\rangle^2 - \left\langle \mathbf{u}^{(\tau)}, \mathbf{a}^{(\tau+1)} \right\rangle^2 \middle| \mathcal{F}^{(\tau)} \right] = \mathbb{E} \left[\left\langle \mathbf{u}^{(\tau+1)} - \mathbf{u}^{(\tau)}, \mathbf{a}^{(\tau+1)} \right\rangle^2 \middle| \mathcal{F}^{(\tau)} \right].$$

Since Alice is in step 3(a), her message fixes $\langle \mathbf{x}, \mathbf{a}^{(\tau+1)} \rangle$ at time τ for every $\mathbf{x} \in \mathbf{X}^{(\tau+1)}$. Thus,

$$\begin{aligned} \mathbb{E} \left[\left\langle \mathbf{u}^{(\tau+1)} - \mathbf{u}^{(\tau)}, \mathbf{a}^{(\tau+1)} \right\rangle^2 \middle| \mathcal{F}^{(\tau)} \right] &= \mathbb{E} \left[\left\langle \mathbb{E}_{\mathbf{x} \sim \gamma} [\mathbf{x} \mid \mathbf{x} \in \mathbf{X}^{(\tau+1)}] - \mathbf{u}^{(\tau)}, \mathbf{a}^{(\tau+1)} \right\rangle^2 \middle| \mathcal{F}^{(\tau)} \right] \\ &= \mathbb{E} \left[\mathbb{E}_{\mathbf{x} \sim \gamma} \left[\left\langle \mathbf{x} - \mathbf{u}^{(\tau)}, \mathbf{a}^{(\tau+1)} \right\rangle^2 \middle| \mathbf{x} \in \mathbf{X}^{(\tau+1)} \right] \middle| \mathcal{F}^{(\tau)} \right] \\ &= \mathbb{E} \left[\left\langle \mathbf{x} - \mathbf{u}^{(\tau)}, \mathbf{a}^{(\tau+1)} \right\rangle^2 \middle| \mathcal{F}^{(\tau)} \right], \end{aligned} \quad (5.7)$$

where the last line follows from the tower property of conditional expectation.

Recall that $\mathbf{u}^{(\tau)} = \mu(\mathbf{X}^{(\tau)})$ is the center of mass. Moreover, the unit vector $\mathbf{a}^{(\tau+1)}$ is determined by $\mathcal{F}^{(\tau)}$ and also the conditional distribution of \mathbf{x} conditioned on $\mathcal{F}^{(\tau)}$ is that of $\mathbf{x} \sim \gamma$ conditioned on $\mathbf{x} \in \mathbf{X}^{(\tau)}$. Thus, using the fact that $\mathbf{X}^{(\tau)}$ is pairwise clean since Alice is in step 3(a), the right hand side in (5.7) is at most λ .

Final Bound. Substituting the above in (5.5), we have

$$\mathbb{E} \left[\left(\Delta \mathbf{z}_1^{(\tau+1)} \right)^2 \middle| \mathcal{F}^{(\tau)} \right] \leq \lambda \cdot \left(\boldsymbol{\beta}^{(\tau+1)} \right)^2 \leq \lambda \cdot \left\| \mathbf{v}^{(\tau)} - \mathbf{v}^{(\tau_{m-1})} \right\|^2,$$

where the second inequality follows from (5.6). This completes the proof of the first statement.

For the moreover part, let us condition on the event $\tau_{m-1} < t < \tau_m$ where Alice speaks at time t . Note that such t must all lie in the same phase of the protocol where Alice is the only one speaking. So, Bob's center of mass does not change from the time τ_{m-1} till t , i.e., $\mathbf{v}^{(t+1)} = \mathbf{v}^{(\tau_{m-1})}$. Thus we have $\Delta \mathbf{z}_1^{(t+1)} = \langle \mathbf{u}^{(t+1)} - \mathbf{u}^{(t)}, \eta \odot \mathbf{v}^{(\tau_{m-1})} \rangle$. Analogous to (5.4), the component of Alice's center of mass along the previous directions are fixed. Thus $\langle \mathbf{u}^{(t+1)}, \mathbf{a}^{(r)} \rangle = \langle \mathbf{u}^{(t)}, \mathbf{a}^{(r)} \rangle$ for all $r \leq t$. Furthermore, by construction, $\eta \odot \mathbf{v}^{(\tau_{m-1})}$ lies in the linear subspace spanned by $\mathbf{a}^{(0)}, \dots, \mathbf{a}^{(\tau_{m-1}+1)}$. Therefore, since $\tau_{m-1} + 1 \leq t$, it follows that $\Delta \mathbf{z}_1^{(t+1)} = 0$.

5.4 Expected Norm of Final Center of Mass (Proof of Lemma 5.4)

Let $\mathbf{H}_A = \mathbf{H}_A^{(d)}$ be the (random) linear subspace spanned by the vectors $\mathbf{a}^{(0)}, \dots, \mathbf{a}^{(d)}$ and similarly, let $\mathbf{H}_B = \mathbf{H}_B^{(d)}$ be the linear subspace spanned by the vectors $\mathbf{b}^{(0)}, \dots, \mathbf{b}^{(d)}$. For any linear subspace V of \mathbb{R}^n , we denote by $\mathbf{\Pi}_V$ and $\mathbf{\Pi}_{V^\perp}$ the projectors on the subspace V and its orthogonal complement V^\perp respectively. Then, we have that

$$\|\mathbf{u}^{(d)}\|^2 = \|\mathbf{\Pi}_{\mathbf{H}_A} \mathbf{u}^{(d)}\|^2 + \|\mathbf{\Pi}_{\mathbf{H}_A^\perp} \mathbf{u}^{(d)}\|^2 \quad \text{and} \quad \|\mathbf{v}^{(d)}\|^2 = \|\mathbf{\Pi}_{\mathbf{H}_B} \mathbf{v}^{(d)}\|^2 + \|\mathbf{\Pi}_{\mathbf{H}_B^\perp} \mathbf{v}^{(d)}\|^2.$$

Note that the non-zero vectors in $(\mathbf{a}^{(t)})_t$ and $(\mathbf{b}^{(t)})_t$ form an orthonormal basis for the subspaces \mathbf{H}_A and \mathbf{H}_B respectively. Moreover, for each $t \leq d$, the inner product $\langle x, \mathbf{a}^{(t)} \rangle$ is fixed for every $x \in \mathbf{X}^{(d)}$ and the inner product $\langle y, \mathbf{b}^{(t)} \rangle$ is also fixed for every $y \in \mathbf{Y}^{(d)}$ where $\mathbf{X}^{(d)} \times \mathbf{Y}^{(d)}$ is the current rectangle determined by $\mathcal{F}^{(d)}$. In particular, since $\mathbf{u}^{(d)}$ is the center of mass of $\mathbf{X}^{(d)}$, this implies that

$$\begin{aligned} \|\mathbf{\Pi}_{\mathbf{H}_A} \mathbf{u}^{(d)}\|^2 &= \sum_{t=1}^d \langle \mathbf{u}^{(d)}, \mathbf{a}^{(t)} \rangle^2 = \sum_{t=1}^d \left(\mathbb{E}_{\mathbf{x} \sim \gamma} \left[\langle \mathbf{x}, \mathbf{a}^{(t)} \rangle \mid \mathbf{x} \in \mathbf{X}^{(d)} \right] \right)^2 \\ &= \sum_{t=1}^d \mathbb{E}_{\mathbf{x} \sim \gamma} \left[\langle \mathbf{x}, \mathbf{a}^{(t)} \rangle^2 \mid \mathbf{x} \in \mathbf{X}^{(d)} \right], \end{aligned}$$

where the second line follows from the inner product being fixed in $\mathbf{X}^{(d)}$. Therefore, we have

$$\|\mathbf{u}^{(d)}\|^2 = \underbrace{\sum_{t=1}^d \mathbb{E}_{\mathbf{x} \sim \gamma} \left[\langle \mathbf{x}, \mathbf{a}^{(t)} \rangle^2 \mid \mathbf{x} \in \mathbf{X}^{(d)} \right]}_{\mathbf{p}_A} + \underbrace{\|\mathbf{\Pi}_{\mathbf{H}_A^\perp} \mathbf{u}^{(d)}\|^2}_{\mathbf{q}_A}.$$

In an analogous fashion,

$$\|\mathbf{v}^{(d)}\|^2 = \underbrace{\sum_{t=1}^d \mathbb{E}_{\mathbf{y} \sim \gamma} \left[\langle \mathbf{y}, \mathbf{b}^{(t)} \rangle^2 \mid \mathbf{y} \in \mathbf{Y}^{(d)} \right]}_{\mathbf{p}_B} + \underbrace{\|\mathbf{\Pi}_{\mathbf{H}_B^\perp} \mathbf{v}^{(d)}\|^2}_{\mathbf{q}_B}.$$

We next show that both $\mathbb{E}[\mathbf{p}_A + \mathbf{p}_B]$ and $\mathbb{E}[\mathbf{q}_A + \mathbf{q}_B]$ are at most $O(d)$. The former follows from stopping time and concentration arguments laid out in the overview that there cannot be too many orthogonal directions where $\mathbb{E} \left[\langle \mathbf{x}, \mathbf{a}^{(t)} \rangle^2 \right]$ is large. The latter follows from an application of level-one inequalities.

We will bound the norm of the projection on the subspaces \mathbf{H}_A and \mathbf{H}_B , which corresponds to the quantity $\mathbb{E}[\mathbf{p}_A + \mathbf{p}_B]$, in [Subsection 5.4.1](#) and bound the norm of the projection on the orthogonal subspaces \mathbf{H}_A^\perp and \mathbf{H}_B^\perp , which corresponds to the quantity $\mathbb{E}[\mathbf{q}_A + \mathbf{q}_B]$, in [Subsection 5.4.2](#).

5.4.1 Projection on the Subspaces \mathbf{H}_A and \mathbf{H}_B

We shall show that the expected norm of the final center of mass when projected on the subspaces \mathbf{H}_A and \mathbf{H}_B is

$$\mathbb{E}[\mathbf{p}_A + \mathbf{p}_B] = O(d).$$

Towards this end, define the random variable $\mathbf{k}_t = \mathbf{k}_t(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{a}^{(t)} \rangle^2 + \langle \mathbf{y}, \mathbf{b}^{(t)} \rangle^2$ for each $t \in \mathbb{N}$. Note that the vectors $\mathbf{a}^{(t)}$'s are being chosen adaptively depending on the previous inner

products $\langle \mathbf{x}, \mathbf{a}^{(\tau)} \rangle$ for $\tau < t$, as well as the Boolean communication bits from step 3(b), thus they are functions of \mathbf{x} and \mathbf{y} as well here. Observe that

$$\mathbb{E}[\mathbf{p}_A + \mathbf{p}_B] = \mathbb{E} \left[\sum_{t=1}^{\mathbf{d}} \mathbb{E}[\mathbf{k}_t \mid \mathcal{F}^{(\mathbf{d})}] \right] = \mathbb{E}_{\mathbf{x}, \mathbf{y} \sim \gamma} \left[\sum_{t=1}^{\mathbf{d}} \mathbf{k}_t \right].$$

We now divide the time sequence into successive intervals of different lengths $r \cdot 4d$ for $r = 1, 2, \dots$. Then we bound the expected sum of \mathbf{k}_t within each time interval by $O(rd)$. We further argue that the probability that the stopping time \mathbf{d} lies in the r -th interval is at most $2 \cdot 2^{-r}$. In particular, for $r \in \mathbb{N}$, letting interval $I_r = \left\{ \binom{r}{2} \cdot 4d + 1, \dots, \binom{r+1}{2} \cdot 4d \right\}$, which is of length $4dr$, we show the following.

Claim 5.5. For any $r \in \mathbb{N}$, we have

$$\mathbb{E}_{\mathbf{x}, \mathbf{y} \sim \gamma} \left[\sum_{t \in I_r} \mathbf{k}_t \mid \mathbf{d} > \binom{r}{2} \cdot 4d \right] \leq 20dr + 4 \ln \left(\frac{1}{\Pr[\mathbf{d} > \binom{r}{2} \cdot 4d]} \right).$$

We shall prove the above claim later since it is the most involved part of the proof. The previous claim readily implies the following probability bounds.

Claim 5.6. For any $r \in \mathbb{N}$, we have $\Pr[\mathbf{d} > \binom{r}{2} \cdot 4d] \leq 2 \cdot 2^{-r}$.

Proof of Claim 5.6. We bound $\Pr[\mathbf{d} > \binom{r}{2} \cdot 4d]$ by induction on r . The claim trivially holds for $r = 1$.

Now we proceed to analyze the event $\mathbf{d} \geq \binom{r+1}{2} \cdot 4d$. Observe that Claim 5.1 implies that there are at most $2d$ many step 3(a) and 3(b) throughout the protocol. Thus if the event above occurs, there are at least $4dr - 2d \geq 2dr$ many time steps $t \in I_r$ where the process is in step 3(c).

By the definition of the cleanup step, if $X \times Y$ is a rectangle determined¹⁰ by $\mathcal{F}^{(t-1)} \cap \{\mathbf{d} > \binom{r}{2} \cdot 4d\}$ where the process is in step 3(c) and Alice speaks, then

$$\mathbb{E}_{\mathbf{x} \sim \gamma} [\mathbf{k}_t \mid (\mathbf{x}, \mathbf{y}) \in X \times Y] = \mathbb{E}_{\mathbf{x} \sim \gamma} \left[\langle \mathbf{x}, \mathbf{a}^{(t)} \rangle^2 \mid \mathbf{x} \in X \right] \geq \mathbb{E}_{\mathbf{x} \sim \gamma} \left[\langle \mathbf{x} - \mu(X), \mathbf{a}^{(t)} \rangle^2 \mid \mathbf{x} \in X \right] \geq \lambda,$$

where $\lambda = 100$ is the cleanup parameter and $\mu(X) = \mathbb{E}_{\mathbf{x} \sim \gamma}[\mathbf{x} \mid \mathbf{x} \in X]$ is the center of mass. This is because $\mathbf{a}^{(t)}$ is chosen to be a unit vector in a direction where the current set (conditioned on the history) is not pairwise clean. A similar statement holds if Bob speaks in step 3(c) for the random variable $\langle \mathbf{y}, \mathbf{b}^{(t)} \rangle^2$ where \mathbf{y} is sampled from γ conditioned on Y .

By the tower property of conditional expectation, the above implies that

$$100 \cdot 2dr \cdot \Pr[\mathbf{d} > \binom{r+1}{2} \cdot 4d \mid \mathbf{d} > \binom{r}{2} \cdot 4d] \leq \mathbb{E} \left[\sum_{t \in I_r} \mathbf{k}_t \mid \mathbf{d} > \binom{r}{2} \cdot 4d \right].$$

Recall that Claim 5.5 implies that the right hand side is at most $\leq 20dr + 4 \ln \left(\frac{1}{\Pr[\mathbf{d} > \binom{r}{2} \cdot 4d]} \right)$. We consider two cases:

- (i) if $\Pr[\mathbf{d} > \binom{r}{2} \cdot 4d] \leq 2^{-r}$, then clearly $\Pr[\mathbf{d} > \binom{r+1}{2} \cdot 4d] \leq 2^{-r}$ as well as required;

¹⁰It suffices to consider such events since we have a product measure on $\mathbf{X}^{(t)} \times \mathbf{Y}^{(t)}$ conditioned on $\mathcal{F}^{(t)}$ and \mathbf{d} is a stopping time and is $\mathcal{F}^{(t)}$ -measurable (i.e., determined by the randomness in $\mathcal{F}^{(t)}$).

(ii) otherwise $\Pr[\mathbf{d} > \binom{r}{2} \cdot 4d] \geq 2^{-r}$ and $20dr + 4 \left(\frac{1}{\Pr[\mathbf{d} > \binom{r}{2} \cdot 4d]} \right) \leq 20dr + 4r$, then it follows that

$$\Pr[\mathbf{d} > \binom{r+1}{2} \cdot 4d \mid \mathbf{d} > \binom{r}{2} \cdot 4d] \leq 1/2,$$

and by induction this implies that $\Pr[\mathbf{d} > \binom{r+1}{2} \cdot 4d] \leq 1/2 \cdot \Pr[\mathbf{d} > \binom{r}{2} \cdot 4d] \leq 2^{-r}$. \square

These claims imply that

$$\begin{aligned} \mathbb{E}[\mathbf{p}_A + \mathbf{p}_B] &\leq \mathbb{E} \left[\sum_{r=0}^{\infty} 1[\mathbf{d} > \binom{r}{2} \cdot 4d] \cdot \sum_{t \in I_r} \mathbf{k}_t \right] \\ &= \sum_{r=0}^{\infty} \Pr[\mathbf{d} > \binom{r}{2} \cdot 4d] \cdot \mathbb{E} \left[\sum_{t \in I_r} \mathbf{k}_t \mid \mathbf{d} > \binom{r}{2} \cdot 4d \right] \\ &\leq \sum_{r=0}^{\infty} \left(2^{1-r} \cdot O(rd) + 4 \cdot \Pr[\mathbf{d} > \binom{r}{2} \cdot 4d] \cdot \ln \left(\frac{1}{\Pr[\mathbf{d} > \binom{r}{2} \cdot 4d]} \right) \right) \\ &\leq \sum_{r=0}^{\infty} (2^{1-r} \cdot O(rd) + O((r+1)2^{-r})) \leq O(d), \end{aligned}$$

where the last line uses the fact that $x \ln(1/x) \leq O((r+1)2^{-r})$ for $0 \leq x \leq 2 \cdot 2^{-r}$ and $r \in \mathbb{N}$. This proves the desired bound on $\mathbb{E}[\mathbf{p}_A + \mathbf{p}_B]$ assuming [Claim 5.5](#) which we prove next.

Proof of Claim 5.5. To prove the claim, we need to analyze the expectation of $\sum_{t \in I_r} \mathbf{k}_t$ under \mathbf{x}, \mathbf{y} sampled from γ conditioned on the event $\mathbf{d} \geq \binom{r}{2} \cdot 4d$.

We first describe an equivalent way of sampling from this distribution which will be easier for analysis. First, we recall that the definition of the cleanup protocol implies that the Boolean communication in $\tilde{\mathcal{C}}$ is solely determined by the previous Boolean communication, since it is specified by the original protocol $\tilde{\mathcal{C}}$ (and thus \mathcal{C}) before the cleanup.

Let us fix any Boolean string $c \in \{0,1\}^*$ that is a valid Boolean transcript in the original communication protocol $\tilde{\mathcal{C}}$. This defines a rectangle $X_c \times Y_c \subseteq \mathbb{R}^n \times \mathbb{R}^n$ consisting of all pairs of inputs to Alice and Bob that result in the Boolean transcript c in $\tilde{\mathcal{C}}$. If we sample $\mathbf{x}, \mathbf{y} \sim \gamma$ conditioned on $\mathbf{d} > \binom{r}{2} \cdot 4d$ and output the unique $(\mathbf{X}_c, \mathbf{Y}_c)$ such that $(\mathbf{x}, \mathbf{y}) \in \mathbf{X}_c \times \mathbf{Y}_c$, we obtain a distribution on rectangles. We use $\gamma(X_c \times Y_c \mid \mathbf{d} > \binom{r}{2} \cdot 4d)$ to denote the probability of obtaining $X_c \times Y_c$ by this sampling process so that $\sum_c \gamma(X_c \times Y_c \mid \mathbf{d} > \binom{r}{2} \cdot 4d) = 1$.

Now consider the following two-stage sampling process. First, we sample a rectangle $X_c \times Y_c$ according to the above distribution, and then we sample the inputs \mathbf{x}, \mathbf{y} sampled from γ_n conditioned on the event that $\{(\mathbf{x}, \mathbf{y}) \in X_c \times Y_c\} \wedge \{\mathbf{d} > \binom{r}{2} \cdot 4d\}$. We shall show the following claim for any rectangle $X_c \times Y_c$ that could be sampled in the first step.

Claim 5.7. $\mathbb{E}_{\mathbf{x}, \mathbf{y} \sim \gamma} [\sum_{t \in I_r} \mathbf{k}_t \mid \mathbf{d} > 4d \binom{r}{2}, (\mathbf{x}, \mathbf{y}) \in X_c \times Y_c] \leq 12dr + 4 \ln \left(\frac{1}{\Pr[\mathbf{d} > 4d \binom{r}{2}, (\mathbf{x}, \mathbf{y}) \in X_c \times Y_c]} \right)$.

Assuming the above, and taking an expectation over $X_c \times Y_c$ drawn with probability $\gamma(X_c \times Y_c \mid \mathbf{d} > \binom{r}{2} \cdot 4d)$, we immediately obtain [Claim 5.5](#):

$$\begin{aligned} &\mathbb{E}_{\mathbf{x}, \mathbf{y} \sim \gamma} \left[\sum_{t \in I_r} \mathbf{k}_t \mid \mathbf{d} > \binom{r}{2} \cdot 4d \right] \\ &\leq 12dr + 4 \cdot \sum_{c \in \{0,1\}^*, |c| \leq d} \gamma(X_c \times Y_c \mid \mathbf{d} > \binom{r}{2} \cdot 4d) \cdot \left(\ln \left(\frac{1}{\gamma(X_c \times Y_c \mid \mathbf{d} > \binom{r}{2} \cdot 4d)} \right) + \ln \left(\frac{1}{\Pr[\mathbf{d} > \binom{r}{2} \cdot 4d]} \right) \right) \end{aligned}$$

$$\begin{aligned}
&\leq 12dr + 4 \cdot \ln(3^d) + 4 \cdot \ln\left(\frac{1}{\Pr[\mathbf{d} > \binom{r}{2} \cdot 4d]}\right) && \text{(by concavity of } \ln(\cdot)\text{)} \\
&\leq 20dr + 4 \cdot \ln\left(\frac{1}{\Pr[\mathbf{d} > \binom{r}{2} \cdot 4d]}\right). && \square
\end{aligned}$$

To complete the proof, we now prove [Claim 5.7](#).

Proof of Claim 5.7. Fix any c such that $\gamma(X_c \times Y_c | \mathbf{d} > \binom{r}{2} \cdot 4d) > 0$. We will bound the expectation of the quantity $\sum_{t \in I_r} \mathbf{k}_t = \sum_{t \in I_r} \langle \mathbf{x}, \mathbf{a}^{(t)} \rangle^2 + \langle \mathbf{y}, \mathbf{b}^{(t)} \rangle^2$ where \mathbf{x}, \mathbf{y} are sampled from γ_n conditioned on the event that $\{(\mathbf{x}, \mathbf{y}) \in X_c \times Y_c\} \wedge \{\mathbf{d} > \binom{r}{2} \cdot 4d\}$. Note that $\mathbf{a}^{(t)}, \mathbf{b}^{(t)}, \mathbf{d}$ are functions of the previous messages of the protocol and hence also the inputs \mathbf{x}, \mathbf{y} . Once we condition on the above event, the Boolean communication is also fixed to be c .

