Quantum Advantages via Fourier Growth Analysis of Boolean Functions

Kewen Wu

Electrical Engineering and Computer Sciences
University of California, Berkeley

Technical Report No. UCB/EECS-2025-39
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2025/EECS-2025-39.html

May 1, 2025

Acknowledgement

My first and foremost thanks go to my advisor Avishay Tal. I also want to thank Jelani Nelson, John Wright, Shirshendu Ganguly, and Ruixiang Zhang for joining my qualifying exam and dissertation committees. Additionally, I was extremely lucky to have many amazing research collaborators and I am grateful for all the inspiring discussions with them. Please find the full acknowledgement in the attached report.

Quantum Advantages via Fourier Growth Analysis of Boolean Functions

By

Kewen Wu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Computer Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Assistant Professor Avishay Tal, Chair
Assistant Professor John Wright
Assistant Professor Ruixiang Zhang

Spring 2025

Quantum Advantages via Fourier Growth Analysis of Boolean Functions

Abstract

Quantum Advantages via Fourier Growth Analysis of Boolean Functions

by

Kewen Wu

Doctor of Philosophy in Computer Science

University of California, Berkeley

Assistant Professor Avishay Tal, Chair

Intuitively quantum computations can easily aggregate signals on the Fourier basis whereas classical algorithms cannot. Raz-Tal (STOC'19, JACM'22) and Bansal-Sinha (STOC'21) formalized this using the notion of Fourier growth from Boolean function analysis. Continuing this line of work, we establish sharp Fourier growth bounds for classical (parity) query algorithms, randomized communication protocols, and parallel quantum query algorithms with limited adaptivity. As such, we obtain unconditional exponential quantum advantages in various query and communication settings.

Contents

Contents i

List of Figures ii

1 Introduction 1
- 1.1 Analysis of Boolean Functions . . . 2
- 1.2 Fourier Growth . . . 3
- 1.3 Quantum Advantages and Forrelation Problem . . . 4
- 1.4 Our Contribution . . . 5

2 Preliminaries 7

3 Quantum Advantages over Parity Query 14
- 3.1 Introduction . . . 15
- 3.2 Our Results . . . 16
- 3.3 Proof Overview . . . 19
- 3.4 Adaptive Azuma's Inequality . . . 24
- 3.5 How to Clean Up Parity Decision Trees . . . 27
- 3.6 Fourier Growth Bounds for Parity Decision Trees . . . 33
- 3.7 Fourier Growth Bounds for Noisy Decision Trees . . . 46

4 Quantum Advantages over Classical Communication 51
- 4.1 Introduction . . . 52
- 4.2 Our Results . . . 54
- 4.3 Proof Overview . . . 60
- 4.4 Concentration for Sum of Squares of Quadratics . . . 68
- 4.5 Fourier Growth via Martingales in Gaussian Space . . . 71
- 4.6 Level-One Fourier Growth . . . 75
- 4.7 Level-Two Fourier Growth . . . 92
- 4.8 Fourier Growth Reductions For General Gadgets . . . 108
- 4.9 Directions Towards Further Improvements . . . 113

5 Power of Adaptivity in Quantum Query 116

5.1 Introduction . . . 116
5.2 Our Results . . . 118
5.3 Proof Overview . . . 121
5.4 Fourier Growth of the Quantum Query Model . . . 132
5.5 Tightness of the Non-Adaptive Case . . . 145
5.6 Quantum Query with Classical Preprocessing . . . 146

6 Other Projects **149**

Bibliography **154**

List of Figures

3.1 An example of the cleanup process for parity decision tree . . . 32

5.1 Quantum algorithm with r adaptive rounds of t parallel queries each. . . . 122
5.2 The constraint $\mathsf{Size}^s(\oplus I, \oplus J) = 1$. . . . 125
5.3 Expressing $g(s)$ as $u^\dagger W R' v$. . . . 125
5.4 Expressing $g(s)$ as $u^\dagger R W' v$. . . . 127
5.5 The constraint $\mathsf{Size}^s(\oplus I, \oplus J) = 1$. . . . 127
5.6 Expressing $g(s)$ as $u^\dagger R_1 W_2 R_2' v$. . . . 128
5.7 Expressing $g(s)$ as $u^\dagger W_1 R_1' R_2' v$ and $u^\dagger R_1 R_2 W_2 v$ respectively. . . . 129
5.8 The summation in $g(s)$ versus $h(s)$. . . . 130
5.9 Expressing $h(s) = u^\dagger Q_1 W_1 Q_1' R_1' Q_2' R_2' Q_3' v$. . . . 131