To analyze the above conditioning, we first do a thought experiment and consider a different protocol that takes standard Gaussian inputs (without any conditioning) and show a tail bound for the random variable $\sum_{t \in I_r} \mathbf{k}_t$ for this new protocol. In the last step, we will use it to compute the expectation we ultimately want.

Protocol \mathcal{C}_c . The protocol \mathcal{C}_c always communicates according to the fixed transcript c in a Boolean communication step and otherwise according to the cleanup protocol $\bar{\mathcal{C}}$ on any input x, y . Consider the random walk on this new protocol tree where the inputs $\mathbf{x}, \mathbf{y} \sim \gamma$ (without any conditioning). Let $(\mathcal{G}^{(t)})_t$ be the associated filtration of the new protocol \mathcal{C}_c which can be identified with the collection of all partial transcripts till time t . Note that the vectors $\mathbf{a}^{(t)}$ and $\mathbf{b}^{(t)}$ in this new protocol are determined only by the previous real communication since the Boolean communication is fixed to c . This also implies that the vectors $\mathbf{a}^{(t)}$ and $\mathbf{b}^{(t)}$ form a predictable sequence with respect to the filtration $(\mathcal{G}^{(t)})_t$. Moreover, by the definition of the protocol the next non-zero vector $\mathbf{a}^{(\cdot)}$ is chosen to be a unit vector orthogonal to the previously chosen $\mathbf{a}^{(\cdot)}$'s and the same holds for the vectors $\mathbf{b}^{(\cdot)}$.

We denote by $\mathbf{k}_t^{(c)}$ the random variable that captures \mathbf{k}_t for the protocol \mathcal{C}_c , i.e., $\mathbf{k}_t^{(c)} = \langle \mathbf{x}, \mathbf{a}^{(t)} \rangle^2 + \langle \mathbf{y}, \mathbf{b}^{(t)} \rangle^2$ for $\mathbf{x}, \mathbf{y} \sim \gamma$ and $\mathbf{a}^{(t)}, \mathbf{b}^{(t)}$ defined by the protocol \mathcal{C}_c . Observe that if $(\mathbf{x}, \mathbf{y}) \in X_c \times Y_c$ then $\mathbf{k}_t^{(c)} = \mathbf{k}_t$.

Consider the behavior of the protocol \mathcal{C}_c at some fixed time t . The nice thing about the protocol \mathcal{C}_c is that conditioned on all previous real messages for $\tau < t$, both \mathbf{x} and \mathbf{y} are standard Gaussian distributions on an affine subspace of \mathbb{R}^n (defined by the previous messages). Then, at time t , since $\mathbf{a}^{(t)}$ is orthogonal to the directions used in all previous real messages, it follows that the distribution of $\langle \mathbf{x}, \mathbf{a}^{(t)} \rangle$ conditioned on any event in $\mathcal{G}^{(t-1)}$ is an independent standard Gaussian for every t if $\mathbf{a}^{(t)}$ is non-zero. The same holds for $\langle \mathbf{y}, \mathbf{b}^{(t)} \rangle$ as well. This last fact uses that the projection of a multi-variate standard Gaussian γ_n in orthonormal directions yields independent real-valued standard Gaussians.

This implies that each new $\langle \mathbf{x}, \mathbf{a}^{(t)} \rangle^2$ and $\langle \mathbf{y}, \mathbf{b}^{(t)} \rangle^2$ is an independent chi-squared random variable conditioned on the history (up to depth $\binom{r}{2} \cdot 4d$) of the random walk. Therefore, [Fact 3.2](#) implies that

$$\Pr_{\mathbf{x}, \mathbf{y} \sim \gamma} \left[\sum_{t \in I_r} \mathbf{k}_t^{(c)}(\mathbf{x}, \mathbf{y}) \geq 2|I_r| + s \mid \mathcal{G}^{(\binom{r}{2} \cdot 4d)} \right] \leq e^{-s/4}.$$

Since $|I_r| \leq 4dr$, we have $\Pr_{\mathbf{x}, \mathbf{y} \sim \gamma} \left[\sum_{t \in I_r} \mathbf{k}_t^{(c)}(\mathbf{x}, \mathbf{y}) \geq 8dr + s \right] \leq e^{-s/4}$.

Computing the Original Expectation. Let us compare the probability of the above tail event in the original protocol $\bar{\mathcal{C}}$ where inputs \mathbf{x}, \mathbf{y} are sampled from γ conditioned on the event that $\{(\mathbf{x}, \mathbf{y}) \in X_c \times Y_c\} \wedge \{\mathbf{d} > \binom{r}{2} \cdot 4d\}$. We can write

$$\begin{aligned} & \Pr_{(\mathbf{x}, \mathbf{y}) \sim \gamma} \left[\sum_{t \in I_r} \mathbf{k}_t(\mathbf{x}, \mathbf{y}) \geq 8dr + s \mid \mathbf{d} > \binom{r}{2} \cdot 4d, (\mathbf{x}, \mathbf{y}) \in X_c \times Y_c \right] \\ &= \frac{\Pr_{\mathbf{x}, \mathbf{y} \sim \gamma} \left[\sum_{t \in I_r} \mathbf{k}_t(\mathbf{x}, \mathbf{y}) \geq 8dr + s, (\mathbf{x}, \mathbf{y}) \in X_c \times Y_c, \mathbf{d} > \binom{r}{2} \cdot 4d \right]}{\Pr_{\mathbf{x}, \mathbf{y} \sim \gamma} \left[(\mathbf{x}, \mathbf{y}) \in X_c \times Y_c, \mathbf{d} > \binom{r}{2} \cdot 4d \right]}. \end{aligned} \quad (5.8)$$

We then bound the numerator by

$$\begin{aligned} & \Pr_{\mathbf{x}, \mathbf{y} \sim \gamma} \left[\sum_{t \in I_r} \mathbf{k}_t(\mathbf{x}, \mathbf{y}) \geq 8dr + s, (\mathbf{x}, \mathbf{y}) \in X_c \times Y_c, \mathbf{d} > \binom{r}{2} \cdot 4d \right] \\ &= \Pr_{\mathbf{x}, \mathbf{y} \sim \gamma} \left[\sum_{t \in I_r} \mathbf{k}_t^{(c)}(\mathbf{x}, \mathbf{y}) \geq 8dr + s, (\mathbf{x}, \mathbf{y}) \in X_c \times Y_c, \mathbf{d} > \binom{r}{2} \cdot 4d \right] \\ & \hspace{15em} \text{(if } (\mathbf{x}, \mathbf{y}) \in X_c \times Y_c \text{ then } \mathbf{k}_t^{(c)} = \mathbf{k}_t) \\ &\leq \Pr_{\mathbf{x}, \mathbf{y} \sim \gamma} \left[\sum_{t \in I_r} \mathbf{k}_t^{(c)}(\mathbf{x}, \mathbf{y}) \geq 8dr + s \right] \leq e^{-s/4}. \end{aligned}$$

Note that the inequality gives us an exponential tail on (5.8):

$$(5.8) \leq e^{-s/4} \cdot \left(\Pr_{\mathbf{x}, \mathbf{y} \sim \gamma} \left[(\mathbf{x}, \mathbf{y}) \in X_c \times Y_c, \mathbf{d} > \binom{r}{2} \cdot 4d \right] \right)^{-1}.$$

We can now integrate the above inequality to get an upper bound on the expected value of $\sum_{t \in I_r} \mathbf{k}_t$ under the distribution of interest. In particular, since for any non-negative random variable \mathbf{w} , the following holds for any parameter $\alpha \geq 0$:

$$\mathbb{E}[\mathbf{w}] = \int_0^{+\infty} \Pr[\mathbf{w} \geq z] dz \leq \alpha + \int_\alpha^{+\infty} \Pr[\mathbf{w} \geq z] dz = \alpha + \int_0^{+\infty} \Pr[\mathbf{w} \geq \alpha + z] dz,$$

we derive the following by taking $\alpha = 8dr + 4 \ln \left(\frac{1}{\Pr_{\mathbf{x}, \mathbf{y} \sim \gamma} [(\mathbf{x}, \mathbf{y}) \in X_c \times Y_c, \mathbf{d} > \binom{r}{2} \cdot 4d]} \right)$:

$$\begin{aligned} & \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \gamma} \left[\sum_{i \in I_r} \mathbf{k}_i(\mathbf{x}, \mathbf{y}) \mid \mathbf{d} > \binom{r}{2} \cdot 4d, (\mathbf{x}, \mathbf{y}) \in X_c \times Y_c \right] \\ &\leq \alpha + \int_0^{+\infty} e^{-z/4} dz = \alpha + 4 \\ &\leq 12dr + 4 \ln \left(\frac{1}{\Pr_{\mathbf{x}, \mathbf{y} \sim \gamma} [(\mathbf{x}, \mathbf{y}) \in X_c \times Y_c, \mathbf{d} > \binom{r}{2} \cdot 4d]} \right). \end{aligned}$$

This completes the proof of Claim 5.7. □

5.4.2 Projection on the Orthogonal Subspaces H_A^\perp and H_B^\perp

We shall show that the expected norm of the final center of mass when projected on the subspaces H_A^\perp and H_B^\perp is

$$\mathbb{E}[\mathbf{q}_A + \mathbf{q}_B] = O(d).$$

Recall that $\mathbf{q}_A = \left\| \Pi_{H_A^\perp} \mathbf{u}^{(d)} \right\|^2$ where H_A is the (random) linear subspace spanned by the orthonormal set of vectors $\mathbf{a}^{(0)}, \dots, \mathbf{a}^{(d)}$ and H_A^\perp its orthogonal complement. Moreover, the vectors $\mathbf{a}^{(t)}$ are determined by the previous Boolean and real communication. A similar statement holds for \mathbf{q}_B and the vectors $\mathbf{b}^{(t)}$ as well.

The proof will follow in two steps. We will first show that one can bound the norm of the projection $\Pi_{H_A^\perp} \mathbf{u}^{(d)}$, which turns out to be the Gaussian center of mass of a set that lives in the subspace H_A^\perp , in terms of the logarithm of the inverse relative measure with respect to the subspace. Note that the Gaussian measure here is the Gaussian measure $\gamma_{H_A^\perp}$ on the subspace H_A^\perp . The case for $\Pi_{H_B^\perp} \mathbf{u}^{(d)}$ will be similar. The second step will use information theory-esque convexity argument to show that on average the logarithm of the inverse relative measure is small.

For the first part, we observe that if we sample $\mathbf{x}, \mathbf{y} \sim \gamma$ and take a random walk on this protocol tree, we obtain a probability measure over transcripts which includes both real and Boolean values. Recall that the Boolean transcript is determined by the original protocol and only depends on the previous Boolean communication and the real transcript is sandwiched between the Boolean communication. Let $\ell = (\mathbf{c}, \mathbf{r})$ denote the random variable representing the full transcript of the generalized protocol where \mathbf{c} is the Boolean communication and \mathbf{r} is the real communication. For any given transcript ℓ , let $\mathbf{X}_\ell \times \mathbf{Y}_\ell$ denote the corresponding rectangle consists of inputs reaching the leaf, and let $\mathbf{X}_\mathbf{c} \times \mathbf{Y}_\mathbf{c}$ (for $\mathbf{X}_\mathbf{c}, \mathbf{Y}_\mathbf{c} \subseteq \mathbb{R}^n$) denote the rectangle consisting of all pairs of inputs to Alice and Bob that result in the Boolean transcript \mathbf{c} . Note that the real communication \mathbf{r} together with \mathbf{c} fixes the subspaces H_A and H_B and particular affine shifts \mathbf{s}_A and \mathbf{s}_B of those subspaces depending on the value of the inner products determined by the full transcript. In particular, the rectangle $\mathbf{X}_\ell \times \mathbf{Y}_\ell$ consistent with the full transcript $\ell = (\mathbf{c}, \mathbf{r})$ is given by $\mathbf{X}_\ell = \mathbf{X}_\mathbf{c} \cap (H_A + \mathbf{s}_A)$ and $\mathbf{Y}_\ell = \mathbf{Y}_\mathbf{c} \cap (H_B + \mathbf{s}_B)$, i.e., taking (random) affine slices of the original sets.

Note that $\mathbf{u}^{(d)}$ and $\mathbf{v}^{(d)}$ are distributed as the center of masses of the final rectangle $\mathbf{X}_\ell \times \mathbf{Y}_\ell$, and thus it suffices to look at the rectangles for the rest of the argument. Since \mathbf{X}_ℓ (resp., \mathbf{Y}_ℓ) lies in some affine shift of H_A^\perp (resp., H_B^\perp), defining the relative center of mass for a set A that lives in the ambient linear subspace V , as $\mu_V(A) = \mathbb{E}_{\mathbf{x} \sim \gamma_V}[\mathbf{x} \mid \mathbf{x} \in A]$ where the Gaussian measure γ_V is on the subspace V , it follows that

$$\mathbb{E}[\mathbf{q}_A + \mathbf{q}_B] = \mathbb{E} \left[\left\| \Pi_{H_A^\perp} \mathbf{u}^{(d)} \right\|^2 + \left\| \Pi_{H_B^\perp} \mathbf{u}^{(d)} \right\|^2 \right] = \mathbb{E}_\ell \left[\left\| \mu_{H_A^\perp}(\Pi_{H_A^\perp} \mathbf{X}_\ell) \right\|^2 + \left\| \mu_{H_B^\perp}(\Pi_{H_B^\perp} \mathbf{Y}_\ell) \right\|^2 \right].$$

Recalling that γ_{rel} is the Gaussian measure of a set relative to its ambient space, we will show:

Claim 5.8. $\left\| \mu_{H_A^\perp}(\Pi_{H_A^\perp} \mathbf{X}_\ell) \right\|^2 \leq 2e^2 \ln \left(\frac{e}{\gamma_{\text{rel}}(\mathbf{X}_\ell)} \right)$ and $\left\| \mu_{H_B^\perp}(\Pi_{H_B^\perp} \mathbf{Y}_\ell) \right\|^2 \leq 2e^2 \ln \left(\frac{e}{\gamma_{\text{rel}}(\mathbf{Y}_\ell)} \right)$.

Note that we can ignore the case when $\gamma_{\text{rel}}(\mathbf{X}_\ell)$ is zero above, since we will eventually take an expectation over ℓ and almost surely this measure is non-zero.

Using the previous claim,

$$\mathbb{E}[\mathbf{q}_A + \mathbf{q}_B] = \mathbb{E} \left[\left\| \Pi_{H_A^\perp} \mathbf{u}^{(d)} \right\|^2 + \left\| \Pi_{H_B^\perp} \mathbf{u}^{(d)} \right\|^2 \right] \leq 2e^2 \cdot \mathbb{E}_\ell \left[\ln \left(\frac{e}{\gamma_{\text{rel}}(\mathbf{X}_\ell \times \mathbf{Y}_\ell)} \right) \right].$$

For the second step of the proof, we show the next claim which relies on convexity arguments to bound the right hand side above by $O(d)$. This is similar in spirit to chain-style arguments from information theory.

Claim 5.9. $\mathbb{E}_\ell \left[\ln \left(\frac{e}{\gamma_{\text{rel}}(\mathbf{X}_\ell \times \mathbf{Y}_\ell)} \right) \right] = O(d)$.

This gives us the final bound $\mathbb{E}[\mathbf{q}_A + \mathbf{q}_B] = O(d)$ assuming the claims which we now prove.

Proof of Claim 5.8. We can bound the norm of the above projection by an application of the Gaussian level-one inequality (Theorem 3.1), which, by rotational symmetry, implies that if A is a subset of a linear subspace V with non-zero measure, then

$$\|\mu_V(A)\|^2 \leq 2e^2 \ln \left(\frac{e}{\gamma_V(A)} \right), \quad (5.9)$$

where recall that $\mu_V(A) = \mathbb{E}_{\mathbf{x} \sim \gamma_V}[\mathbf{x} \mid \mathbf{x} \in A]$ is the center of mass with respect to the Gaussian measure γ_V on the subspace V .

If we run the generalized protocol on $\mathbf{x}, \mathbf{y} \sim \gamma$ and condition on getting the full transcript ℓ , the conditional probability measure on $\Pi_{\mathbf{H}_A^\perp} \mathbf{x}$ is that of the Gaussian measure $\gamma_{\mathbf{H}_A^\perp}$ conditioned on $\mathbf{x} \in \mathbf{X}_\ell - \mathbf{s}_A$ and $\Pi_{\mathbf{H}_B^\perp} \mathbf{y}$ is that of the Gaussian measure $\gamma_{\mathbf{H}_B^\perp}$ conditioned on $\mathbf{y} \in \mathbf{Y}_\ell - \mathbf{s}_B$ and they are independent. This follows from the fact that so far the parties have fixed inner products along a basis for the orthogonal subspaces \mathbf{H}_A and \mathbf{H}_B and the fact the projection of a standard Gaussian on orthogonal subspaces are independent.

Thus, applying (5.9), we have

$$\|\mu_{\mathbf{H}_A^\perp}(\Pi_{\mathbf{H}_A^\perp} \mathbf{X}_\ell)\|^2 \leq 2e^2 \ln \left(\frac{e}{\gamma_{\mathbf{H}_A^\perp}(\mathbf{X}_\ell - \mathbf{s}_A)} \right) = 2e^2 \ln \left(\frac{e}{\gamma_{\text{rel}}(\mathbf{X}_\ell)} \right),$$

where the last line follows since $\mathbf{H}_A + \mathbf{s}_A$ is the ambient space for \mathbf{X}_ℓ (this holds almost surely) and $\gamma_{\text{rel}}(S) = \gamma_V(S - t)$ if $V + t$ is the ambient space of S . A similar argument proves the bound on $\|\mu_{\mathbf{H}_B^\perp}(\Pi_{\mathbf{H}_B^\perp} \mathbf{Y}_\ell)\|^2$. \square

Proof of Claim 5.9. For this claim, it will be convenient to consider a different generalized protocol \mathcal{C}' that generates the same distribution on the leaves ℓ . In particular, since the Boolean messages in the generalized protocol $\bar{\mathcal{C}}$ only depend on the previous Boolean messages, one can first send all the Boolean messages \mathbf{c} , and then send all the real messages \mathbf{r} choosing them according to the protocol $\bar{\mathcal{C}}$ depending on the previous real messages and the (partial) Boolean transcript. Note that the protocol \mathcal{C}' generates the same distribution on the leaves ℓ when the inputs $\mathbf{x}, \mathbf{y} \sim \gamma_n$. In particular, the real communication only partitions¹¹ each rectangles $X_c \times Y_c$ that corresponds to the Boolean transcript c into affine slices.

For rest of the claim, we now work with the protocol \mathcal{C}' where the Boolean communication happens first. To prove the claim, we condition on a Boolean transcript $\mathbf{c} = c$ and by induction show that

$$\mathbb{E}_{\mathbf{r}} \left[\ln \left(\frac{e}{\gamma_{\text{rel}}(\mathbf{X}_{(c,\mathbf{r})} \times \mathbf{Y}_{(c,\mathbf{r})})} \right) \mid \mathbf{c} = c \right] \leq \ln \left(\frac{e}{\gamma_{\text{rel}}(X_c \times Y_c)} \right), \quad (5.10)$$

¹¹We remark that this protocol \mathcal{C}' suffices for proving this claim since we are looking only at the leaves. However, unlike Lemma 5.3, directly bounding the expected quadratic variation of the martingale corresponding to the protocol \mathcal{C}' is difficult.

where (c, r) is the full transcript and $X_c \times Y_c$ is the rectangle containing all the inputs such that Boolean transcript is c . Note that $\gamma_{\text{rel}}(X_c \times Y_c)$ is the probability of obtaining the Boolean transcript c since the ambient space of X_c and Y_c is \mathbb{R}^n .

Then, taking expectation over the Boolean transcript c ,

$$\begin{aligned} \mathbb{E}_{\ell} \left[\ln \left(\frac{e}{\gamma_{\text{rel}}(\mathbf{X}_{\ell} \times \mathbf{Y}_{\ell})} \right) \right] &\leq \mathbb{E}_c \left[\ln \left(\frac{e}{\gamma_{\text{rel}}(\mathbf{X}_c \times \mathbf{Y}_c)} \right) \right] \\ &= \sum_{c \in \{0,1\}^*, |c| \leq d} \Pr[c = c] \ln \left(\frac{e}{\Pr[c = c]} \right) \\ &\leq \ln(2e \cdot 2^d) = O(d), \end{aligned}$$

where the last line follows from concavity.

Induction. To complete the proof, we now show (5.10) by induction. For this, let us look at an intermediate step t in \mathcal{C}' where the Boolean communication is fixed to c and Alice and Bob have exchanged some real messages $r_{<t} := r_1, \dots, r_{t-1}$. Let the current rectangle be $X_{(c, r_{<t})} \times Y_{(c, r_{<t})}$ and it is Alice's turn to speak. Note that $X_{(c, r_{<t})}$ and $Y_{(c, r_{<t})}$ live in some affine subspaces at this point and in the current round, Alice sends the inner product of her input x with a vector $a^{(t)}$ that is determined by the previous messages and orthogonal to the ambient space of $X_{(c, r_{<t})}$. At this step, Bob's set $Y_{(c, r_{<t})}$ does not change at all. We shall show that in each step, the log of the inverse of the relative measure of the current rectangle does not increase on average over the next message:

$$\mathbb{E}_{r_{\leq t}} \left[\ln \left(\frac{e}{\gamma_{\text{rel}}(\mathbf{X}_{(c, r_{\leq t})})} \right) \middle| c = c, r_{<t} = r_{<t} \right] \leq \ln \left(\frac{e}{\gamma_{\text{rel}}(X_{(c, r_{<t})})} \right), \quad (5.11)$$

and an analogous statement holds when Bob speaks. Taking an expectation over $r_{<t}$, the above directly applies (5.10) by a straightforward backward induction:

$$\begin{aligned} \mathbb{E}_{r_{\leq t}} \left[\ln \left(\frac{e}{\gamma_{\text{rel}}(\mathbf{X}_{(c, r_{\leq t})} \times \mathbf{Y}_{(c, r_{\leq t})})} \right) \middle| c = c \right] &\leq \mathbb{E}_{r_{<t}} \left[\ln \left(\frac{e}{\gamma_{\text{rel}}(\mathbf{X}_{(c, r_{<t})} \times \mathbf{Y}_{(c, r_{<t})})} \right) \middle| c = c \right] \\ &\leq \dots \leq \ln \left(\frac{e}{\gamma_{\text{rel}}(X_c \times Y_c)} \right). \end{aligned}$$

To see (5.11), let us write $X := X_{(c, r_{<t})}$ for Alice's current set. Recall that since we have fixed the history, Alice has fixed inner product with some orthogonal directions $a^{(1)}, \dots, a^{(t-1)}$ and she has decided on the next direction $a := a^{(t)}$ along which she will send the next inner product. Thus, X lives in some fixed affine subspace $V^{\perp} + s$ where V is the span of $a^{(1)}, \dots, a^{(t-1)}$ and the next message $r := r_t = \langle x, a \rangle$. Moreover, conditioned on the history till this point, the conditional probability distribution on Alice's input $x \in \mathbb{R}^n$ can be described as follows: the projections corresponding to the non-zero vectors in the sequence $a^{(1)}, \dots, a^{(t-1)}$, i.e., the inner products $\langle x, a^{(\tau)} \rangle$ where $a^{(\tau)} \neq 0$ for $\tau < t$, are fixed according to the shift s , while the distribution on the orthogonal complement V^{\perp} is that of the Gaussian measure $\gamma_{V^{\perp}}$ on the subspace V^{\perp} after conditioning on the event that $x \in X - s$ (which lives in V^{\perp}). This uses that projections of a standard n -dimensional Gaussian in orthogonal directions are independent.

Let k be the dimension of V where $k < n$. Then, by doing a linear transformation, we may assume that $V^{\perp} = \mathbb{R}^{n-k}$ (and thus $X \subseteq \mathbb{R}^{n-k}$ and the shift s fixes the coordinates $n-k+1$ through

n) and $a = e_1$, i.e., in the current message Alice reveals the first coordinate of $\mathbf{x} \in \mathbb{R}^{n-k}$ where \mathbf{x} is sampled from γ_{n-k} conditioned on $\mathbf{x} \in X$. In this case, γ_{rel} in the left hand side of (5.11) is exactly $\gamma_{\text{rel}}(X \cap \{x_1 = r\})$ if Alice sends r as the message, while for the right hand side of (5.11), we have $\gamma_{\text{rel}}(X) = \gamma_{n-k}(X)$. Denoting by $d\mu_{x_1}$ the probability density function of x_1 , our statement boils down to showing

$$\int_{\mathbb{R}} \ln \left(\frac{e}{\gamma_{\text{rel}}(X \cap \{x_1 = r\})} \right) d\mu_{x_1}(r) \leq \ln \left(\frac{e}{\gamma_{n-k}(X)} \right).$$

We show the above by explicitly writing the probability density function $d\mu_{x_1}$. Denote by $d\gamma_{n-k}(x_1, \dots, x_{n-k})$ the standard Gaussian density function¹² in \mathbb{R}^{n-k} . The density function of the random vector \mathbf{x} sampled from γ_{n-k} conditioned on $\mathbf{x} \in X$, is given $\gamma_{n-k}(X)^{-1} \cdot d\gamma_{n-k}(x_1, \dots, x_{n-k})$ for $\mathbf{x} \in X$ and zero outside. Thus, we have

$$\begin{aligned} d\mu_{x_1}(r) &= \frac{\int_{X \cap \{x_1=r\}} d\gamma_{n-k}(x_1, \dots, x_{n-k})}{\gamma_{n-k}(X)} \\ &= d\gamma_1(r) \cdot \frac{\int_{X \cap \{x_1=r\}} d\gamma_{n-k-1}(x_2, \dots, x_{n-k})}{\gamma_{n-k}(X)} = d\gamma_1(r) \cdot \frac{\gamma_{\text{rel}}(X \cap \{x_1 = r\})}{\gamma_{n-k}(X)}. \end{aligned}$$

Then, by concavity, the left hand side of (5.11) is exactly given by

$$\begin{aligned} \int_{\mathbb{R}} \ln \left(\frac{e}{\gamma_{\text{rel}}(X \cap \{x_1 = r\})} \right) d\mu_{x_1}(r) &\leq \ln \left(\int_{\mathbb{R}} \frac{e}{\gamma_{\text{rel}}(X \cap \{x_1 = r\})} d\mu_{x_1}(r) \right) \\ &= \ln \left(\frac{e}{\gamma_{n-k}(X)} \int_{\mathbb{R}} d\gamma_1(r) \right) = \ln \left(\frac{e}{\gamma_{n-k}(X)} \right). \quad \square \end{aligned}$$

6 Level-Two Fourier Growth

In this section, we prove [Theorem 1.3](#) that $L_{1,2}(h) = O(d^{3/2} \log^3(n))$. Similar to the proof of level-one bound [Theorem 1.2](#), we start with a d -round communication protocol $\tilde{\mathcal{C}}$ over the Gaussian space as defined in [Section 4](#). Note that $\tilde{\mathcal{C}}$ in turn comes from the original Boolean communication protocol \mathcal{C} . Thus in the following we assume without loss of generality $d \leq n$.

Given the discussion in [Subsection 4.3](#), to bound the second-level Fourier growth, one can attempt to bound the expected quadratic variation of the martingale that results from the protocol $\tilde{\mathcal{C}}$ directly, but similar to the case of level-one it is hard to leverage cancellations here to prove the bound we aim for. So, starting from $\tilde{\mathcal{C}}$, we will define a communication protocol $\bar{\mathcal{C}}$ that computes the same function as $\tilde{\mathcal{C}}$, but satisfies some additional “clean” property where it is easier to control the quadratic variation. This new protocol will differ from $\tilde{\mathcal{C}}$ in two ways. Firstly, the protocol $\bar{\mathcal{C}}$ will consist of additional “cleanup steps” where Alice and Bob reveal certain *quadratic forms* of their input. Secondly, the protocol $\bar{\mathcal{C}}$ will send the real value of the quadratic form *with certain precision*. Note that this protocol will not involve sending real messages at all, instead, any potential real messages will be truncated to a few bits of precision and be sent as Boolean messages.

We emphasize that the main difference in the protocol $\bar{\mathcal{C}}$ from the corresponding level-one variant comes from the precision control, which is not needed there due to the fact that Gaussian distribution remains a (lower-dimensional) Gaussian under linear projections. For technical reasons we shall also need to analyze the martingale under a truncated Gaussian distribution, where all

¹²Explicitly $d\gamma_m(x_1, \dots, x_m) = \prod_{i=1}^m d\gamma_1(x_i)$ where $d\gamma_1(r) = \frac{1}{\sqrt{2\pi}} e^{-r^2/2}$ is the density function for one-dimensional standard Gaussian.

coordinates are bounded in some large interval $[-T, T]$. This intuitively doesn't incur a noticeable difference on the distribution since it is highly unlikely that coordinates drawn from Gaussian distribution will be outside such intervals and recalling [Remark 4.2](#) and [Proposition 4.4](#), it still suffices to analyze the corresponding martingale under the truncated Gaussian distribution.

We next define the notion of a 4-wise clean protocol.

6.1 4-Wise Clean Protocols

Consider an intermediate node in the protocol and let $X \subseteq \mathbb{R}^n$ refer to the set of Alice's inputs reaching this node. We denote by $\mathbb{S}^{n \times n-1}$ the set of all matrices in $\mathbb{R}^{n \times n}$ with zero diagonal and unit norm (when viewed as n^2 -dimensional vectors). For a parameter $\lambda > 0$, we say that the set X is *4-wise clean in a direction* $a \in \mathbb{S}^{n \times n-1}$ if

$$\mathbb{E}_{\mathbf{x} \sim \gamma} \left[\left\langle \mathbf{x} \dot{\otimes} \mathbf{x} - \sigma(X), a \right\rangle^2 \mid \mathbf{x} \in X \right] < \lambda,$$

where we recall that $\sigma(X) = \mathbb{E}_{\mathbf{x} \sim \gamma} \left[\mathbf{x} \dot{\otimes} \mathbf{x} \mid \mathbf{x} \in X \right]$ is the level-two center of mass of X under the Gaussian measure. We say that the set X is *4-wise clean* if it is 4-wise clean in *every direction* a . Our new protocol will consist of the original protocol, interspersed by cleaning steps. Once Alice sends her bit as in the original protocol, she cleans X by revealing $\langle \mathbf{x} \dot{\otimes} \mathbf{x}, a \rangle$ with a few bits of precision while there exists direction $a \in \mathbb{S}^{n \times n-1}$ such that X not clean in direction a . Once X becomes clean, Alice proceeds to the next round and Bob does an analogous cleanup. We now describe this formally.

Communication with Finite Precision. Let positive integer L be a precision parameter that we will use for truncation. In our new communication protocol, we will send real numbers with precision 2^{-L} . This is formalized as the $\text{trunc}_L(z)$ function defined at $z \in \mathbb{R}$ as

$$\text{trunc}_L(z) = \lfloor z \cdot 2^L \rfloor / 2^L.$$

Construct $\bar{\mathcal{C}}$ from $\tilde{\mathcal{C}}$. As described before, $\bar{\mathcal{C}}$ will consist of the original protocol along with extra steps where Alice or Bob reveal the (approximate) value of a quadratic form on their input. Consider an intermediate node of this new protocol at depth t . We always use the random variable $\mathbf{X}^{(t)}$ (resp., $\mathbf{Y}^{(t)}$) to denote the set of inputs of Alice (resp., Bob) reaching the node. If Alice is revealing a quadratic form in this step, we use $\mathbf{a}^{(t)}$ to denote the matrix of the quadratic form revealed at this node, otherwise set $\mathbf{a}^{(t)}$ to be the all-zeroes matrix. We define $\mathbf{b}^{(t)}$ similarly for Bob. Throughout the protocol, we will always set $\mathbf{u}^{(t)}$ and $\mathbf{v}^{(t)}$ to denote $\sigma(\mathbf{X}^{(t)})$ and $\sigma(\mathbf{Y}^{(t)})$ respectively.

Recall that $\lambda > 0$ is the parameter for cleanup to be optimized later. Since we will now send real numbers (with certain precision) as bit-strings, their magnitudes should also be well controlled to guarantee bounded message length. This is managed by a parameter $T > 0$ and we will restrict the inputs to the parties in $\bar{\mathcal{C}}$ to come from the box $[-T, T]^n$. Note that, by Gaussian concentration, $T = \Theta\left(\sqrt{\log(n)}\right)$ suffices.

1. At the beginning, Alice receives an input $x \in [-T, T]^n$ and Bob receives an input $y \in [-T, T]^n$.
2. We initialize $t \leftarrow 0$, $\mathbf{X}^{(0)}, \mathbf{Y}^{(0)} \leftarrow [-T, T]^n$, and $\mathbf{a}^{(0)}, \mathbf{b}^{(0)} \leftarrow 0^{n \times n}$.

3. For each phase $i = 1, 2, \dots, d$: suppose we are starting the cleanup for a node at depth i in the original protocol $\tilde{\mathcal{C}}$ and suppose we are at a node of depth t in the new protocol $\bar{\mathcal{C}}$. If it is Alice's turn to speak in $\bar{\mathcal{C}}$:

- (a) **Orthogonalization by revealing the correlation with Bob's center of mass.**
 Alice begins by revealing the inner product of her input x with Bob's current (signed) level-two center of mass $\eta \odot \mathbf{v}^{(t)}$. Since in the previous steps, she has already revealed the inner product with Bob's previous centers of mass, for technical reasons, we will only have Alice announce the inner product with the component of $\eta \odot \mathbf{v}^{(t)}$ that is orthogonal to the previous directions along which Alice announced the inner product. More formally, let $\mathbf{a}^{(t+1)}$ be the component of $\eta \odot \mathbf{v}^{(t)}$ that is orthonormal to the span of the previous directions $\mathbf{a}^{(\tau)}$ for $\tau \leq t$, i.e.,

$$\mathbf{a}^{(t+1)} = \text{unit} \left(\eta \odot \mathbf{v}^{(t)} - \sum_{\tau=1}^t \langle \eta \odot \mathbf{v}^{(t)}, \mathbf{a}^{(\tau)} \rangle \cdot \mathbf{a}^{(\tau)} \right).$$

Alice computes $\bar{\mathbf{c}}^{(t+1)} \leftarrow \text{trunc}_L \left(\left\langle x \dot{\otimes} x, \mathbf{a}^{(t+1)} \right\rangle \right)$ and sends $\bar{\mathbf{c}}^{(t+1)}$ to Bob. Set $\mathbf{b}^{(t+1)} \leftarrow 0^{n \times n}$. Increment t by 1 and go to step (b).

- (b) **Original communication.** Alice sends the bit $\bar{\mathbf{c}}^{(t+1)}$ that she was supposed to send in $\tilde{\mathcal{C}}$ based on previous messages and x . Set $\mathbf{a}^{(t+1)}, \mathbf{b}^{(t+1)} \leftarrow 0^{n \times n}$. Increment t by 1 and go to step (c).
- (c) **Cleanup steps.** While there exists some direction $a \in \mathbb{S}^{n \times n - 1}$ orthogonal to previous directions, i.e., $\langle a, \mathbf{a}^{(\tau)} \rangle = 0$ for all $\tau \leq t$, and $\mathbf{X}^{(t)}$ is *not 4-wise clean* in direction a , Alice computes $\bar{\mathbf{c}}^{(t+1)} \leftarrow \text{trunc}_L \left(\left\langle x \dot{\otimes} x, a \right\rangle \right)$ and sends $\bar{\mathbf{c}}^{(t+1)}$ to Bob. Set $\mathbf{a}^{(t+1)} \leftarrow a$ and $\mathbf{b}^{(t+1)} \leftarrow 0^{n \times n}$. Increment t by 1. Repeat step (c) while $\mathbf{X}^{(t)}$ is not 4-wise clean; otherwise, increment i by 1 and go back to the for-loop in step 3 which starts a new phase.

If it is Bob's turn to speak, we define everything similarly with the role of $x, \mathbf{a}, \mathbf{X}, \mathbf{u}$ switched with $y, \mathbf{b}, \mathbf{Y}, \mathbf{v}$.

4. Finally at the end of the protocol, the value $\bar{\mathcal{C}}(x, y)$ is determined based on all the previous communication and the corresponding output it defines in $\tilde{\mathcal{C}}$.

Remark 6.1. Note that by construction, the non-zero matrices among $\mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \dots$ form an orthonormal set when viewed as n^2 -dimensional vectors (similarly for $\mathbf{b}^{(1)}, \mathbf{b}^{(2)}, \dots$) and moreover, their diagonals are zero. Lastly, $\mathbf{a}^{(t)}$ and $\mathbf{b}^{(t)}$ are known to both Alice and Bob as they are canonically determined by previous messages.

We remark that the steps 3(a), 3(b), and 3(c) always occur in sequence for each party and we refer to such a sequence of steps as a *phase* for that party. Note that there are at most d phases. If a new phase starts at time t , then the current rectangle $\mathbf{X}^{(t)} \times \mathbf{Y}^{(t)}$ is 4-wise clean for both parties by construction.

Now we formalize a few useful properties regarding the communication protocol $\bar{\mathcal{C}}$. The first fact below follows since each $\mathbf{u}^{(t)}$ is an expectation of $\mathbf{x} \dot{\otimes} \mathbf{x}$ over some distribution and $\mathbf{x} \dot{\otimes} \mathbf{x}$ has zero diagonal.

Fact 6.2. $\mathbf{u}^{(0)} = \mathbf{v}^{(0)} = 0^{n \times n}$ and each $\mathbf{u}^{(t)}, \mathbf{v}^{(t)}$ has zero diagonal.

The following follows from tail bounds for the univariate standard normal distribution.

Fact 6.3. Let $\gamma^* = \gamma(\mathbf{X}^{(0)}) \cdot \gamma(\mathbf{Y}^{(0)})$. Then $\gamma^* \geq 1 - O\left(n \cdot e^{-T^2/2}\right)$.

The next fact says that when a node fixes a quadratic form with 2^{-L} precision, for any two inputs that reach this node, the quadratic forms differ by at most 2^{-L} .

Fact 6.4. In step 3(a) and 3(c), any $x, x' \in \mathbf{X}^{(t+1)}$ satisfies $\left| \langle x \dot{\otimes} x, \mathbf{a}^{(t+1)} \rangle - \langle x' \dot{\otimes} x', \mathbf{a}^{(t+1)} \rangle \right| < 2^{-L}$. Similarly any $y, y' \in \mathbf{Y}^{(t+1)}$ satisfies $\left| \langle y \dot{\otimes} y, \mathbf{b}^{(t+1)} \rangle - \langle y' \dot{\otimes} y', \mathbf{b}^{(t+1)} \rangle \right| < 2^{-L}$.

The next claim bounds the maximum attainable norms for Alice and Bob's level-two center of masses at any point in the protocol. This uses the fact that the inputs come from the truncated Gaussian distribution.

Claim 6.5. $\|\mathbf{u}^{(t)}\| = \|\eta \odot \mathbf{u}^{(t)}\| < nT$ and $\|\mathbf{v}^{(t)}\| = \|\eta \odot \mathbf{v}^{(t)}\| < nT$ for all possible t and $\mathbf{u}^{(t)}, \mathbf{v}^{(t)}$ throughout the communication.

Proof. Since η is a matrix with zero diagonal and $\{\pm 1\}$ entries off diagonal and $\mathbf{u}^{(t)}$ also has zero diagonal, $\|\mathbf{u}^{(t)}\| = \|\eta \odot \mathbf{u}^{(t)}\|$. In addition, since $\mathbf{X}^{(t)} \subseteq \mathbf{X}^{(0)} = [-T, T]^n$, we have

$$\|\mathbf{u}^{(t)}\| \leq \mathbb{E}_{\mathbf{x} \sim \gamma} \left[\left\| \left(\mathbf{x} \dot{\otimes} \mathbf{x} \right) \right\| \mid \mathbf{x} \in \mathbf{X}^{(t)} \right] \leq \sqrt{(n^2 - n) \cdot T^2} < nT.$$

A similar analysis works for $\mathbf{v}^{(t)}$. □

The next claim gives a bound on the length of any message in the protocol $\bar{\mathcal{C}}$.

Claim 6.6. For any $x \in \mathbf{X}^{(0)}$ and $y \in \mathbf{Y}^{(0)}$, any message in $\bar{\mathcal{C}}(x, y)$ consists of at most $L + \log(Tn)$ many bits.

Proof. Assume without loss of generality it is Alice's turn to speak. On step 3(b) she sends one bits. On steps 3(a) and 3(c), she computes $\text{trunc}_L(\langle x \dot{\otimes} x, a \rangle)$ for some $a \in \mathbb{S}^{n \times n-1}$ and send the result. Since

$$\left| \langle x \dot{\otimes} x, a \rangle \right| \leq \left\| x \dot{\otimes} x \right\| \cdot \|a\| \leq \sqrt{(n^2 - n) \cdot T^2} < nT,$$

and the message is a multiple of 2^{-L} that means trunc_L yields a message with $L + \log(nT)$ many bits. □

The last claim bounds the maximum depth of the new protocol $\bar{\mathcal{C}}$.

Claim 6.7. Let ℓ be an arbitrary leaf of the protocol $\bar{\mathcal{C}}$ and $D(\ell)$ be its depth. Then $D(\ell) \leq 2n^2$. Moreover, along this path there are at most $n^2 - n$ many non-zero $\mathbf{a}^{(t)}$ and at most $n^2 - n$ many non-zero $\mathbf{b}^{(t)}$ for $t \in \{1, \dots, D(\ell)\}$.

Proof. We count the number of communication steps separately:

- **Steps 3(a) and 3(b).** Steps 3(a) and 3(b) occur once in every phase, thus at most d times.
- **Step 3(c).** For Alice, each time she communicates at step 3(c), the direction $a \in \mathbb{S}^{n \times n-1}$ is non-zero and orthogonal to all previous $\mathbf{a}^{(t)}$'s. Since the dimension of $\mathbb{S}^{n \times n-1}$ is $n^2 - n$, this happens at most $n^2 - n$ times. Similar argument works for Bob.

Thus in total we have at most $2(n^2 - n) + 2d \leq 2n^2$ steps. □

We will eventually show that, with suitable choice of λ, T, L , typically $D(\ell)$ is at most $d \cdot \text{polylog}(n)$.

6.2 Bounding the Expected Quadratic Variation

Consider the martingale process defined in (4.5) from a random walk on the protocol tree generated by $\bar{\mathcal{C}}$ where the inputs \mathbf{x}, \mathbf{y} are sampled from γ_n conditioned on being in the bounded cube $[-T, T]^n$. Recall that Proposition 4.3 still holds (see Remark 4.5).

Formally, at time t the process is defined by

$$\mathbf{z}_2^{(t)} = \left\langle \mathbf{u}^{(t)}, \eta \odot \mathbf{v}^{(t)} \right\rangle,$$

where we recall that $\mathbf{u}^{(t)} = \sigma(\mathbf{X}^{(t)})$ and $\mathbf{v}^{(t)} = \sigma(\mathbf{Y}^{(t)})$ and η is a fixed sign matrix with a zero diagonal. The martingale process stops once it hits a leaf of $\bar{\mathcal{C}}$. Let \mathbf{d} denote the (stopping) time when this happens. Note that $\mathbb{E}[\mathbf{d}]$ is exactly the expected depth of the protocol $\bar{\mathcal{C}}$.

In light of Remark 4.2 and Proposition 4.4, to prove Theorem 1.3, it suffices to prove the following.

Lemma 6.8. $\mathbb{E} \left[\sum_{t=1}^{\mathbf{d}} \left(\Delta \mathbf{z}_2^{(t)} \right)^2 \right] = O \left(d^3 \log^6(n) \right).$

Lemma 6.8 is proved in three steps. We first show that essentially the only change in the value of the martingale is the orthogonalization step 3(a). The reason is the same as the level-one bound: Alice's messages sent in step 3(b) and 3(c) are always near-orthogonal to Bob's current level-two center of mass, thus they do not change the value of the martingale $\mathbf{z}_2^{(t)}$ much. Moreover, by level-two analog of (2.2), since Alice's current node was clean just before Alice sent the message in step 3(a), the expected change $\mathbb{E} \left[\left(\Delta \mathbf{z}_2^{(t+1)} \right)^2 \right]$ can be bounded in terms of the squared norm of the change that occurred in $\mathbf{u}^{(t)}$ (or $\mathbf{v}^{(t)}$) between the current round and the last round where Alice was in step 3(a). Similar argument works for Bob.

Formally, this is encapsulated by the next lemma for which we need some additional definitions. Let $(\mathcal{F}^{(t)})_t$ denote the natural filtration induced by the random walk on the generalized protocol tree with respect to which $\mathbf{z}_2^{(t)}$ is a Doob martingale and also $\mathbf{u}^{(t)}, \mathbf{v}^{(t)}$ form vector-valued martingales (recall Proposition 4.3). Note that $\mathcal{F}^{(t)}$ fixes all the rectangles encountered during times $0, \dots, t$ and thus for $\tau \leq t$, the random variables $\mathbf{u}^{(\tau)}, \mathbf{v}^{(\tau)}, \mathbf{z}_2^{(\tau)}$ are determined, in particular, they are $\mathcal{F}^{(t)}$ -measurable. Recalling that λ is the cleanup parameter to be optimized later, we then have the following. Below we assume without any loss of generality that Alice speaks first and, in particular, we note that Alice speaks in step 3(a) for the first time at time zero when both Alice and Bob's center of masses are at zero: $\mathbf{u}^{(0)} = \mathbf{v}^{(0)} = 0$.

Lemma 6.9 (Step Size). *Let $0 = \tau_1 < \tau_2 < \dots \leq \mathbf{d}$ be a sequence of stopping times with τ_m being the index of the round where Alice speaks in step 3(a) for the m^{th} time or \mathbf{d} if there is no such round. Then, for any integer $m \geq 2$,*

$$\mathbb{E} \left[\left(\Delta \mathbf{z}_2^{(\tau_{m+1})} \right)^2 \middle| \mathcal{F}^{(\tau_m)} \right] \leq \lambda \cdot \left\| \mathbf{v}^{(\tau_m)} - \mathbf{v}^{(\tau_{m-1})} \right\|^2 + 16n^7 T^3 \cdot 2^{-L}.$$

and moreover, for any $t \in \mathbb{N}$, we have that

$$\mathbb{E} \left[\left(\Delta \mathbf{z}_2^{(t+1)} \right)^2 \middle| \mathcal{F}^{(t)}, \tau_{m-1} < t < \tau_m, \text{ Alice speaks at time } t \right] \leq 4n^6 T^2 \cdot 2^{-2L}$$

A similar statement also holds if Bob speaks where \mathbf{v} is replaced by \mathbf{u} and the sequence (τ_m) is replaced by (τ'_m) where τ'_m is the index of the round where Bob speaks in step 3(a) for the m^{th} time or \mathbf{d} if there is no such round.

We indeed see that, if $L = \Omega(\log(n))$ and $T = O(\sqrt{\log(n)})$, then $\text{poly}(T, n) \cdot 2^{-L} = o(1)$, and steps 3(b) and 3(c) do not contribute much to the quadratic variation and only the steps 3(a) do. Also, since the first time Alice and Bob speak, they start in step 3(a), we also note that $\mathbf{u}^{(\tau_1)}$ and $\mathbf{v}^{(\tau'_1)}$ are their initial centers of mass which are both zero.

We shall prove the above lemma in [Subsection 6.3](#) and continue with the bound on the quadratic variation here. Using the bounds on the step sizes from [Lemma 6.9](#),

$$\begin{aligned} \mathbb{E} \left[\sum_{t=1}^d \left(\Delta z_2^{(t)} \right)^2 \right] &\leq \lambda \cdot \mathbb{E} \left[\sum_{m \geq 2} \left\| \mathbf{v}^{(\tau_m)} - \mathbf{v}^{(\tau_{m-1})} \right\|^2 + \left\| \mathbf{u}^{(\tau'_m)} - \mathbf{u}^{(\tau'_{m-1})} \right\|^2 \right] + 16n^7 T^3 \cdot 2^{-L} \cdot \mathbb{E}[d] \\ &\leq \lambda \cdot \mathbb{E} \left[\sum_{m \geq 2} \left\| \mathbf{v}^{(\tau_m)} - \mathbf{v}^{(\tau_{m-1})} \right\|^2 + \left\| \mathbf{u}^{(\tau'_m)} - \mathbf{u}^{(\tau'_{m-1})} \right\|^2 \right] + 16n^7 T^3 \cdot 2^{-L} \cdot 2n^2. \end{aligned} \quad (\text{by Claim 6.7})$$

On the other hand, using the orthogonality of vector-valued martingale differences from [\(3.2\)](#),

$$\mathbb{E} \left[\sum_{m \geq 2} \left\| \mathbf{v}^{(\tau_m)} - \mathbf{v}^{(\tau_{m-1})} \right\|^2 \right] = \mathbb{E} \left[\left\| \mathbf{v}^{(d)} \right\|^2 \right].$$

A similar statement holds for $(\mathbf{u}^{(t)})$ as well. Therefore,

$$\mathbb{E} \left[\sum_{t=1}^d \left(\Delta z_2^{(t)} \right)^2 \right] \leq \lambda \cdot \left(\mathbb{E} \left[\left\| \mathbf{u}^{(d)} \right\|^2 \right] + \mathbb{E} \left[\left\| \mathbf{v}^{(d)} \right\|^2 \right] \right) + 64n^9 T^3 \cdot 2^{-L}. \quad (6.1)$$

Then in [Subsection 6.4](#) we will apply level-two inequalities (see [Theorem 3.1](#)) to convert the bounding $\mathbb{E} \left[\left\| \mathbf{u}^{(d)} \right\|^2 + \left\| \mathbf{v}^{(d)} \right\|^2 \right]$ into bounding the second moment $\mathbb{E}[d^2]$. This reduction is formalized as [Lemma 6.10](#) below and its proof is similar to [[GRT21](#), Claim 1].

For each leaf ℓ , let $\gamma(\ell) = \gamma(\mathbf{X}^{(D(\ell))}) \cdot \gamma(\mathbf{Y}^{(D(\ell))})$ be the Gaussian measure of the rectangle at ℓ . Recall $\gamma^* = \gamma(\mathbf{X}^{(0)}) \times \gamma(\mathbf{Y}^{(0)})$.

Lemma 6.10. $\mathbb{E} \left[\left\| \mathbf{u}^{(d)} \right\|^2 + \left\| \mathbf{v}^{(d)} \right\|^2 \right] \leq O \left(\frac{1}{\gamma^*} + L^2 \mathbb{E}[d^2] \right)$.

Finally, in [Subsection 6.5](#), we bound the second moment $\mathbb{E}[d^2]$ for a suitable choice of parameters.

Lemma 6.11. *It holds that $\mathbb{E}[d^2] = O(d^2)$ and $\gamma^* \geq \frac{3}{4}$ for $L = \Theta(\log(n))$, $T = \Theta(\sqrt{\log(n)})$, and $\lambda = \Theta(d \log^4(n))$.*

Given [Lemmas 6.10](#) and [6.11](#), the proof of [Lemma 6.8](#) naturally follows.

Proof of Lemma 6.8. With the parameters chosen in [Lemma 6.11](#), we have

$$\begin{aligned} \mathbb{E} \left[\sum_{t=1}^d \left(\Delta z_2^{(t)} \right)^2 \right] &\leq O(d \log^4(n)) \cdot \left(\mathbb{E} \left[\left\| \mathbf{u}^{(d)} \right\|^2 \right] + \mathbb{E} \left[\left\| \mathbf{v}^{(d)} \right\|^2 \right] \right) + 1 && (\text{by (6.1)}) \\ &\leq O(d \log^4(n)) \cdot (1 + \log^2(n) \cdot \mathbb{E}[d^2]) + 1 && (\text{by Lemma 6.10}) \\ &\leq O(d \log^4(n)) \cdot (1 + \log^2(n) \cdot d^2) + 1 && (\text{by Lemma 6.11}) \\ &= O(d^3 \log^6(n)). && \square \end{aligned}$$

Remark 6.12. Note that our proof for level-two Fourier growth actually holds for a slightly more general setting, where Alice and Bob are allowed to send $O(L) = O(\log(n))$ bits during each original communication round. This can be viewed as balancing the length of the messages in step 3(b) with step 3(a) and step 3(c).

Since one can always convert a d -round 1-bit communication protocol into a $\frac{2d}{\log \log(n)}$ -round $\log(n)$ -bit communication protocol, we obtain a slightly better level-two Fourier growth bound of

$$O\left(\frac{d^{3/2} \log^3(n)}{(\log \log(n))^{3/2}}\right).$$

The conversion is done by Alice (resp., Bob) enumerating the next $\log \log(n)/2$ bits from Bob (resp., Alice), and providing the corresponding $\log \log(n)/2$ bits responses for each possibility.

It is also possible to improve the $\log^3(n)$ factor to $\log^2(n)$ by varying the cleanup parameter λ with depth. For example, for depth in the interval $[4rd, 4(r+1)d]$, one could pick $\lambda_r = \Theta(d \cdot \log^2(n) \cdot r^2)$. Since our focus is mostly on improving the polynomial dependence in d where there is still room for improvement, we do not make an effort here to improve the polylog terms.

6.3 Bounds on Step Sizes (Proof of Lemma 6.9)

Let us abbreviate $\tau = \tau_m$ and note that at time τ a new phase starts for Alice. By construction, this means that the current rectangle $\mathbf{X}^{(\tau)} \times \mathbf{Y}^{(\tau)}$ determined by $\mathcal{F}^{(\tau)}$ is 4-wise clean with parameter λ , and since Alice is in step 3(a) at the start of a new phase, $\mathbf{a}^{(\tau+1)}$ is chosen to be the (normalized) component of $\eta \odot \mathbf{v}^{(\tau)}$ that is orthogonal to previous directions $\mathbf{a}^{(1)}, \dots, \mathbf{a}^{(\tau)}$.

For each $r = 1, \dots, \tau + 1$, let $\beta^{(r)} := \langle \eta \odot \mathbf{v}^{(\tau)}, \mathbf{a}^{(r)} \rangle$ be the length of $\eta \odot \mathbf{v}^{(\tau)}$ along direction $\mathbf{a}^{(r)}$. Each $\beta^{(r)}$ is $\mathcal{F}^{(\tau)}$ -measurable (i.e., it is determined by $\mathcal{F}^{(\tau)}$) and $\eta \odot \mathbf{v}^{(\tau)} = \sum_{r \leq \tau+1} \beta^{(r)} \cdot \mathbf{a}^{(r)}$. In this case, we have

$$\begin{aligned} \mathbb{E} \left[\left(\Delta z_2^{(\tau+1)} \right)^2 \middle| \mathcal{F}^{(\tau)} \right] &= \mathbb{E} \left[\left\langle \mathbf{u}^{(\tau+1)} - \mathbf{u}^{(\tau)}, \eta \odot \mathbf{v}^{(\tau)} \right\rangle^2 \middle| \mathcal{F}^{(\tau)} \right] \\ &= \mathbb{E} \left[\left(\sum_{r=1}^{\tau+1} \beta^{(r)} \cdot \left\langle \mathbf{u}^{(\tau+1)} - \mathbf{u}^{(\tau)}, \mathbf{a}^{(r)} \right\rangle \right)^2 \middle| \mathcal{F}^{(\tau)} \right]. \end{aligned} \quad (6.2)$$

Similar to the level-one proof, the components of $\mathbf{u}^{(\tau+1)}$ and $\mathbf{u}^{(\tau)}$ are roughly the same along any of the previous directions $\mathbf{a}^{(1)}, \dots, \mathbf{a}^{(\tau)}$ and so they almost cancel out and the major quantity is in the direction $\mathbf{a}^{(\tau+1)}$. This follows since, in all the previous steps $r \leq \tau$, Alice has already fixed $\langle x \dot{\otimes} x, \mathbf{a}^{(r)} \rangle$ with precision 2^{-L} . This implies that for any $\mathbf{X}^{(\tau)}$ and $\mathbf{X}^{(\tau+1)}$ that are determined by $\mathcal{F}^{(\tau+1)}$, the inner product with all the previous $\mathbf{a}^{(1)}, \dots, \mathbf{a}^{(\tau)}$ is fixed with precision 2^{-L} over the choice of x . Formally, by Fact 6.4, we have that for any $x \in \mathbf{X}^{(\tau)}$ and $x' \in \mathbf{X}^{(\tau+1)}$, it holds that $\left| \langle x \dot{\otimes} x, \mathbf{a}^{(r)} \rangle - \langle x' \dot{\otimes} x', \mathbf{a}^{(r)} \rangle \right| \leq 2^{-L}$ for all $r \leq \tau$. In particular, since $\mathbf{u}^{(\tau)} = \sigma(\mathbf{X}^{(\tau)})$ and $\mathbf{u}^{(\tau+1)} = \sigma(\mathbf{X}^{(\tau+1)})$ are the corresponding centers of mass, we have that

$$\left| \left\langle \mathbf{u}^{(\tau+1)} - \mathbf{u}^{(\tau)}, \mathbf{a}^{(r)} \right\rangle \right| \leq 2^{-L} \quad \text{for all } r \leq \tau. \quad (6.3)$$

On the other hand, since $\mathbf{X}^{(\tau+1)} \subseteq \mathbf{X}^{(\tau)} \subseteq \mathbf{X}^{(0)} = [-T, T]^n$ and $\mathbf{a}^{(\tau+1)}$ is a unit direction, we have

$$\left| \left\langle \mathbf{u}^{(\tau+1)} - \mathbf{u}^{(\tau)}, \mathbf{a}^{(\tau+1)} \right\rangle \right| \leq \left\| \mathbf{u}^{(\tau+1)} - \mathbf{u}^{(\tau)} \right\| \leq 2nT. \quad (6.4)$$

Similarly, noting that η is a sign matrix, we can bound

$$|\beta^{(r)}| = \left| \left\langle \eta \odot \mathbf{v}^{(r)}, \mathbf{a}^{(r)} \right\rangle \right| \leq \left\| \eta \odot \mathbf{v}^{(r)} \right\| \leq \left\| \mathbf{v}^{(r)} \right\| \leq nT \quad \text{for all } r \leq \tau + 1. \quad (6.5)$$

Expanding the square in (6.2) and plugging these estimates to each one of the $(\tau + 1)^2$ terms gives

$$\begin{aligned} \mathbb{E} \left[\left(\Delta \mathbf{z}_2^{(\tau+1)} \right)^2 \middle| \mathcal{F}^{(\tau)} \right] &\leq \mathbb{E} \left[\left(\beta^{(\tau+1)} \right)^2 \left\langle \mathbf{u}^{(\tau+1)} - \mathbf{u}^{(\tau)}, \mathbf{a}^{(\tau+1)} \right\rangle^2 + ((\tau + 1)^2 - 1) \cdot \frac{2(nT)^3}{2^L} \middle| \mathcal{F}^{(\tau)} \right] \\ &\leq \left(\beta^{(\tau+1)} \right)^2 \mathbb{E} \left[\left\langle \mathbf{u}^{(\tau+1)} - \mathbf{u}^{(\tau)}, \mathbf{a}^{(\tau+1)} \right\rangle^2 \middle| \mathcal{F}^{(\tau)} \right] + 12n^7 T^3 \cdot 2^{-L}, \end{aligned} \quad (6.6)$$

where the second line follows from [Claim 6.7](#).

We now bound the term outside the expectation by the change in the center of mass $\mathbf{v}^{(\cdot)}$ and the term inside the expectation by the fact that the set is 4-wise clean.

Term Outside the Expectation. Recall that $\mathbf{a}^{(\tau+1)}$ is chosen to be the (normalized) component of $\eta \odot \mathbf{v}^{(\tau)}$ that is orthogonal to the span of $\mathbf{a}^{(1)}, \dots, \mathbf{a}^{(\tau)}$. Since $\eta \odot \mathbf{v}^{(\tau_{m-1})}$ is in the span of $\mathbf{a}^{(1)}, \dots, \mathbf{a}^{(\tau_{m-1}+1)}$ and $\tau_{m-1} + 1 \leq \tau = \tau_m$, it is orthogonal to $\mathbf{a}^{(\tau+1)}$. Hence

$$\beta^{(\tau+1)} = \left\langle \eta \odot \mathbf{v}^{(\tau)}, \mathbf{a}^{(\tau+1)} \right\rangle = \left\langle \eta \odot \left(\mathbf{v}^{(\tau)} - \mathbf{v}^{(\tau_{m-1})} \right), \mathbf{a}^{(\tau+1)} \right\rangle.$$

Since $\mathbf{a}^{(\tau+1)}$ is a unit direction and η is a sign matrix, this implies that

$$\left(\beta^{(\tau+1)} \right)^2 \leq \left\| \mathbf{v}^{(\tau)} - \mathbf{v}^{(\tau_{m-1})} \right\|^2. \quad (6.7)$$

Term Inside the Expectation. Recall that Alice is in step 3(a), she sends $\left\langle x \dot{\otimes} x, \mathbf{a}^{(\tau+1)} \right\rangle$ with precision 2^{-L} at time τ , and thus the same inner product with $\mathbf{a}^{(\tau+1)}$ is fixed with precision 2^{-L} for every point in $\mathbf{X}^{(\tau+1)}$ determined by $\mathcal{F}^{(\tau+1)}$. Thus

$$\begin{aligned} \left\langle \mathbf{u}^{(\tau+1)}, \mathbf{a}^{(\tau+1)} \right\rangle^2 &= \left(\mathbb{E}_{\mathbf{x} \sim \gamma} \left[\left\langle \mathbf{x} \dot{\otimes} \mathbf{x}, \mathbf{a}^{(\tau+1)} \right\rangle \middle| \mathbf{x} \in \mathbf{X}^{(\tau+1)} \right] \right)^2 \\ &= \left(\left\langle \mathbf{x} \dot{\otimes} \mathbf{x}, \mathbf{a}^{(\tau+1)} \right\rangle + \mathbb{E}_{\mathbf{x} \sim \gamma} \left[\varepsilon_{\mathbf{x}} \middle| \mathbf{x} \in \mathbf{X}^{(\tau+1)} \right] \right)^2 \\ &\quad \left(|\varepsilon_{\mathbf{x}}| \leq 2^{-L} \text{ is the truncation error by } \text{Fact 6.4} \right) \\ &\leq \left\langle \mathbf{x} \dot{\otimes} \mathbf{x}, \mathbf{a}^{(\tau+1)} \right\rangle^2 + 2^{-2L} + 2^{1-L} \cdot \left| \left\langle \mathbf{x} \dot{\otimes} \mathbf{x}, \mathbf{a}^{(\tau+1)} \right\rangle \right| \\ &\leq \left\langle \mathbf{x} \dot{\otimes} \mathbf{x}, \mathbf{a}^{(\tau+1)} \right\rangle^2 + nT \cdot 2^{2-L}, \end{aligned} \quad (6.8)$$

where the last line follows from $\left| \left\langle \mathbf{x} \dot{\otimes} \mathbf{x}, \mathbf{a}^{(\tau+1)} \right\rangle \right| \leq \left\| \mathbf{x} \dot{\otimes} \mathbf{x} \right\|$ and $\mathbf{x} \in \mathbf{X}^{(0)} = [-T, T]^n$.

Final Bound. Since $(\mathbf{u}^{(r)})_r$ is a matrix-valued martingale and thus $\mathbb{E} \left[\mathbf{u}^{(\tau+1)} \middle| \mathcal{F}^{(\tau)} \right] = \mathbf{u}^{(\tau)}$, we have

$$\mathbb{E} \left[\left\langle \mathbf{u}^{(\tau+1)} - \mathbf{u}^{(\tau)}, \mathbf{a}^{(\tau+1)} \right\rangle^2 \middle| \mathcal{F}^{(\tau)} \right] = \mathbb{E} \left[\left\langle \mathbf{u}^{(\tau+1)}, \mathbf{a}^{(\tau+1)} \right\rangle^2 - \left\langle \mathbf{u}^{(\tau)}, \mathbf{a}^{(\tau+1)} \right\rangle^2 \middle| \mathcal{F}^{(\tau)} \right]$$

Then by (6.8), we upper bound the right hand side by

$$nT \cdot 2^{2-L} + \mathbb{E}_{\mathbf{x} \sim \gamma} \left[\left\langle \mathbf{x} \otimes \mathbf{x}, \mathbf{a}^{(\tau+1)} \right\rangle^2 - \left\langle \mathbf{u}^{(\tau)}, \mathbf{a}^{(\tau+1)} \right\rangle^2 \middle| \mathcal{F}(\tau) \right].$$

Since $\mathbf{X}^{(\tau)}$ is 4-wise clean with parameter λ , it can be bounded by $nT \cdot 2^{2-L} + \lambda$:

$$\mathbb{E} \left[\left\langle \mathbf{u}^{(\tau+1)} - \mathbf{u}^{(\tau)}, \mathbf{a}^{(\tau+1)} \right\rangle^2 \middle| \mathcal{F}(\tau) \right] \leq nT \cdot 2^{2-L} + \lambda \quad (6.9)$$

Putting everything together, we have

$$\begin{aligned} \mathbb{E} \left[\left(\Delta \mathbf{z}_2^{(\tau+1)} \right)^2 \middle| \mathcal{F}(\tau) \right] &\leq \left(\beta^{(\tau+1)} \right)^2 \mathbb{E} \left[\left\langle \mathbf{u}^{(\tau+1)} - \mathbf{u}^{(\tau)}, \mathbf{a}^{(\tau+1)} \right\rangle^2 \middle| \mathcal{F}(\tau) \right] + 12n^7 T^3 \cdot 2^{-L} \quad (\text{by (6.6)}) \\ &\leq \left(\beta^{(\tau+1)} \right)^2 \cdot (nT \cdot 2^{2-L} + \lambda) + 12n^7 T^3 \cdot 2^{-L} \quad (\text{by (6.9)}) \\ &\leq \lambda \cdot \left(\beta^{(\tau+1)} \right)^2 + n^3 T^3 \cdot 2^{2-L} + 12n^7 T^3 \cdot 2^{-L} \quad (\text{by (6.5)}) \\ &\leq \lambda \cdot \left\| \mathbf{v}^{(\tau)} - \mathbf{v}^{(\tau_{m-1})} \right\|^2 + n^3 T^3 \cdot 2^{2-L} + 12n^7 T^3 \cdot 2^{-L} \quad (\text{by (6.7)}) \\ &\leq \lambda \cdot \left\| \mathbf{v}^{(\tau)} - \mathbf{v}^{(\tau_{m-1})} \right\|^2 + 16n^7 T^3 \cdot 2^{-L}. \end{aligned}$$

This completes the proof of the first statement in the lemma.

For the moreover part, let us condition on the event $\tau_{m-1} < t < \tau_m$ where Alice speaks at time t . Note that such t must all lie in the same phase of the protocol where Alice is the only one speaking. So, Bob's center of mass does not change from the time τ_{m-1} till t , i.e., $\mathbf{v}^{(t+1)} = \mathbf{v}^{(\tau_{m-1})}$. Thus we have

$$\Delta \mathbf{z}_2^{(t+1)} = \left\langle \mathbf{u}^{(t+1)} - \mathbf{u}^{(t)}, \eta \odot \mathbf{v}^{(\tau_{m-1})} \right\rangle. \quad (6.10)$$

Analogous to (6.3), the component of Alice's center of mass along the previous directions are fixed with precision 2^{-L} . Thus by Fact 6.4,

$$\left| \left\langle \mathbf{u}^{(t+1)} - \mathbf{u}^{(t)}, \mathbf{a}^{(r)} \right\rangle \right| \leq 2^{-L} \quad \text{for all } r \leq t. \quad (6.11)$$

Furthermore, by construction, $\eta \odot \mathbf{v}^{(\tau_{m-1})}$ lies in the space spanned by $\mathbf{a}^{(1)}, \dots, \mathbf{a}^{(\tau_{m-1}+1)}$. Note that $\tau_{m-1} + 1 \leq t$. Similar to the previous analysis, for each $r = 1, \dots, t$, let $\beta^{(r)} := \left\langle \eta \odot \mathbf{v}^{(t)}, \mathbf{a}^{(r)} \right\rangle$ be the length of $\eta \odot \mathbf{v}^{(t)}$ along direction $\mathbf{a}^{(r)}$. Then (6.5) also holds here. Therefore

$$\begin{aligned} \left| \Delta \mathbf{z}_2^{(t+1)} \right| &= \left| \sum_{r=1}^t \beta^{(r)} \cdot \left\langle \mathbf{u}^{(t+1)} - \mathbf{u}^{(t)}, \mathbf{a}^{(r)} \right\rangle \right| \quad (\text{by (6.10)}) \\ &\leq \sum_{r=1}^t \left| \beta^{(r)} \right| \cdot \left| \left\langle \mathbf{u}^{(t+1)} - \mathbf{u}^{(t)}, \mathbf{a}^{(r)} \right\rangle \right| \leq \sum_{r=1}^t nT \cdot 2^{-L} \quad (\text{by (6.5) and (6.11)}) \\ &\leq 2n^3 T \cdot 2^{-L}. \quad (\text{by Claim 6.7}) \end{aligned}$$

6.4 Conversion to Second Moment Bounds of the Depth (Proof of Lemma 6.10)

Recall $\gamma^* = \gamma(\mathbf{X}^{(0)}) \times \gamma(\mathbf{Y}^{(0)})$ and $\gamma(\ell) = \gamma(\mathbf{X}^{(D(\ell))}) \cdot \gamma(\mathbf{Y}^{(D(\ell))})$ for each leaf ℓ . The goal of this subsection is to prove Lemma 6.10.

We first note the following basic fact.

Fact 6.13. $\sum_{\ell} \gamma(\ell) = \gamma^*$ and

$$\Pr_{\mathbf{x} \sim X^{(0)}, \mathbf{y} \sim Y^{(0)}} [\bar{\mathcal{C}}(\mathbf{x}, \mathbf{y}) \text{ reaches leaf } \ell] = \gamma(\ell)/\gamma^*.$$

Now we apply [Theorem 3.1](#) with $k = 2$ to relate the LHS of [Lemma 6.10](#) with an entropy-type bound.

Lemma 6.14. $\mathbb{E} \left[\|\mathbf{u}^{(d)}\|^2 + \|\mathbf{v}^{(d)}\|^2 \right] \leq \frac{4e^2}{\gamma^*} \sum_{\ell} \gamma(\ell) \cdot \ln^2 \left(\frac{e}{\gamma(\ell)} \right).$

Proof. Let ℓ be a fixed leaf and $D = D(\ell)$ be its depth. Note that this also fixes the rectangle $X^{(D)} \times Y^{(D)}$ and thus the centers of mass $u^{(D)}, v^{(D)}$. Define the indicator function $\mathbf{1}_{\ell}: \mathbb{R}^{2n} \rightarrow \{0, 1\}$ by

$$\mathbf{1}_{\ell}(x, y) = \begin{cases} 1 & (x, y) \in X^{(D)} \times Y^{(D)}, \\ 0 & \text{otherwise.} \end{cases}$$

Then we have

$$\begin{aligned} & \left\| u^{(D)} \right\|^2 + \left\| v^{(D)} \right\|^2 \\ &= \left\| \mathbb{E}_{\mathbf{x} \sim \gamma} [\mathbf{x} \otimes \mathbf{x} \mid \mathbf{x} \in X^{(D)}] \right\|^2 + \left\| \mathbb{E}_{\mathbf{y} \sim \gamma} [\mathbf{y} \otimes \mathbf{y} \mid \mathbf{y} \in Y^{(D)}] \right\|^2 \\ &= \sum_{\substack{i, j=1 \\ i \neq j}}^n \left(\mathbb{E}_{\mathbf{x} \sim \gamma} [\mathbf{x}_i \mathbf{x}_j \mid \mathbf{x} \in X^{(D)}] \right)^2 + \sum_{\substack{i, j=1 \\ i \neq j}}^n \left(\mathbb{E}_{\mathbf{y} \sim \gamma} [\mathbf{y}_i \mathbf{y}_j \mid \mathbf{y} \in Y^{(D)}] \right)^2 \\ &= \sum_{\substack{i, j=1 \\ i \neq j}}^n \left(\mathbb{E}_{\mathbf{x}, \mathbf{y} \sim \gamma} [\mathbf{x}_i \mathbf{x}_j \mid (\mathbf{x}, \mathbf{y}) \in X^{(D)} \times Y^{(D)}] \right)^2 + \sum_{\substack{i, j=1 \\ i \neq j}}^n \left(\mathbb{E}_{\mathbf{x}, \mathbf{y} \sim \gamma} [\mathbf{y}_i \mathbf{y}_j \mid (\mathbf{x}, \mathbf{y}) \in X^{(D)} \times Y^{(D)}] \right)^2 \\ &= \frac{2}{\gamma(\ell)^2} \left(\sum_{S \in \binom{[n]}{2}} \left(\mathbb{E}_{\mathbf{x} \sim \gamma, \mathbf{y} \sim \gamma} [\mathbf{1}_{\ell}(\mathbf{x}, \mathbf{y}) \mathbf{x}_S] \right)^2 + \sum_{S \in \binom{[n]}{2}} \left(\mathbb{E}_{\mathbf{x} \sim \gamma, \mathbf{y} \sim \gamma} [\mathbf{1}_{\ell}(\mathbf{x}, \mathbf{y}) \mathbf{y}_S] \right)^2 \right) \\ &\leq \frac{2}{\gamma(\ell)^2} \sum_{S \in \binom{[2n]}{2}} \left(\mathbb{E}_{\mathbf{w} \sim \gamma_n \times \gamma_n} [\mathbf{1}_{\ell}(\mathbf{w}) \mathbf{w}_S] \right)^2 \\ &\leq \frac{2}{\gamma(\ell)^2} \cdot 2e^2 \gamma(\ell)^2 \cdot \ln^2 \left(\frac{e}{\gamma(\ell)} \right) \quad (\text{by } \text{Theorem 3.1}) \\ &= 4e^2 \cdot \ln^2 \left(\frac{e}{\gamma(\ell)} \right). \end{aligned}$$

Therefore taking expectation over a random ℓ , by [Fact 6.13](#), we have

$$\mathbb{E} \left[\|\mathbf{u}^{(d)}\|^2 + \|\mathbf{v}^{(d)}\|^2 \right] \leq 4e^2 \cdot \mathbb{E}_{\ell} \left[\ln^2 \left(\frac{e}{\gamma(\ell)} \right) \right] = \frac{4e^2}{\gamma^*} \sum_{\ell} \gamma(\ell) \cdot \ln^2 \left(\frac{e}{\gamma(\ell)} \right). \quad \square$$

Now in the next lemma, we bound the right hand side of [Lemma 6.14](#) in terms of the second moment of the depth, which immediately proves [Lemma 6.10](#).

Lemma 6.15. *Assume that $Tn \leq 2^L$. Then, $\sum_{\ell} \gamma(\ell) \cdot \ln^2(e/\gamma(\ell)) \leq O(1 + \gamma^* \cdot L^2 \mathbb{E}[d^2])$.*

Proof. By [Claim 6.6](#), and the assumption $Tn \leq 2^L$ each message is of length at most $L + \log(Tn) \leq 2L$. We divide ℓ into two cases based on $\gamma(\ell)$:

$$\begin{aligned}
& \sum_{\ell: \gamma(\ell) < 2^{-3L \cdot D(\ell)}} \gamma(\ell) \cdot \ln^2 \left(\frac{e}{\gamma(\ell)} \right) \\
& \leq \sum_{\ell: \gamma(\ell) < 2^{-3L \cdot D(\ell)}} 2^{-3L \cdot D(\ell)} \cdot \ln^2 \left(e \cdot 2^{3L \cdot D(\ell)} \right) \quad (x \ln^2(e/x) \text{ is increasing when } 0 \leq x \leq 0.2) \\
& \leq \sum_{t=1}^{\infty} 2^{-3L \cdot t} \cdot 2(9L^2 t^2 + 1) \cdot |\{\ell : D(\ell) = t\}| \quad (\text{since } \ln^2(ab) \leq 2 \ln^2(a) + 2 \ln^2(b)) \\
& \leq \sum_{t=1}^{\infty} 2^{-3L \cdot t} \cdot 2(9L^2 t^2 + 1) \cdot 2^{(2L) \cdot t} \quad (\text{each message is of length } \leq 2L) \\
& \leq \sum_{t=1}^{\infty} 2(9L^2 t^2 + 1) \cdot 2^{-Lt} = O(1) \quad (\text{since } L \geq 2)
\end{aligned}$$

and

$$\begin{aligned}
\sum_{\ell: \gamma(\ell) \geq 2^{-3L \cdot D(\ell)}} \gamma(\ell) \cdot \ln^2 \left(\frac{e}{\gamma(\ell)} \right) & \leq \sum_{\ell: \gamma(\ell) \geq 2^{-3L \cdot D(\ell)}} \gamma(\ell) \cdot \ln^2 \left(e \cdot 2^{3L \cdot D(\ell)} \right) \\
& \leq 2 \cdot 9L^2 \sum_{\ell} \gamma(\ell) D(\ell)^2 + 2 \sum_{\ell} \gamma(\ell) \\
& = 18L^2 \gamma^* \cdot \mathbb{E}_{\ell} [D(\ell)^2] + 2 \\
& = 18L^2 \gamma^* \cdot \mathbb{E} [\mathbf{d}^2] + 2.
\end{aligned}$$

Adding up the two estimates above gives the desired bound. \square

6.5 Second Moment Bounds for the Depth (Proof of [Lemma 6.11](#))

The final ingredient is an estimate for the second moment $\mathbb{E}[\mathbf{d}^2]$. This subsection is devoted to this goal and proving [Lemma 6.11](#).

For messages $\ell' = (\bar{\mathbf{c}}^{(1)}, \dots, \bar{\mathbf{c}}^{(t)})$, we define $\gamma(\ell') = \gamma(\mathbf{X}^{(t)}) \cdot \gamma(\mathbf{Y}^{(t)})$ where $\mathbf{X}^{(t)}, \mathbf{Y}^{(t)}$ is defined by the protocol using the messages ℓ' . Note that this definition is consistent with $\gamma(\ell)$ from [Subsection 6.4](#) for a leaf ℓ .

Lemma 6.16. *There exists a universal constant $\alpha > 0$ such that the following holds. Let $0 \leq d_1 < d_2$ be two arbitrary integers with $d_2 - d_1 \geq 2d + 1$. Let $\ell^* = (\bar{\mathbf{c}}^{(1)}, \dots, \bar{\mathbf{c}}^{(d_1)})$ be arbitrary messages of the first d_1 communication steps. Assume $2^L \geq 8n^4 T^2$. Then*

$$\Pr[\mathbf{d} \geq d_2 \mid \ell^*] \leq \frac{\alpha \cdot d_2^2 L^2}{\lambda \cdot (d_2 - d_1 - 2d)} + \frac{1}{4} \cdot \frac{2^{-3L \cdot d_1}}{\gamma(\ell^*)}.$$

Proof. Let \mathbf{x}, \mathbf{y} be sampled from γ conditioned on $\mathbf{x} \in \mathbf{X}^{(0)}, \mathbf{y} \in \mathbf{Y}^{(0)}$. Let ℓ be its corresponding leaf in $\bar{\mathbf{C}}$ and \mathbf{d} be the depth of ℓ . By [Claim 6.7](#), ℓ always has finite depth. We extend $\mathbf{a}^{(t)} = \mathbf{b}^{(t)} = \mathbf{0}^{n \times n}$ and $\mathbf{X}^{(t)} = \mathbf{X}^{(d)}, \mathbf{Y}^{(t)} = \mathbf{Y}^{(d)}$ for all $t > d$. Then define

$$\mathbf{k}(\mathbf{x}, \mathbf{y}) = \sum_{t=d_1+1}^{d_2} \left(\left\langle \mathbf{x} \dot{\otimes} \mathbf{x}, \mathbf{a}^{(t)} \right\rangle^2 + \left\langle \mathbf{y} \dot{\otimes} \mathbf{y}, \mathbf{b}^{(t)} \right\rangle^2 \right) \quad \text{and} \quad K = \mathbb{E}_{\mathbf{x}, \mathbf{y} \sim \gamma} [\mathbf{k}(\mathbf{x}, \mathbf{y}) \mid \ell^*],$$

where $\bar{\mathbf{a}}^{(\cdot)}$'s and $\bar{\mathbf{b}}^{(\cdot)}$'s depend only on ℓ .¹³ Equivalently, we can write K as

$$K = \mathbb{E}_{\mathbf{x}, \mathbf{y} \sim \gamma} \left[\mathbf{k}(\mathbf{x}, \mathbf{y}) \mid (\mathbf{x}, \mathbf{y}) \in X^{(d_1)} \times Y^{(d_1)} \right],$$

where $X^{(d_1)}$ and $Y^{(d_1)}$ are fixed due to ℓ^* .

Observe that for any fixed $t \geq d_1$, $\mathbf{X}^{(t)} \times \mathbf{Y}^{(t)}$ induced by different ℓ , conditioned on ℓ^* , is a disjoint partition of $X^{(d_1)} \times Y^{(d_1)}$. Therefore sampling $\mathbf{x}, \mathbf{y} \sim \gamma$ conditioned on $(\mathbf{x}, \mathbf{y}) \in X^{(d_1)} \times Y^{(d_1)}$ is equivalent to

- first sample random messages $\ell' = (\bar{\mathbf{c}}^{(d_1+1)}, \dots, \bar{\mathbf{c}}^{(t)})$ conditioned on ℓ^* ,
- then sample $\mathbf{x}, \mathbf{y} \sim \gamma$ conditioned on $(\mathbf{x}, \mathbf{y}) \in \mathbf{X}^{(t)} \times \mathbf{Y}^{(t)}$ given ℓ' .

Note that we can further expand ℓ' to a leaf ℓ as a full communication path, and obtain the following equivalent sampling process:

- Sample a random leaf ℓ conditioned on ℓ^* .
- Sample $\mathbf{x}, \mathbf{y} \sim \gamma$ conditioned on $(\mathbf{x}, \mathbf{y}) \in \mathbf{X}^{(t)} \times \mathbf{Y}^{(t)}$ defined by the first t messages of ℓ .

As a result, we have

$$\begin{aligned} K &= \sum_{t=d_1+1}^{d_2} \mathbb{E}_{\ell} \left[\mathbb{E}_{\mathbf{x}, \mathbf{y} \sim \gamma} \left[\left\langle \mathbf{x} \dot{\otimes} \mathbf{x}, \mathbf{a}^{(t)} \right\rangle^2 + \left\langle \mathbf{y} \dot{\otimes} \mathbf{y}, \mathbf{b}^{(t)} \right\rangle^2 \mid (\mathbf{x}, \mathbf{y}) \in \mathbf{X}^{(t)} \times \mathbf{Y}^{(t)} \right] \mid \ell^* \right] \\ &= \mathbb{E}_{\ell} \left[\sum_{t=d_1+1}^{d_2} \mathbb{E}_{\mathbf{x} \sim \gamma} \left[\left\langle \mathbf{x} \dot{\otimes} \mathbf{x}, \mathbf{a}^{(t)} \right\rangle^2 \mid \mathbf{x} \in \mathbf{X}^{(t)} \right] + \mathbb{E}_{\mathbf{y} \sim \gamma} \left[\left\langle \mathbf{y} \dot{\otimes} \mathbf{y}, \mathbf{b}^{(t)} \right\rangle^2 \mid \mathbf{y} \in \mathbf{Y}^{(t)} \right] \mid \ell^* \right]. \end{aligned}$$

Observe that there are at most $2d$ many step 3(a) and 3(b) in ℓ . This means, if $d \geq d_2$, then from the $(d_1 + 1)$ -th to the d_2 -th communication steps, there are at least $d_2 - d_1 - 2d$ cleanup steps (i.e., step 3(c)), each of which contributes at least λ to K . Thus we can lower bound K by

$$K \geq \lambda \cdot (d_2 - d_1 - 2d) \cdot \Pr [d \geq d_2 \mid \ell^*]. \quad (6.12)$$

On the other hand by [Claim 6.7](#), there are at most n^2 non-zero $\mathbf{a}^{(\cdot)}$'s and at most n^2 non-zero $\mathbf{b}^{(\cdot)}$'s in each communication path. Thus

$$\mathbf{k}(\mathbf{x}, \mathbf{y}) \leq n^2 \cdot \left(\max_{x \in \mathbf{X}^{(0)}} \left\| x \dot{\otimes} x \right\|^2 + \max_{y \in \mathbf{Y}^{(0)}} \left\| y \dot{\otimes} y \right\|^2 \right) < 2n^4 T^2. \quad (6.13)$$

We now obtain another upper bound using [Theorem 3.3](#). Let $\bar{\ell} = (\bar{\mathbf{c}}^{(1)}, \dots, \bar{\mathbf{c}}^{(d_2)})$ extend ℓ^* for the next $d_2 - d_1$ messages.¹⁴ Then $K = \mathbb{E}_{\bar{\ell}} [\mathbf{k}(\bar{\ell}) \mid \ell^*]$ where $\mathbf{k}(\bar{\ell}) := \mathbb{E}_{\mathbf{x}, \mathbf{y} \sim \gamma} [\mathbf{k}(\mathbf{x}, \mathbf{y}) \mid \bar{\ell}]$. Note that $\bar{\ell}$ fixes $\mathbf{a}^{(\cdot)}$'s and $\mathbf{b}^{(\cdot)}$'s in $\mathbf{k}(\mathbf{x}, \mathbf{y})$. Therefore we use $\mathbf{k}_{\bar{\ell}}(\mathbf{x}, \mathbf{y})$ to denote $\mathbf{k}(\mathbf{x}, \mathbf{y})$ with the directions $\mathbf{a}^{(\cdot)}$'s and $\mathbf{b}^{(\cdot)}$'s fixed by $\bar{\ell}$. We now continue the bound on $\mathbf{k}(\bar{\ell})$:

$$\mathbf{k}(\bar{\ell}) \leq \sum_{t=0}^{\infty} \Pr_{\mathbf{x}, \mathbf{y} \sim \gamma} [\mathbf{k}_{\bar{\ell}}(\mathbf{x}, \mathbf{y}) \geq t \mid \bar{\ell}] = \sum_{t=0}^{\infty} \frac{\Pr_{\mathbf{x}, \mathbf{y} \sim \gamma} [\mathbf{k}_{\bar{\ell}}(\mathbf{x}, \mathbf{y}) \geq t, \bar{\ell}]}{\Pr_{\mathbf{x}, \mathbf{y} \sim \gamma} [\bar{\ell}]}$$

¹³Note that ℓ specifies all the communication messages, which allows us to simulate the protocol and obtain each $\mathbf{a}^{(\cdot)}$ and $\mathbf{b}^{(\cdot)}$.

¹⁴If $\bar{\ell}$ becomes a leaf before d_2 , then we can simply pad dummy messages to it.

$$\begin{aligned}
&= \sum_{t=0}^{\infty} \min \left\{ 1, \frac{\Pr_{\mathbf{x}, \mathbf{y} \sim \gamma} [\mathbf{k}_{\bar{\ell}}(\mathbf{x}, \mathbf{y}) \geq t, \bar{\ell}]}{\gamma(\bar{\ell})} \right\} && \text{(by the definition of } \gamma(\cdot)\text{)} \\
&\leq \sum_{t=0}^{\infty} \min \left\{ 1, \frac{\Pr_{\mathbf{x}, \mathbf{y} \sim \gamma} [\mathbf{k}_{\bar{\ell}}(\mathbf{x}, \mathbf{y}) \geq t]}{\gamma(\bar{\ell})} \right\}. && (6.14)
\end{aligned}$$

We now analyze $\Pr_{\mathbf{x}, \mathbf{y} \sim \gamma} [\mathbf{k}_{\bar{\ell}}(\mathbf{x}, \mathbf{y}) \geq t]$ using [Theorem 3.3](#). Since $a^{(t)}, b^{(t)}$ cannot be non-zero simultaneously, we rearrange the matrices and assume $a^{(d_1+1)}, \dots, a^{(d')}, b^{(d'+1)}, \dots, b^{(d'')}$ are the only non-zero matrices where $d'' \leq d_2$. Then

$$\mathbf{k}_{\bar{\ell}}(\mathbf{x}, \mathbf{y}) = \sum_{t=d_1+1}^{d'} \langle \mathbf{x} \dot{\otimes} \mathbf{x}, a^{(t)} \rangle^2 + \sum_{t=d'+1}^{d''} \langle \mathbf{y} \dot{\otimes} \mathbf{y}, b^{(t)} \rangle^2.$$

Note that a 's (resp., b 's) satisfy the condition in [Theorem 3.3](#). Let $1/\kappa$ be the constant¹⁵ in Ω in [Theorem 3.3](#). Hence

$$\begin{aligned}
\Pr [\mathbf{k}_{\bar{\ell}}(\mathbf{x}, \mathbf{y}) \geq t] &\leq \Pr \left[\sum_{t=d_1+1}^{d'} \langle \mathbf{x} \dot{\otimes} \mathbf{x}, a^{(t)} \rangle^2 \geq t/2 \right] + \Pr \left[\sum_{t=d'+1}^{d''} \langle \mathbf{y} \dot{\otimes} \mathbf{y}, b^{(t)} \rangle^2 \geq t/2 \right] \\
&\leq 2 \exp \left\{ -\frac{1}{\kappa} \cdot \frac{t/2}{d' - d_1 + \sqrt{t/2}} \right\} + 2 \exp \left\{ -\frac{1}{\kappa} \cdot \frac{t/2}{d'' - d' + \sqrt{t/2}} \right\} \\
&\quad \text{(by [Theorem 3.3](#) and assuming } t \geq 196 \cdot \max \{d' - d_1, d'' - d'\}\text{)} \\
&\leq 4 \exp \left\{ -\frac{1}{\kappa} \cdot \frac{t/2}{d_2 - d_1 + \sqrt{t/2}} \right\}. && \text{(since } d_1 \leq d' \leq d'' \leq d_2\text{)}
\end{aligned}$$

Thus for any $t \geq 196 \cdot (d_2 - d_1) \geq 196 \cdot \max \{d' - d_1, d'' - d'\}$, we have

$$\Pr [\mathbf{k}_{\bar{\ell}}(\mathbf{x}, \mathbf{y}) \geq t] \leq 4 \exp \left\{ -\frac{1}{\kappa} \cdot \frac{t/2}{d_2 - d_1 + \sqrt{t/2}} \right\}. \quad (6.15)$$

For $\gamma(\bar{\ell}) \geq 2^{-3L \cdot d_2}$, we plug (6.15) into (6.14) and obtain

$$\begin{aligned}
\mathbf{k}(\bar{\ell}) &\leq \sum_{t=0}^{196 \cdot (d_2 - d_1)^2} 1 + \sum_{t > 196 \cdot (d_2 - d_1)^2} \min \left\{ 1, 2^{3L \cdot d_2 + 1} \cdot \exp \left\{ -\frac{1}{\kappa} \cdot \frac{t/2}{d_2 - d_1 + \sqrt{t/2}} \right\} \right\} && \text{(by (6.15))} \\
&\leq 196 \cdot (d_2 - d_1)^2 + 1 + \sum_{t \geq 196 \cdot (d_2 - d_1)^2} \min \left\{ 1, 2^{3L \cdot d_2 + 1} \cdot e^{-\frac{1}{\kappa} \cdot \frac{t/2}{2\sqrt{t/2}}} \right\} \\
&\leq 197 \cdot d_2^2 + \sum_{t \geq 1} \min \left\{ 1, 2^{3L \cdot d_2 + 1} \cdot e^{-\frac{\sqrt{t/2}}{2\kappa}} \right\} \\
&\leq \alpha \cdot d_2^2 L^2, && (6.16)
\end{aligned}$$

where α is another universal constant. Now we have

$$K = \mathbb{E}_{\bar{\ell}} [\mathbf{k}(\bar{\ell}) \mid \ell^*] = \sum_{\bar{\ell}} \frac{\gamma(\bar{\ell})}{\gamma(\ell^*)} \cdot \mathbf{k}(\bar{\ell}) = \sum_{\bar{\ell}: \gamma(\bar{\ell}) < 2^{-3L \cdot d_2}} \frac{\gamma(\bar{\ell})}{\gamma(\ell^*)} \cdot \mathbf{k}(\bar{\ell}) + \sum_{\bar{\ell}: \gamma(\bar{\ell}) \geq 2^{-3L \cdot d_2}} \frac{\gamma(\bar{\ell})}{\gamma(\ell^*)} \cdot \mathbf{k}(\bar{\ell}),$$

¹⁵In particular $\kappa = 56448$ suffices from our proof in [Appendix B](#).

where the first summation can be bounded by

$$\begin{aligned}
\sum_{\bar{\ell}:\gamma(\bar{\ell})<2^{-3L\cdot d_2}} \frac{\gamma(\bar{\ell})}{\gamma(\ell^*)} \cdot \mathbf{k}(\bar{\ell}) &\leq \frac{2^{-3L\cdot d_1}}{\gamma(\ell^*)} \cdot \sum_{\bar{\ell}} 2^{-3L\cdot(d_2-d_1)} \cdot n^4 T^2 && \text{(by (6.13))} \\
&\leq \frac{2^{-3L\cdot d_1}}{\gamma(\ell^*)} \cdot 2^{2L\cdot(d_2-d_1)} \cdot 2^{-3L\cdot(d_2-d_1)} \cdot n^4 T^2 \\
&\quad \text{(since } \ell^* \text{ is fixed and each message is at most } 2L \text{ bits)} \\
&= \frac{2^{-3L\cdot d_1}}{\gamma(\ell^*)} \cdot \frac{2n^4 T^2}{2^L} && \text{(since } d_2 - d_1 \geq 1)
\end{aligned}$$

and the second summation is bounded by

$$\sum_{\bar{\ell}:\gamma(\bar{\ell})\geq 2^{-3L\cdot d_2}} \frac{\gamma(\bar{\ell})}{\gamma(\ell^*)} \cdot \mathbf{k}(\bar{\ell}) \leq \sum_{\bar{\ell}} \frac{\gamma(\bar{\ell})}{\gamma(\ell^*)} \cdot \alpha \cdot d_2^2 L^2 = \alpha \cdot d_2^2 L^2. \quad \text{(by (6.16))}$$

Then combining (6.12), we have

$$\lambda \cdot (d_2 - d_1 - 2d) \cdot \Pr[\mathbf{d} \geq d_2 \mid \ell^*] \leq \alpha \cdot d_2^2 L^2 + \frac{2^{-3L\cdot d_1}}{\gamma(\ell^*)} \cdot \frac{2n^4 T^2}{2^L}.$$

Assume $2^L \geq 8n^4 T^2$ and $d_2 - d_1 \geq 2d + 1$. Then

$$\Pr[\mathbf{d} \geq d_2 \mid \ell^*] \leq \frac{\alpha \cdot d_2^2 L^2}{\lambda \cdot (d_2 - d_1 - 2d)} + \frac{1}{4} \cdot \frac{2^{-3L\cdot d_1}}{\gamma(\ell^*)}. \quad \square$$

Corollary 6.17. *Assume $\gamma^* \geq 3/4$, $T \leq n$, $L \geq \Theta(\log(n))$, and $\lambda \geq \Theta(dL^2 \log^2(n))$. Then for each $k = 0, 1, \dots, 4\log(n)$, we have*

$$\Pr[\mathbf{d} \geq 4kd] \leq 2^{-k} + \frac{k}{n^5}.$$

Proof. We prove the bound by induction on k . The base case $k = 0$ is trivial. For the inductive case, let ℓ^* be the first $4(k-1)d$ communication messages. Then we bound

$$P := \sum_{\ell^*:\gamma(\ell^*)/\gamma^* < 2^{-3L\cdot 4(k-1)d}} \frac{\gamma(\ell^*)}{\gamma^*} \cdot \Pr[\mathbf{d} \geq 4kd \mid \ell^*]$$

and

$$Q := \sum_{\ell^*:\gamma(\ell^*)/\gamma^* \geq 2^{-3L\cdot 4(k-1)d}} \frac{\gamma(\ell^*)}{\gamma^*} \cdot \Pr[\mathbf{d} \geq 4kd \mid \ell^*]$$

separately.

For P , observe that if $k = 1$ then ℓ^* is root of the protocol, thus $\gamma(\ell^*) = \gamma^*$ and $P = 0$. On the other hand, if $k \geq 2$, then

$$\begin{aligned}
P &\leq \sum_{\ell^*:\gamma(\ell^*)/\gamma^* < 2^{-3L\cdot 4(k-1)d}} 2^{-3L\cdot 4(k-1)d} \leq \sum_{\ell^*} 2^{-3L\cdot 4(k-1)d} \\
&\leq 2^{2L\cdot 4(k-1)d} \cdot 2^{-3L\cdot 4(k-1)d} && \text{(each communication message is at most } 2L \text{ bits)} \\
&= 2^{-L\cdot 4(k-1)d} \leq n^{-5}. && \text{(since } k \geq 2 \text{ and } L \geq \Theta(\log(n)))
\end{aligned}$$

Now we turn to Q . Applying [Lemma 6.16](#) with ℓ^* and $d_1 = 4(k-1)d, d_2 = 4kd$, we have

$$\begin{aligned}
Q &\leq \sum_{\ell^*: \gamma(\ell^*)/\gamma^* \geq 2^{-3L \cdot 4(k-1)d}} \frac{\gamma(\ell^*)}{\gamma^*} \cdot \left(\frac{16\alpha \cdot k^2 d^2 L^2}{2dR} + \frac{1}{4} \cdot \frac{2^{-3L \cdot 4(k-1)d}}{\gamma(\ell^*)} \right) \\
&\leq \sum_{\ell^*} \frac{\gamma(\ell^*)}{\gamma^*} \cdot \left(\frac{8\alpha \cdot k^2 d L^2}{\lambda} + \frac{1}{4\gamma^*} \right) \\
&= \Pr[\mathbf{d} \geq 4(k-1)d] \cdot \left(\frac{8\alpha \cdot k^2 d L^2}{\lambda} + \frac{1}{4\gamma^*} \right) \\
&\leq \Pr[\mathbf{d} \geq 4(k-1)d] \cdot \frac{1}{2} \quad (\text{since } \gamma^* \geq 3/4 \text{ and } \lambda \geq \Theta(dL^2 \log^2(n)), k \leq 4 \log(n)) \\
&\leq \left(2^{-(k-1)} + \frac{k-1}{n^5} \right) \cdot \frac{1}{2} \leq 2^{-k} + \frac{k-1}{n^5}. \quad (\text{by induction hypothesis})
\end{aligned}$$

By adding up P and Q , we complete the induction. \square

Given [Corollary 6.17](#) and suitable choice of the parameters, we now prove the second moment bound.

Proof of [Lemma 6.11](#). With $L = \Theta(\log(n))$, $T = \Theta(\sqrt{\log(n)})$, and $\lambda = \Theta(d \log^4(n))$, by [Fact 6.3](#), we have $\gamma^* \geq 3/4$. Therefore the second moment of \mathbf{d} is

$$\begin{aligned}
\mathbb{E}[\mathbf{d}^2] &\leq \sum_{k=0}^{4 \log(n)} (4(k+1)d)^2 \cdot \Pr[\mathbf{d} \geq 4kd] + \Pr[\mathbf{d} \geq 16d \log(n)] \cdot (2n^2)^2 \quad (\text{by [Claim 6.7](#)}) \\
&\leq \sum_{k=0}^{4 \log(n)} (4(k+1)d)^2 \cdot \left(2^{-k} + \frac{k}{n^5} \right) + \left(n^{-4} + \frac{4 \log(n)}{n^5} \right) \cdot (2n^2)^2 \quad (\text{by [Corollary 6.17](#)}) \\
&= O(d^2). \quad \square
\end{aligned}$$

7 Fourier Growth Reductions For General Gadgets

In this section, we show that Fourier growth bounds of communication protocols for general (constant-sized) gadgets can be reduced to the bounds of XOR-fiber, and vice versa. This implies that in the study of Fourier growth, they are all equivalent.

Let m_1, m_2 be two positive integers. Let $g: \{\pm 1\}^{m_1} \times \{\pm 1\}^{m_2} \rightarrow \{\pm 1\}$ be a gadget. Recall that ν is the uniform distribution over $\{\pm 1\}^n$. We now use $\nu_1, \nu_2, \bar{\nu}_1, \bar{\nu}_2$ to denote the uniform distributions over $\{\pm 1\}^{m_1}, \{\pm 1\}^{m_2}, (\{\pm 1\}^{m_1})^n, (\{\pm 1\}^{m_2})^n$ respectively. We define the g -fiber of communication protocols similar to the XOR-fiber:

Definition 7.1. For any randomized two-party protocol $\mathcal{C}: (\{\pm 1\}^{m_1})^n \times (\{\pm 1\}^{m_2})^n \rightarrow [-1, 1]$, its g -fiber, denoted by $\mathcal{C}_{\downarrow g}: \{\pm 1\}^n \rightarrow [-1, 1]$, is defined by

$$\mathcal{C}_{\downarrow g}(z) = \mathbb{E}_{\mathbf{x} \sim \bar{\nu}_1, \mathbf{y} \sim \bar{\nu}_2} [\mathcal{C}(\mathbf{x}, \mathbf{y}) \mid g(\mathbf{x}_i, \mathbf{y}_i) = z_i, \forall i],$$

where the expectation is also over the internal randomness of \mathcal{C} .

To compare the Fourier growth bounds between gadgets, we use $L_{1,k}(g, d, m_1, m_2, n)$ to denote the upper bound of the level- k Fourier growth for the g -fiber of an arbitrary randomized communication protocol $\mathcal{C}: (\{\pm 1\}^{m_1})^n \times (\{\pm 1\}^{m_2})^n \rightarrow [-1, 1]$ with at most d bits of communication, where

$g: \{\pm 1\}^{m_1} \times \{\pm 1\}^{m_2} \rightarrow \{\pm 1\}$ is the gadget. Since randomized protocols are convex combinations of deterministic protocols of the same cost, using this notation, our main results [Theorems 1.2](#) and [1.3](#) can be rephrased as

$$L_{1,1}(\text{XOR}, d, 1, 1, n) \leq O\left(\sqrt{d}\right) \quad \text{and} \quad L_{1,2}(\text{XOR}, d, 1, 1, n) \leq O\left(d^{3/2} \log^3(n)\right).$$

For any set $S \subseteq [m_1]$, define $x_S = \prod_{i \in S} x_i$, and similarly for y_T with $T \subseteq [m_2]$. Similar to the standard Fourier representation of Boolean functions, the gadget g , which is a two-party function, also has Fourier representation:

$$g(x, y) = \sum_{S \subseteq [m_1], T \subseteq [m_2]} \widehat{g}(S, T) \cdot x_S y_T, \quad \text{where} \quad \widehat{g}(S, T) = \mathbb{E}_{\mathbf{x} \sim \nu_1, \mathbf{y} \sim \nu_2} [g(\mathbf{x}, \mathbf{y}) \cdot \mathbf{x}_S \mathbf{y}_T].$$

For convenience, we will assume g satisfies the following assumption. It's easy to see that the XOR gadget satisfies it.

Assumption 7.2. $\widehat{g}(S, T) = 0$ if $S = \emptyset$ or $T = \emptyset$.

Remark 7.3. This assumption is equivalent to the fact that, restricted on any input to Alice's side, the remaining function on Bob's side is balanced, and vice versa.

Even if g does not satisfy the assumption, then we can embed it inside a similar gadget $g': \{\pm 1\}^{m_1+1} \times \{\pm 1\}^{m_2+1} \rightarrow \{\pm 1\}$, where we XOR the last bit of Alice and the last bit of Bob to the old gadget g applied to Alice's first m_1 bits and Bob's first m_2 bits, i.e.,

$$g'(x, y) = x_{m_1+1} y_{m_2+1} \cdot g(x_{\leq m_1}, y_{\leq m_2}).$$

Then g' satisfies the assumption and inherits most properties of g sufficient for studies in communication complexity tasks.

Now for a protocol $\mathcal{C}: (\{\pm 1\}^{m_1})^n \times (\{\pm 1\}^{m_2})^n \rightarrow [-1, 1]$, it is also a two-party function and thus admitting similar Fourier representation. We view an input from $(\{\pm 1\}^{m_1})^n$ as indexed by a tuple in $[n] \times [m_1]$. Therefore any subset of $(\{\pm 1\}^{m_1})^n$ is uniquely identified as $\bigcup_{i \in [n]} \{i\} \times S_i$, where each $S_i \subseteq [m_1]$. We use $S^{[n]}$ to denote $(S_i)_{i \in [n]}$. Thus the Fourier coefficients of \mathcal{C} can be written as

$$\widehat{\mathcal{C}}(S^{[n]}, T^{[n]}) := \widehat{\mathcal{C}}\left(\bigcup_{i \in [n]} \{i\} \times S_i, \bigcup_{i \in [n]} \{i\} \times T_i\right),$$

and the Fourier representation of \mathcal{C} is

$$\mathcal{C}(x, y) = \sum_{S^{[n]}, T^{[n]}} \widehat{\mathcal{C}}(S^{[n]}, T^{[n]}) \cdot \prod_{i \in [n]} x_{i, S_i} \cdot \prod_{j \in [n]} y_{j, T_j},$$

where $x_{i, S} = \prod_{j \in S} x_{i, j}$ and similar for $y_{j, T}$.

Under this notation and assuming [Assumption 7.2](#), we can effectively compute the Fourier coefficients of any g -fiber.

Fact 7.4. Assume gadget $g: \{\pm 1\}^{m_1} \times \{\pm 1\}^{m_2} \rightarrow \{\pm 1\}$ satisfies [Assumption 7.2](#). Then we have

$$\widehat{\mathcal{C}}_{\downarrow g}(I) = \sum_{\substack{S^I, T^I \\ S_i \neq \emptyset, T_i \neq \emptyset, \forall i \in I}} \widehat{\mathcal{C}}(S^I, T^I) \cdot \prod_{i \in I} \widehat{g}(S_i, T_i) \quad \text{for any } I \subseteq [n],$$

where we use S^I to denote $S^{[n]}$ with S_j fixed to \emptyset for all $j \notin I$.

Proof. Observe that

$$\begin{aligned}
\widehat{\mathcal{C}}_{\downarrow g}(I) &= \mathbb{E}_{\mathbf{z} \sim \nu} \left[\mathcal{C}_{\downarrow g}(\mathbf{z}) \cdot \prod_{i \in I} z_i \right] \\
&= \mathbb{E}_{\mathbf{z} \sim \nu} \left[\mathbb{E}_{\mathbf{x} \sim \bar{\nu}_1, \mathbf{y} \sim \bar{\nu}_2} [\mathcal{C}(\mathbf{x}, \mathbf{y}) \mid g(\mathbf{x}_i, \mathbf{y}_i) = z_i, \forall i] \cdot \prod_{i \in I} z_i \right] \\
&= \mathbb{E}_{\mathbf{z} \sim \nu} \left[\mathbb{E}_{\mathbf{x} \sim \bar{\nu}_1, \mathbf{y} \sim \bar{\nu}_2} \left[\mathcal{C}(\mathbf{x}, \mathbf{y}) \cdot \prod_{i \in I} g(\mathbf{x}_i, \mathbf{y}_i) \mid g(\mathbf{x}_i, \mathbf{y}_i) = z_i, \forall i \right] \right].
\end{aligned}$$

Since $\widehat{g}(\emptyset, \emptyset) = 0$ by [Assumption 7.2](#), every pair (x, y) is sampled with the same probability under the conditional distribution. Thus we get

$$\widehat{\mathcal{C}}_{\downarrow g}(I) = \mathbb{E}_{\mathbf{x} \sim \bar{\nu}_1, \mathbf{y} \sim \bar{\nu}_2} \left[\mathcal{C}(\mathbf{x}, \mathbf{y}) \cdot \prod_{i \in I} g(\mathbf{x}_i, \mathbf{y}_i) \right].$$

Now we expand \mathcal{C} and g in the Fourier basis and obtain

$$\begin{aligned}
\widehat{\mathcal{C}}_{\downarrow g}(I) &= \mathbb{E}_{\mathbf{x} \sim \bar{\nu}_1, \mathbf{y} \sim \bar{\nu}_2} \left[\left(\sum_{S^{[n]}, T^{[n]}} \widehat{\mathcal{C}}(S^{[n]}, T^{[n]}) \prod_{i \in [n]} \mathbf{x}_{i, S_i} \prod_{j \in [n]} \mathbf{y}_{j, T_j} \right) \cdot \prod_{i \in I} \left(\sum_{S_i, T_i} \widehat{g}(S_i, T_i) \mathbf{x}_{i, S_i} \mathbf{y}_{i, T_i} \right) \right] \\
&= \mathbb{E}_{\mathbf{x} \sim \bar{\nu}_1, \mathbf{y} \sim \bar{\nu}_2} \left[\left(\sum_{S^{[n]}, T^{[n]}} \widehat{\mathcal{C}}(S^{[n]}, T^{[n]}) \prod_{i \in [n]} \mathbf{x}_{i, S_i} \prod_{j \in [n]} \mathbf{y}_{j, T_j} \right) \left(\sum_{S^I, T^I} \prod_{i \in I} \widehat{g}(S_i, T_i) \mathbf{x}_{i, S_i} \mathbf{y}_{i, T_i} \right) \right] \\
&= \sum_{S^I, T^I} \widehat{\mathcal{C}}(S^I, T^I) \cdot \prod_{i \in I} \widehat{g}(S_i, T_i) \\
&= \sum_{\substack{S^I, T^I \\ S_i \neq \emptyset, T_i \neq \emptyset, \forall i \in I}} \widehat{\mathcal{C}}(S^I, T^I) \cdot \prod_{i \in I} \widehat{g}(S_i, T_i), \tag{by [Assumption 7.2](#)}
\end{aligned}$$

as desired. \square

Now we present the reduction from XOR-fiber to a general g -fiber.

Theorem 7.5. *Assume gadget $g: \{\pm 1\}^{m_1} \times \{\pm 1\}^{m_2} \rightarrow \{\pm 1\}$ satisfies [Assumption 7.2](#). Then*

$$\begin{aligned}
L_{1,k}(\text{XOR}, d, 1, 1, n) &\leq \left(\max_{S, T} |\widehat{g}(S, T)| \right)^{-k} \cdot L_{1,k}(g, d, m_1, m_2, n) \\
&\leq 2^{(m_1 + m_2) \cdot k/2} \cdot L_{1,k}(g, d, m_1, m_2, n).
\end{aligned}$$

Proof. Let $\mathcal{C}: \{\pm 1\}^n \times \{\pm 1\}^n \rightarrow [-1, 1]$ be an arbitrary protocol of cost at most d . Then for a fixed set $I \subseteq [n]$, by [Fact 7.4](#) applied to the XOR gadget, we have

$$\widehat{\mathcal{C}}_{\downarrow \text{XOR}}(I) = \widehat{\mathcal{C}}(1^I, 1^I). \tag{7.1}$$

Let $S \subseteq [m_1]$ and $T \subseteq [m_2]$ maximize $|\widehat{g}(S, T)|$. Since g satisfies [Assumption 7.2](#), we know S and T are not empty sets.

Now define a different protocol $\mathcal{C}' : (\{\pm 1\}^{m_1})^n \times (\{\pm 1\}^{m_2})^n \rightarrow [-1, 1]$ as follows: After receiving input x , Alice computes $x'_i = x_{i, S}$ for each block x_i ; and Bob computes similarly $y'_i = y_{i, T}$ upon

receiving input y . Then they execute the protocol \mathcal{C} on x' and y' . That is, $\mathcal{C}'(x, y) = \mathcal{C}(x', y')$. Therefore, for any $I \subseteq [n]$ and S^I, T^I satisfying $S_i \neq \emptyset, T_i \neq \emptyset$ for $i \in I$, we have

$$\widehat{\mathcal{C}'}(S^I, T^I) = \begin{cases} \widehat{\mathcal{C}}(1^I, 1^I) & S_i = S, T_i = T, \forall i \in I, \\ 0 & \text{otherwise.} \end{cases}$$

Then by (7.1) and Fact 7.4 applied to \mathcal{C}' with gadget g , we have

$$\widehat{\mathcal{C}'_{\downarrow g}}(I) = \widehat{\mathcal{C}}(1^I, 1^I) \cdot \widehat{g}(S, T)^{|I|} = \widehat{\mathcal{C}_{\downarrow \text{XOR}}}(I) \cdot \widehat{g}(S, T)^{|I|}.$$

Now summing over all $I \subseteq [n]$ of size k , we have

$$\begin{aligned} L_{1,k}(\mathcal{C}_{\downarrow \text{XOR}}) &= \sum_{I \subseteq [n]: |I|=k} \left| \widehat{\mathcal{C}_{\downarrow \text{XOR}}}(I) \right| = |\widehat{g}(S, T)|^{-k} \cdot \sum_{I \subseteq [n]: |I|=k} \left| \widehat{\mathcal{C}'_{\downarrow g}}(I) \right| = |\widehat{g}(S, T)|^{-k} \cdot L_{1,k}(\mathcal{C}'_{\downarrow g}) \\ &\leq |\widehat{g}(S, T)|^{-k} \cdot L_{1,k}(g, d, m_1, m_2, n). \end{aligned} \quad (\text{since } \mathcal{C}' \text{ has cost at most } d)$$

Since \mathcal{C} is arbitrary, this proves the first half of Theorem 7.5. To prove the second half, we use an averaging argument and Parseval's identity on g :

$$|\widehat{g}(S, T)| \geq \sqrt{2^{-m_1-m_2} \sum_{S', T'} \widehat{g}(S', T')^2} = \sqrt{2^{-m_1-m_2}}. \quad \square$$

Using similar analysis, we also have a reduction from a general g -fiber to XOR-fiber.

Theorem 7.6. *Assume gadget $g: \{\pm 1\}^{m_1} \times \{\pm 1\}^{m_2} \rightarrow \{\pm 1\}$ satisfies Assumption 7.2. Then*

$$\begin{aligned} L_{1,k}(g, d, m_1, m_2, n) &\leq \left(\sum_{S, T} |\widehat{g}(S, T)| \right)^k \cdot L_{1,k}(\text{XOR}, d, 1, 1, n) \\ &\leq 2^{(m_1+m_2) \cdot k/2} \cdot L_{1,k}(\text{XOR}, d, 1, 1, n). \end{aligned}$$

Proof. Let $\mathcal{C}: (\{\pm 1\}^{m_1})^n \times (\{\pm 1\}^{m_2})^n \rightarrow [-1, 1]$ be an arbitrary protocol of cost at most d . Then for a fixed set $I \subseteq [n]$, by Fact 7.4 applied to gadget g and using Assumption 7.2, we have

$$\widehat{\mathcal{C}_{\downarrow g}}(I) = \sum_{S^I, T^I} \widehat{\mathcal{C}}(S^I, T^I) \cdot \prod_{i \in I} \widehat{g}(S_i, T_i).$$

Therefore

$$L_{1,k}(\mathcal{C}_{\downarrow g}) \leq \sum_{I \subseteq [n]: |I|=k} \sum_{S^I, T^I} \left| \widehat{\mathcal{C}}(S^I, T^I) \right| \cdot \left| \prod_{i \in I} \widehat{g}(S_i, T_i) \right|.$$

Now let $M = \sum_{S, T} |\widehat{g}(S, T)|$. Let ρ be a distribution over subsets of $[m_1] \times [m_2]$ and its probability density function is defined as:

$$\rho(S, T) = |\widehat{g}(S, T)|/M.$$

Then we can rewrite $L_{1,k}(\mathcal{C}_{\downarrow g})$ as

$$L_{1,k}(\mathcal{C}_{\downarrow g}) \leq \sum_{I \subseteq [n]: |I|=k} \mathbb{E}_{(S^I, T^I) \sim \rho^I} \left[\left| \widehat{\mathcal{C}}(S^I, T^I) \right| \cdot M^k \right]$$

$$= M^k \cdot \mathbb{E}_{(S^{[n]}, T^{[n]}) \sim \rho^{[n]}} \left[\sum_{I \subseteq [n]: |I|=k} \left| \widehat{\mathcal{C}}(S^I, T^I) \right| \right]. \quad (7.2)$$

Now we fix an arbitrary $(S^{[n]}, T^{[n]})$ sampled from $\rho^{[n]}$. Note that S_i and T_i are not empty by the definition of ρ and [Assumption 7.2](#). Then define a different protocol $\mathcal{C}' : \{\pm 1\}^n \times \{\pm 1\}^n \rightarrow [-1, 1]$ as follows: After receiving input x , Alice samples $x' \in (\{\pm 1\}^{m_1})^n$ uniformly conditioned on $x'_{i, S_i} = x_i$ for all $i \in [n]$; and Bob samples similarly $y' \in (\{\pm 1\}^{m_2})^n$ conditioned on $y'_{i, T_i} = x_i$ for all $i \in [n]$. Then they execute the protocol \mathcal{C} on x' and y' . That is, $\mathcal{C}'(x, y) = \mathbb{E}_{x', y'}[\mathcal{C}(x', y')]$. Therefore, for any $I \subseteq [n]$, we have

$$\widehat{\mathcal{C}'}(1^I, 1^I) = \widehat{\mathcal{C}}(S^I, T^I).$$

By [Fact 7.4](#) applied to \mathcal{C}' and the XOR gadget, we have

$$\widehat{\mathcal{C}'_{\downarrow \text{XOR}}}(I) = \widehat{\mathcal{C}'}(1^I, 1^I) = \widehat{\mathcal{C}}(S^I, T^I).$$

Since \mathcal{C}' has cost at most d , we have

$$\sum_{I \subseteq [n]: |I|=k} \left| \widehat{\mathcal{C}}(S^I, T^I) \right| = \sum_{I \subseteq [n]: |I|=k} \left| \widehat{\mathcal{C}'_{\downarrow \text{XOR}}}(I) \right| = L_{1,k}(\mathcal{C}'_{\downarrow \text{XOR}}) \leq L_{1,k}(\text{XOR}, d, 1, 1, n).$$

Putting back to [\(7.2\)](#), we have

$$L_{1,k}(\mathcal{C}_{\downarrow g}) \leq M^k \cdot L_{1,k}(\text{XOR}, d, 1, 1, n),$$

which proves the first half of [Theorem 7.6](#) since \mathcal{C} is arbitrary. To prove the second half, we use Cauchy-Schwarz inequality and Parseval's identity on g :

$$M = \sum_{S, T} |\widehat{g}(S, T)| \leq \sqrt{2^{m_1+m_2} \sum_{S, T} \widehat{g}(S, T)^2} = \sqrt{2^{m_1+m_2}}. \quad \square$$

As a corollary, to study the Fourier growth bounds, we can switch between gadgets conveniently, as long as the gadgets have small size.

Corollary 7.7. *Assume gadgets $g : \{\pm 1\}^{m_1} \times \{\pm 1\}^{m_2} \rightarrow \{\pm 1\}$ and $g' : \{\pm 1\}^{m'_1} \times \{\pm 1\}^{m'_2} \rightarrow \{\pm 1\}$ satisfy [Assumption 7.2](#). Then*

$$L_{1,k}(g, d, m_1, m_2, n) \leq 2^{(m_1+m_2+m'_1+m'_2) \cdot k/2} \cdot L_{1,k}(g', d, m'_1, m'_2, n).$$

8 Directions Towards Further Improvements

In this section we propose potential directions for further improving our second level bounds. In [Subsection 8.1](#), we show that better Fourier growth bounds can be obtained from strong lifting theorems in a black-box way. This relies on the Fourier growth reductions in [Section 7](#). In [Subsection 8.2](#), we examine the bottleneck in our analysis and identify major obstacles within.

8.1 Better Lifting Theorems Imply Better Fourier Growth

Let $f : \{\pm 1\}^n \rightarrow \{\pm 1\}$ be a Boolean function. Let $g : \{\pm 1\}^{m_1} \times \{\pm 1\}^{m_2} \rightarrow \{\pm 1\}$ be a gadget. A lifting theorem connects the communication complexity of $f \circ g$ with the query complexity of f . Some lifting theorems show that a low-cost communication protocol can be simulated by a low-cost query algorithm.

To be more precise, let $\mathcal{C} : (\{\pm 1\}^{m_1})^n \times (\{\pm 1\}^{m_2})^n \rightarrow [-1, 1]$ be a randomized two-party protocol. Recall [Definition 7.1](#), the g -fiber of \mathcal{C} , denoted $\mathcal{C}_{\downarrow g}(z) : \{\pm 1\}^n \rightarrow [-1, 1]$, is defined by

$$\mathcal{C}_{\downarrow g}(z) = \mathbb{E}_{\mathbf{x} \sim \bar{\nu}_1, \mathbf{y} \sim \bar{\nu}_2} [\mathcal{C}(\mathbf{x}, \mathbf{y}) \mid g(\mathbf{x}_i, \mathbf{y}_i) = z_i, \forall i].$$

We say that g satisfies a strong lifting theorem if for all randomized protocols \mathcal{C} of small communication bits, there is a randomized decision tree of small depth that approximates $\mathcal{C}_{\downarrow g}$ on each input with error $1/\text{poly}(n)$ (see e.g., [\[GPW20\]](#)).

Theorem 8.1. *Assume gadget $g : \{\pm 1\}^{m_1} \times \{\pm 1\}^{m_2} \rightarrow \{\pm 1\}$ satisfies [Assumption 7.2](#). Assume for any randomized protocol $\mathcal{C} : (\{\pm 1\}^{m_1})^n \times (\{\pm 1\}^{m_2})^n \rightarrow [-1, 1]$ with at most d bits of communication, there exists a randomized decision tree \mathcal{T} of depth at most D that approximates $\mathcal{C}_{\downarrow g}$ with pointwise error at most $1/n^k$, i.e.,*

$$|\mathcal{T}(z) - \mathcal{C}_{\downarrow g}(z)| \leq n^{-k} \quad \forall z \in \{\pm 1\}^n.$$

Then, for any randomized protocol $\mathcal{C}' : \{\pm 1\}^n \times \{\pm 1\}^n \rightarrow [-1, 1]$ with at most d bits of communication, its XOR-fiber $\mathcal{C}'_{\downarrow \text{XOR}}$ has level- k Fourier growth

$$\begin{aligned} L_{1,k}(\mathcal{C}'_{\downarrow \text{XOR}}) &\leq \left(\max_{S,T} |\widehat{g}(S,T)| \right)^{-k} \cdot \sqrt{D^k \cdot O(\log(n))^{k-1}} \\ &\leq 2^{(m_1+m_2) \cdot k/2} \cdot \sqrt{D^k \cdot O(\log(n))^{k-1}}. \end{aligned}$$

As a simple corollary, we see that if the assumption of [Theorem 8.1](#) holds with $k = 2$, $D = d \cdot \text{polylog}(n)$, and a polylogarithmic-sized gadget g (i.e., $2^{m_1}, 2^{m_2} \leq \text{polylog}(n)$), then the second level Fourier growth of the XOR-fiber of any randomized protocol of cost d is at most $d \cdot \text{polylog}(n)$ as desired.

We also remark that state-of-the-art lifting results hold with the gadget g being either:

- The inner product on $m_1 = m_2 = O(\log(n))$ bits [\[CFK⁺19\]](#). However, for such g the largest Fourier coefficient squared is $1/\text{poly}(n)$, which yields a trivial bound in [Theorem 8.1](#).
- The index function with $m_1 = \text{poly}(n)$, $m_2 = \log(m_1)$ [\[GPW20\]](#).¹⁶ In this case the largest Fourier coefficient squared is $1/m_1^2$, which again yields a trivial bound in [Theorem 8.1](#). Nonetheless, even a polynomial improvement on m_1 , say $m_1 = n^{0.01}$, would give new non-trivial bounds in [Theorem 8.1](#) and in turn improves our lower bound on the XOR-lift of Forrelation.

Proof of [Theorem 8.1](#). Let $\mathcal{C} : (\{\pm 1\}^{m_1})^n \times (\{\pm 1\}^{m_2})^n \rightarrow [-1, 1]$ be a randomized protocol of cost at most d . Then by assumption, $\mathcal{C}_{\downarrow g}$ can be approximated up to error $1/n^k$ by a randomized decision tree \mathcal{T} of depth at most D . Thus any Fourier coefficient of $\mathcal{C}_{\downarrow g}$ and \mathcal{T} differs by at most

¹⁶For deterministic lifting, a better bound $m_1 = O(n \log(n))$ is known [\[LMM⁺22\]](#), but it doesn't suffice for our reduction.

$1/n^k$. Therefore by the level- k Fourier growth bounds on randomized decision trees [Tal20, SSW21], we have

$$L_{1,k}(\mathcal{C}_{\downarrow g}) \leq \sum_{S \subseteq [n]: |S|=k} \left(n^{-k} + \left| \widehat{\mathcal{J}}(S) \right| \right) \leq \sqrt{D^k \cdot O(\log(n))^{k-1}}.$$

Since \mathcal{C} is arbitrary, the claimed bound for $\mathcal{C}'_{\downarrow \text{XOR}}$ follows from Theorem 7.5. \square

8.2 Sums of Squares of Quadratic Forms for Pairwise Clean Sets

In our analysis for the level-two bound, we showed that one can transform a general protocol to a 4-wise clean protocol with parameter $\lambda = d \cdot \text{polylog}(n)$ by adding $O(d)$ additional cleanup steps in expectation. If one could show that with essentially the same number of steps, one could take $\lambda = \text{polylog}(n)$, then we would obtain the optimal level-two bound of $d \cdot \text{polylog}(n)$.

We recall that to bound the number of cleanup steps, we rely on a concentration inequality for sums of squares of orthonormal quadratic forms (Theorem 3.3), which says that if M_1, \dots, M_m are matrices with zero diagonal and form an orthonormal set when viewed as n^2 dimensional vectors, then the random variable $\mathbf{q} = \sum_{i=1}^m \left\langle \mathbf{x} \dot{\otimes} \mathbf{x}, M_i \right\rangle^2$ satisfies $\Pr_{\mathbf{x} \sim \gamma_n}[\mathbf{q} \geq t] \leq e^{-\Omega(\sqrt{t})}$ for any $t \gtrsim m^2$. Using this tail bound for $m = \Theta(d)$ and conditioning on $\mathbf{x} \in X$ where X is an arbitrary subset of \mathbb{R}^n with Gaussian measure $\approx 2^{-d}$, we obtained a bound $\mathbb{E}_{\mathbf{x} \sim \gamma}[\mathbf{q} \mid \mathbf{x} \in X] \lesssim d^2$. This shows that there can be at most $O(d)$ such quadratic forms M_i 's where the value $\mathbb{E}_{\mathbf{x} \sim \gamma} \left[\left\langle \mathbf{x} \dot{\otimes} \mathbf{x}, M_i \right\rangle^2 \mid \mathbf{x} \in X \right]$ can be larger than d and hence, the reason we can only take $\lambda \approx d$. We note that the argument just described is for the non-adaptive setting, while in our case the M_i 's are also being chosen adaptively, so additional work is needed.

The next example shows that the aforementioned statement is tight even in the non-adaptive setting where the M_i 's are fixed: in particular, there is a set X of large measure and $\approx d$ such orthonormal quadratic forms where the above expectation after conditioning on $\mathbf{x} \in X$ is $\Theta(d^2)$.

Example 8.2. For $1 \leq i < j \leq \sqrt{d}$, let $M_{ij} = E_{ij}$ for $i < j$ where E_{ij} denotes the $n \times n$ matrix where only the (i, j) entry is one. Note that the matrices M_{ij} form an orthonormal set and they all have a zero diagonal. Let $X = \{x \in \mathbb{R}^n \mid |x_i| \gtrsim d^{1/4} \text{ for all } i \leq d^{1/2}\}$. Then, the Gaussian measure $\gamma(X) = 2^{-\Theta(d)}$ but

$$\mathbb{E}_{\mathbf{x} \sim \gamma} \left[\sum_{1 \leq i < j \leq \sqrt{d}} \left\langle \mathbf{x} \dot{\otimes} \mathbf{x}, M_{ij} \right\rangle^2 \mid \mathbf{x} \in X \right] = \Theta(d^2).$$

Note that the set X in the example above is not pairwise clean and for our application, one can get around it by first ensuring that the protocol is pairwise clean and then proceeding with the 4-wise cleanup process. Motivated by this, we speculate that when the set is pairwise clean, then the expected value of the sum of squares of orthonormal quadratic forms is much smaller unlike the example above. Assuming such a statement and combining it with our ideas for handling the adaptivity suggests a potential way of improving the level-two bounds.

References

- [AA18] Scott Aaronson and Andris Ambainis. Forrelation: A problem that optimally separates quantum from classical computing. *SIAM J. Comput.*, 47(3):982–1038, 2018. 6
- [Aar10] Scott Aaronson. BQP and the polynomial hierarchy. In *STOC*, pages 141–150, 2010.

- [ABK23] Scott Aaronson, Harry Buhrman, and William Kretschmer. A qubit, a coin, and an advice string walk into a relational problem. *arXiv preprint arXiv:2302.10332*, 2023. 6
- [Agr20] Rohit Agrawal. Coin theorems and the fourier expansion. *Chic. J. Theor. Comput. Sci.*, 2020, 2020. 1, 4
- [ALM20] Radosław Adamczak, Rafał Latała, and Rafał Meller. Hanson–wright inequality in banach spaces. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 56(4), nov 2020. 14, 16, 60
- [BCW98] Harry Buhrman, Richard Cleve, and Avi Wigderson. Quantum vs. classical communication and computation. In *STOC*, pages 63–68. ACM, 1998. 5
- [BIJ⁺21] Jarosław Błasiok, Peter Ivanov, Yaonan Jin, Chin Ho Lee, Rocco A Servedio, and Emanuele 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. 1
- [Bor75] Christer Borell. The brunn-minkowski inequality in gauss space. *Inventiones mathematicae*, 30(2):207–216, 1975. 60
- [BS21] Nikhil Bansal and Makrand Sinha. k-forrelation optimally separates quantum and classical query complexity. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1303–1316, 2021. 6, 7
- [BTW15] Eric Blais, Li-Yang Tan, and Andrew Wan. An inequality for the fourier spectrum of parity decision trees. *CoRR*, abs/1506.01055, 2015. 1, 2
- [BV10] Joshua Brody and Elad Verbin. The coin problem and pseudorandomness for branching programs. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 30–39, 2010. 4
- [CFK⁺19] Arkadev Chattopadhyay, Yuval Filmus, Sajin Koroth, Or Meir, and Toniann Pitassi. Query-to-communication lifting for bpp using inner product. In *ICALP*, 2019. 5, 6, 53
- [CGR14] Gil Cohen, Anat Ganor, and Ran 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. 4
- [CHHL19] Eshan Chattopadhyay, Pooya Hatami, Kaave Hosseini, and Shachar Lovett. Pseudorandom generators from polarizing random walks. *Theory Comput.*, 15:1–26, 2019. 1, 8
- [CHLT18] Eshan Chattopadhyay, Pooya Hatami, Shachar Lovett, and Avishay 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
- [CHRT18] Eshan Chattopadhyay, Pooya Hatami, Omer Reingold, and Avishay Tal. Improved pseudorandomness for unordered branching programs through local monotonicity. In *STOC*, pages 363–375. ACM, 2018. 1

- [CKLM19] Arkadev Chattopadhyay, Michal Koucký, Bruno Loff, and Sagnik Mukhopadhyay. Simulation theorems via pseudo-random properties. *Comput. Complex.*, 28(4):617–659, 2019. 5
- [CR12] Amit Chakrabarti and Oded Regev. An optimal lower bound on the communication complexity of gap-hamming-distance. *SIAM J. Comput.*, 41(5):1299–1317, 2012. 5, 12
- [dRNV16] Susanna F. de Rezende, Jakob Nordström, and Marc Vinyals. How limited interaction hinders real communication (and what it means for proof and circuit complexity). In *FOCS*, pages 295–304. IEEE Computer Society, 2016. 5
- [EM22] Ronen Eldan and Dana Moshkovitz. Reduction from non-unique games to boolean unique games. In *13th Innovations in Theoretical Computer Science Conference (ITCS 2022)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2022. 16
- [Gav20] Dmitry Gavinsky. Entangled simultaneity versus classical interactivity in communication complexity. *IEEE Trans. Inf. Theory*, 66(7):4641–4651, 2020. 6
- [GKPW19] Mika Göös, Pritish Kamath, Toniann Pitassi, and Thomas Watson. Query-to-communication lifting for P NP. *Comput. Complex.*, 28(1):113–144, 2019. 5
- [GLM⁺15] Mika Göös, Shachar Lovett, Raghu Meka, Thomas Watson, and David Zuckerman. Rectangles are nonnegative juntas. In *STOC*, pages 257–266. ACM, 2015. 5
- [Göo15] Mika Göös. Lower bounds for clique vs. independent set. In *FOCS*, pages 1066–1076. IEEE Computer Society, 2015. 5
- [GPW15] Mika Göös, Toniann Pitassi, and Thomas Watson. Deterministic communication vs. partition number. In *FOCS*, pages 1077–1088. IEEE Computer Society, 2015. 5
- [GPW20] Mika Göös, Toniann Pitassi, and Thomas Watson. Query-to-communication lifting for BPP. *SIAM J. Comput.*, 49(4), 2020. 5, 7, 53
- [GRT21] Uma Girish, Ran Raz, and Avishay Tal. Quantum versus randomized communication complexity, with efficient players. In *ITCS*, volume 185 of *LIPICs*, pages 54:1–54:20, 2021. Presented in QIP, 2020 as a contributed talk. 1, 3, 4, 6, 39
- [GRZ21] Uma Girish, Ran Raz, and Wei Zhan. Lower bounds for xor of forrelations. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, 2021. 1, 3
- [GSTW16] Parikshit Gopalan, Rocco A. Servedio, Avishay Tal, and Avi Wigderson. Degree and sensitivity: tails of two distributions. *CoRR*, abs/1604.07432, 2016. 1
- [GTW21] Uma Girish, Avishay Tal, and Kewen Wu. Fourier growth of parity decision trees. In *36th Computational Complexity Conference (CCC 2021)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2021. 1, 2, 3, 7
- [HHL18] Hamed Hatami, Kaave Hosseini, and Shachar Lovett. Structure of protocols for XOR functions. *SIAM J. Comput.*, 47(1):208–217, 2018. 2, 5, 8
- [INW94] Russell Impagliazzo, Noam Nisan, and Avi Wigderson. Pseudorandomness for network algorithms. In *STOC*, pages 356–364. ACM, 1994. 8

- [IRR⁺21] Siddharth Iyer, Anup Rao, Victor Reis, Thomas Rothvoss, and Amir Yehudayoff. Tight bounds on the fourier growth of bounded functions on the hypercube. *arXiv preprint arXiv:2107.06309*, 2021. [1](#)
- [KKL88] Jeff Kahn, Gil Kalai, and Nathan Linial. The influence of variables on boolean functions (extended abstract). In *29th Annual Symposium on Foundations of Computer Science, White Plains, New York, USA, 24-26 October 1988*, pages 68–80. IEEE Computer Society, 1988. [16](#)
- [KMR17] Pravesh K. Kothari, Raghu Meka, and Prasad Raghavendra. Approximating rectangles by juntas and weakly-exponential lower bounds for LP relaxations of csps. In *STOC*, pages 590–603. ACM, 2017. [5](#)
- [Lee19] Chin Ho Lee. Fourier bounds and pseudorandom generators for product tests. In Amir Shpilka, editor, *34th Computational Complexity Conference, CCC 2019, July 18-20, 2019, New Brunswick, NJ, USA*, volume 137 of *LIPICs*, pages 7:1–7:25. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019. [1](#)
- [LMM⁺22] Shachar Lovett, Raghu Meka, Ian Mertz, Toniann Pitassi, and Jiapeng Zhang. Lifting with sunflowers. In *13th Innovations in Theoretical Computer Science Conference (ITCS 2022)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2022. [5](#), [53](#)
- [LPV22] Chin Ho Lee, Edward Pyne, and Salil P. Vadhan. Fourier growth of regular branching programs. In Amit Chakrabarti and Chaitanya Swamy, editors, *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)*, volume 245 of *LIPICs*, pages 2:1–2:21. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022. [1](#)
- [LRS15] James R. Lee, Prasad Raghavendra, and David Steurer. Lower bounds on the size of semidefinite programming relaxations. In *STOC*, pages 567–576. ACM, 2015. [5](#)
- [LSS⁺19] Nutan Limaye, KartEEK Sreenivasaiah, Srikanth Srinivasan, Utkarsh 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*, pages 442–453, 2019. [4](#)
- [LV18] Chin Ho Lee and Emanuele Viola. The coin problem for product tests. *ACM Transactions on Computation Theory (TOCT)*, 10(3):1–10, 2018. [4](#)
- [Man95] Yishay Mansour. An $O(n^{\log \log n})$ learning algorithm for DNF under the uniform distribution. *J. Comput. Syst. Sci.*, 50(3):543–550, 1995. Appeared in COLT, 1992. [1](#)
- [MO10] Ashley Montanaro and Tobias Osborne. On the communication complexity of xor functions, 2010. [2](#)
- [O’D14] Ryan O’Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. [16](#)
- [OS07] Ryan O’Donnell and Rocco A. Servedio. Learning monotone decision trees in polynomial time. *SIAM Journal on Computing*, 37(3):827–844, 2007. [1](#), [2](#)

- [Raz87] Alexander 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*, 41(4):333–338, 1987. 1
- [Raz95] Ran Raz. Fourier analysis for probabilistic communication complexity. *Comput. Complex.*, 5(3/4):205–221, 1995. 2
- [RM99] Ran Raz and Pierre McKenzie. Separation of the monotone NC hierarchy. *Comb.*, 19(3):403–435, 1999. 5
- [RPRC16] Robert Robere, Toniann Pitassi, Benjamin Rossman, and Stephen A. Cook. Exponential lower bounds for monotone span programs. In *FOCS*, pages 406–415. IEEE Computer Society, 2016. 5
- [RS10] Alexander A. Razborov and Alexander A. Sherstov. The sign-rank of ac^0 . *SIAM J. Comput.*, 39(5):1833–1855, 2010. 5
- [RSV13] Omer Reingold, Thomas Steinke, and Salil P. Vadhan. Pseudorandomness for regular branching programs via Fourier analysis. In *APPROX-RANDOM*, pages 655–670. Springer, 2013. 1
- [RT19] Ran Raz and Avishay Tal. Oracle separation of BQP and PH. In *STOC*, pages 13–23. ACM, 2019. Presented in QIP, 2019 as a plenary talk. Accepted to the Journal of the ACM. 1
- [RY22] Anup Rao and Amir Yehudayoff. Anticoncentration and the exact gap-hamming problem. *SIAM Journal on Discrete Mathematics*, 36(2):1071–1092, 2022. 5
- [She11] Alexander A. Sherstov. The pattern matrix method. *SIAM J. Comput.*, 40(6):1969–2000, 2011. 5
- [She12] Alexander A. Sherstov. The communication complexity of gap hamming distance. *Theory Comput.*, 8(1):197–208, 2012. 5
- [Smo87] Roman Smolensky. Algebraic methods in the theory of lower bounds for boolean circuit complexity. In Alfred V. Aho, editor, *Proceedings of the 19th Annual ACM Symposium on Theory of Computing, 1987, New York, New York, USA*, pages 77–82. ACM, 1987. 1
- [SSW21] Alexander A Sherstov, Andrey A Storozhenko, and Pei Wu. An optimal separation of randomized and quantum query complexity. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1289–1302, 2021. 1, 2, 6, 7, 54
- [ST78] Vladimir N Sudakov and Boris S Tsirel’son. Extremal properties of half-spaces for spherically invariant measures. *Journal of Soviet Mathematics*, 9(1):9–18, 1978. 60
- [SVW17] Thomas Steinke, Salil P. Vadhan, and Andrew Wan. Pseudorandomness and Fourier-growth bounds for width-3 branching programs. *Theory of Computing*, 13(1):1–50, 2017. Appeared in APPROX-RANDOM, 2014. 1
- [SZ08] Yaoyun Shi and Zhiqiang Zhang. Communication complexities of xor functions. *arXiv preprint arXiv:0808.1762*, 2008. 2

- [SZ09] Yaoyun Shi and Yufan Zhu. Quantum communication complexity of block-composed functions. *Quantum Inf. Comput.*, 9(5&6):444–460, 2009. [5](#)
- [Tal96] Michel Talagrand. How much are increasing sets positively correlated? *Comb.*, 16(2):243–258, 1996. [16](#)
- [Tal17] Avishay Tal. Tight bounds on the Fourier spectrum of AC0. In *Computational Complexity Conference*, volume 79 of *LIPICs*, pages 15:1–15:31. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2017. [1](#)
- [Tal20] Avishay Tal. Towards optimal separations between quantum and randomized query complexities. In *FOCS*, pages 228–239. IEEE, 2020. [1](#), [2](#), [7](#), [54](#)
- [TWXZ13] Hing Yin Tsang, Chung Hoi Wong, Ning Xie, and Shengyu Zhang. Fourier sparsity, spectral norm, and the log-rank conjecture. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 658–667, 2013. [2](#)
- [Ver18] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018. [16](#)
- [Vid12] Thomas 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.*, 2012, 2012. [5](#), [12](#)
- [Wu22] Xinyu Wu. A stochastic calculus approach to the oracle separation of BQP and PH. *Theory Comput.*, 18:1–11, 2022. [1](#)
- [WYY17] Xiaodi Wu, Penghui Yao, and Henry S. Yuen. Raz-mckenzie simulation with the inner product gadget. *Electron. Colloquium Comput. Complex.*, 24:10, 2017. [5](#)
- [Zha14] Shengyu Zhang. Efficient quantum protocols for xor functions. In *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*, pages 1878–1885. SIAM, 2014. [2](#)

A Gap-Hamming Lower Bounds

As an immediate consequence of [Theorem 1.5](#), we can derive optimal lower bounds against the Gap-Hamming problem as in [Theorem 1.6](#).

Proof of [Theorem 1.6](#). Set $\rho = 10/\sqrt{n}$. Fix the randomness to be any $r \in \{0, 1\}^*$ and let \mathcal{C}_r refer to the deterministic protocol \mathcal{C} with randomness fixed to r . Suppose $d \leq \tau \cdot n$ for a sufficiently small constant τ , we apply [Theorem 1.5](#) on ρ as well as $-\rho$, and apply triangle inequality to conclude that

$$\left| \mathbb{E}_{\mathbf{z} \sim \pi_{\rho}^{\otimes n}} [h_r(\mathbf{z})] - \mathbb{E}_{\mathbf{z} \sim \pi_{-\rho}^{\otimes n}} [h_r(\mathbf{z})] \right| \leq 2 \cdot O\left(\sqrt{d/n}\right) < 1/9.$$

Let σ_{ρ} be the distribution of (\mathbf{x}, \mathbf{y}) induced by sampling $\mathbf{x} \sim \pi_0^{\otimes n}$ and $\mathbf{z} \sim \pi_{\rho}^{\otimes n}$ and letting $\mathbf{y} = \mathbf{x} \odot \mathbf{z}$, similarly define $\sigma_{-\rho}$ but with $\mathbf{z} \sim \pi_{-\rho}^{\otimes n}$. We now expand $h_r(\mathbf{z})$ in terms of $\mathcal{C}(x, y)$, take an expectation over r and apply triangle inequality to conclude that

$$\left| \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \sigma_{\rho}} [\mathcal{C}(\mathbf{x}, \mathbf{y})] - \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \sigma_{-\rho}} [\mathcal{C}(\mathbf{x}, \mathbf{y})] \right| < 1/9. \tag{A.1}$$

Hoeffding's inequality implies that for $\mathbf{z} \sim \pi_\rho^{\otimes n}$, we have

$$\Pr \left[\left| \sum_i z_i - 10\sqrt{n} \right| \geq 5\sqrt{n} \right] \leq 2 \exp \left\{ \frac{-2 \cdot (5\sqrt{n})^2}{4n} \right\} < 1/18.$$

This implies that a random $(\mathbf{x}, \mathbf{y}) \sim \sigma_\rho$ is a YES instance of the Gap-Hamming problem with probability larger than 17/18. Let $\tilde{\sigma}_\rho$ denote σ_ρ conditioned on YES instances of the Gap-Hamming problem. Similarly define $\tilde{\sigma}_{-\rho}$ to be $\sigma_{-\rho}$ conditioned on NO instances of the Gap-Hamming problem. Since $\mathcal{C}(x, y)$ has outputs in $[-1, 1]$, we have

$$\left| \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \sigma_\rho} [\mathcal{C}(\mathbf{x}, \mathbf{y})] - \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \tilde{\sigma}_\rho} [\mathcal{C}(\mathbf{x}, \mathbf{y})] \right| < 1/9$$

and

$$\left| \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \sigma_{-\rho}} [\mathcal{C}(\mathbf{x}, \mathbf{y})] - \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \tilde{\sigma}_{-\rho}} [\mathcal{C}(\mathbf{x}, \mathbf{y})] \right| < 1/9.$$

This, along with (A.1) and triangle inequality, implies that

$$\left| \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \tilde{\sigma}_\rho} [\mathcal{C}(\mathbf{x}, \mathbf{y})] - \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \tilde{\sigma}_{-\rho}} [\mathcal{C}(\mathbf{x}, \mathbf{y})] \right| < 1/3.$$

However, this contradicts the assumption that the protocol \mathcal{C} solves the Gap-Hamming problem with advantage at least 2/3. \square

B Concentration for Sum of Squares of Quadratic Forms

Here we prove [Theorem 3.3](#). While it follows from [[ALM20](#), Theorem 6] which is a Banach space-valued version of the Hanson-Wright inequality, in our setting a weaker statement suffices, for which we give a self-contained proof following [[ALM20](#)].

For any integer $n \geq 1$, we use $\mathcal{B}^n = \{x \in \mathbb{R}^n \mid \|x\| \leq 1\}$ to denote the unit Euclidean ball in \mathbb{R}^n . For any two sets $A, B \subseteq \mathbb{R}^n$, we define $A + B = \{x + y \mid x \in A, y \in B\}$. For any set $A \subseteq \mathbb{R}^n$ and any number $t \in \mathbb{R}$, we define $tA = \{t \cdot x \mid x \in A\}$. Let $\Phi: \mathbb{R} \rightarrow [0, 1]$ be the cumulative distribution function of the standard Gaussian distribution, i.e., $\Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-u^2/2} du$.

Now we cite the famous Gaussian isoperimetric inequality [[Bor75](#), [ST78](#)].

Theorem B.1 (Gaussian Isoperimetric Inequality). *Let $A \subseteq \mathbb{R}^n$ be a measurable set and assume $\gamma_n(A) \geq \Phi(a)$ for some $a \in \mathbb{R}$. Then for any $t \geq 0$, we have $\gamma_n(A + t\mathcal{B}^n) \geq \Phi(a + t)$.*

In particular, if $\gamma_n(A) \geq 1/2$, then we can pick $a = 0$ in [Theorem B.1](#) and have

$$\gamma_n(A + t\mathcal{B}^n) \geq \Phi(t) \geq 1 - e^{-t^2/2}. \tag{B.1}$$

Now we are ready to prove [Theorem 3.3](#).

Proof of [Theorem 3.3](#). Note that the bound is trivial when $m = 0$. Thus from now on we assume without loss of generality $m \geq 1$.

For each $x \in \mathbb{R}^n$, let $K_x = \sum_{i=1}^m \left\langle x \dot{\otimes} x, M_i \right\rangle^2$. We first write K_x as a squared Euclidean norm of a vector:

- For $i \in [m]$, we view M_i as a length- n^2 row vector.

- Let $M \in \mathbb{R}^{m \times n^2}$ be a matrix where the i -th row is M_i .

Therefore we have

$$K_x = \left\| M(x \otimes x) \right\|^2 = \|M(x \otimes x)\|^2, \quad (\text{B.2})$$

where \otimes is the standard tensor product and the second equality follows since each M_i has zero diagonal.

Define $f(y) = \|M(y \otimes y)\|$, $g(y) = \sup_{z \in \mathbb{S}^{n-1}} \|M(z \otimes y)\|$, and $h(y) = \sup_{z \in \mathbb{S}^{n-1}} \|M(y \otimes z)\|$. Let $F = \mathbb{E}_{\mathbf{y} \sim \gamma_n}[f(\mathbf{y})]$, $G = \mathbb{E}_{\mathbf{y} \sim \gamma_n}[g(\mathbf{y})]$, and $H = \mathbb{E}_{\mathbf{y} \sim \gamma_n}[h(\mathbf{y})]$ be their mean. Define the set

$$A = \{y \in \mathbb{R}^n \mid f(y) < 6F, g(y) < 6G, \text{ and } h(y) < 6H\}.$$

By Markov's inequality and union bound, we have the Gaussian measure of A is $\gamma_n(A) \geq 1/2$. Then by (B.1), we have

$$\gamma_n(A + t\mathcal{B}^n) \geq 1 - e^{-t^2/2} \quad \text{holds for all } t \geq 0. \quad (\text{B.3})$$

Now for an arbitrary $x \in A + t\mathcal{B}^n$, we write $x = y + tz$ where $y \in A$ and $z \in \mathcal{B}^n$. Then

$$\begin{aligned} \|M(x \otimes x)\| &\leq \|M(y \otimes y)\| + t \cdot \|M(y \otimes z)\| + t \cdot \|M(z \otimes y)\| + t^2 \cdot \|M(z \otimes z)\| \\ &< 6F + 6t(G + H) + t^2V, \end{aligned}$$

where $V = \sup_{z \in \mathbb{S}^{n-1}} \|M(z \otimes z)\|$. This, together with (B.2) and (B.3), implies

$$\Pr_{\mathbf{x} \sim \gamma_n} \left[K_{\mathbf{x}} \geq (6F + 6t(G + H) + t^2V)^2 \right] \leq \Pr_{\mathbf{x} \sim \gamma_n} [\mathbf{x} \notin A + t\mathcal{B}^n] = 1 - \gamma_n(A + t\mathcal{B}^n) \leq e^{-t^2/2}. \quad (\text{B.4})$$

Now we calculate F, G, H, V in the following claim, the proof of which will be presented later.

Claim B.2. $F \leq \sqrt{2m}$, $G, H \leq \sqrt{m}$, and $V \leq 1$.

Plugging Claim B.2 into (B.4), we have

$$\Pr_{\mathbf{x} \sim \gamma_n} \left[K_{\mathbf{x}} \geq \left(6\sqrt{2m} + 12t\sqrt{m} + t^2 \right)^2 \right] \leq e^{-t^2/2} \quad \text{holds for any } t \geq 0.$$

Now we set

$$t = \frac{1}{168} \sqrt{\frac{r}{m + \sqrt{r}}} \geq 0$$

and assume $r \geq 98m$. Then $6\sqrt{2m} \leq \frac{6}{7}\sqrt{r}$, $12t\sqrt{m} \leq \frac{1}{14}\sqrt{r}$, and $t^2 \leq \frac{1}{14}\sqrt{r}$. Therefore

$$\Pr_{\mathbf{x} \sim \gamma_n} \left[\sum_{i=1}^m \langle \mathbf{x} \otimes \mathbf{x}, M_i \rangle^2 \geq r \right] = \Pr_{\mathbf{x} \sim \gamma_n} [K_{\mathbf{x}} \geq r] \leq e^{-t^2/2} = \exp \left\{ -\frac{1}{56448} \cdot \frac{r}{m + \sqrt{r}} \right\}. \quad \square$$

Finally we present the missing proof of Claim B.2.

Proof of Claim B.2. First we observe that rows of M are unit vectors, therefore

$$\|M\| = \sqrt{m}. \quad (\text{B.5})$$

In addition, rows of M are orthogonal to each other, therefore the operator norm of M is

$$\|M\|_{\text{op}} \leq 1. \quad (\text{B.6})$$

We index the columns of M by $[n]^2$ and let the column vectors of M be $(b_{i,j})_{i,j \in [n]}$. Since rows of M are flattened matrices with zero diagonal, we have

$$b_{i,i} = 0^m \quad \text{for all } i \in [n]. \quad (\text{B.7})$$

Now we bound F, G, H, V separately.

Bounding F . Observe that

$$\begin{aligned}
F^2 &= \left(\mathbb{E}_{\mathbf{y} \sim \gamma_n} [\|M(\mathbf{y} \otimes \mathbf{y})\|] \right)^2 \leq \mathbb{E}_{\mathbf{y} \sim \gamma_n} [\|M(\mathbf{y} \otimes \mathbf{y})\|^2] = \mathbb{E}_{\mathbf{y} \sim \gamma_n} \left[\left\| \sum_{i,j \in [n]} b_{i,j} \mathbf{y}_i \mathbf{y}_j \right\|^2 \right] \quad (\text{by convexity}) \\
&= \mathbb{E}_{\mathbf{y} \sim \gamma_n} \left[\sum_{i,j,i',j' \in [n]} \langle b_{i,j}, b_{i',j'} \rangle \mathbf{y}_i \mathbf{y}_j \mathbf{y}_{i'} \mathbf{y}_{j'} \right] = \sum_{i,j \in [n]} \left(\|b_{i,j}\|^2 + \langle b_{i,j}, b_{j,i} \rangle \right) \quad (\text{by (B.7)}) \\
&\leq \sum_{i,j \in [n]} \left(\|b_{i,j}\|^2 + \frac{1}{2} (\|b_{i,j}\|^2 + \|b_{j,i}\|^2) \right) = 2 \sum_{i,j \in [n]} \|b_{i,j}\|^2 \\
&= 2 \|M\|^2 = 2m. \quad (\text{by (B.5)})
\end{aligned}$$

Bounding G and H . Fix an arbitrary $\mathbf{y} \in \mathbb{R}^n$ and we first simplify $g(\mathbf{y})$. For each $i \in [n]$, define vector $b_i = \sum_{j \in [n]} b_{i,j} \mathbf{y}_j$ and let B be the matrix with b_i 's as column vectors. Then

$$g(\mathbf{y}) = \sup_{z \in \mathbb{S}^{n-1}} \left\| \sum_{i,j \in [n]} b_{i,j} z_i \mathbf{y}_j \right\| = \sup_{z \in \mathbb{S}^{n-1}} \left\| \sum_{i \in [n]} b_i z_i \right\| = \|B\|_{\text{op}} \leq \|B\| = \sqrt{\sum_{i \in [n]} \left\| \sum_{j \in [n]} b_{i,j} \mathbf{y}_j \right\|^2}. \quad (\text{B.8})$$

Now we bound G :

$$\begin{aligned}
G^2 &= \left(\mathbb{E}_{\mathbf{y} \sim \gamma_n} [g(\mathbf{y})] \right)^2 \leq \mathbb{E}_{\mathbf{y} \sim \gamma_n} [g(\mathbf{y})^2] \quad (\text{by convexity}) \\
&\leq \mathbb{E}_{\mathbf{y} \sim \gamma_n} \left[\sum_{i \in [n]} \left\| \sum_{j \in [n]} b_{i,j} \mathbf{y}_j \right\|^2 \right] = \mathbb{E}_{\mathbf{y} \sim \gamma_n} \left[\sum_{i \in [n]} \sum_{j,j' \in [n]} \langle b_{i,j}, b_{i,j'} \rangle \mathbf{y}_j \mathbf{y}_{j'} \right] \quad (\text{by (B.8)}) \\
&= \sum_{i,j \in [n]} \|b_{i,j}\|^2 = \|M\|^2 = m. \quad (\text{by (B.5)})
\end{aligned}$$

Similar argument works for H .

Bounding V . Note that for any $z \in \mathbb{S}^{n-1}$, we have $\|z \otimes z\| = \|z\|^2 = 1$. Thus, by (B.6), we have

$$V = \sup_{z \in \mathbb{S}^{n-1}} \|M(z \otimes z)\| \leq \|M\|_{\text{op}} \leq 1. \quad \square$$