Acknowledgments

My first and foremost thanks go to my advisor Avishay Tal. I first met Avishay at the STOC'19 poster session. In retrospect, I probably asked some stupid questions and I even forgot to congratulate him for winning the STOC'19 best paper award. But apparently, Avishay didn't care about all these and still gave me a PhD offer afterward. Due to COVID, the first year of my PhD was remote in China. It was challenging due to the time difference, but Avishay was always supportive in arranging meetings and financial help. As a fresh grad student, I didn't know what research problems were promising and doable (and I don't dare to say that I know now). So the first few projects were Avishay's proposal and it was a smooth transition for me from undergrad to grad school. These early works made up my thesis, which freed me from worrying about graduation and allowed me to explore other things. Indeed, thanks to Avishay, I was maximally free on collaboration, research, and internship: I worked on algorithm design, differential privacy, parameterized complexity, quantum computing, cryptography, ..., while being his student on Boolean function analysis. In my fourth and fifth years, I was even mostly not in Berkeley but had lots of travels and stayed largely with my wife in San Diego. During these days, Avishay became more hands-off and trusted me in planning my own life. There are so many stories to share about Avishay's mentoring/help/encouragement regarding research/life/career, but I have limited space. So let me just end here by saying that Avishay is an extraordinary advisor in *all* aspects. I'm sure that Avishay will forgive me for this abrupt ending, just like he forgave me for the awkward conversations at STOC'19.

I also want to thank Jelani Nelson, John Wright, Shirshendu Ganguly, and Ruixiang Zhang for joining my qualifying exam and dissertation committees. I took a lot of career advice from them and I wish I had more chances in doing research with them. I am also grateful to Google Research (Badih Ghazi and Ravi Kumar), NTT Research (Justin Holmgren), and Google Quantum AI (Robin Kothari) for the research intern opportunities on differential privacy, cryptography, and quantum computing. These experiences gave me legit excuses to say that I care about practical impact while doing theoretical research.

During my undergrad and PhD, I was extremely lucky to have many amazing collaborators. While it was not always the case that we produced publications (if you are one of these collaborators, I take full responsibility and I owe you coffee for this), I did always learn new fields and problems and tools. I don't want to offend anyone if I don't expand enough on our collaboration. So I will just list names in alphabetic order: Ishaq Aden-Ali, Ryan Alweiss, Fangqi Dong, Badih Ghazi, Uma Girish, David Gosset, Daniel Grier, Siyao Guo, Venkatesan Guruswami, Kun He, Justin Holmgren, Mingjia Huo, Barnabás Janzer, Jiaqing Jiang, Ce Jin, Zhihan Jin, Pritish Kamath, Daniel Kane, Tarun Kathuria, Robin Kothari, Ravi Kumar, Srijita Kundu, Jiaqi Leng, Qian Li, Chao Liao, Bingkai Lin, Qipeng Liu, Shachar Lovett, Pinyan Lu, Xin Lyu, Pasin Manurangsi, Jackson Morris, Jelani Nelson, Ryan O'Donnell, Anthony Ostuni, Xuandi Ren, Bhaskar Roberts, Makrand Sinha, Zhao Song, Benjamin Sudakov, Xiaoming Sun, Yican Sun, Yuan Sun, Avishay Tal, Shang-Hua Teng, Chunyang

Wang, Jiaheng Wang, Bujiao Wu, Xiaodi Wu, Zhiyu Xia, Xinyuan Xie, Guangxu Yang, Kuan Yang, Qi Ye, Yitong Yin, Chihao Zhang, Jialin Zhang, Jiapeng Zhang, Yufan Zheng.

Many friends helped me in ways beyond collaboration: Noga Alon, Ryan Babbush, Christian Borgs, Lijie Chen, Xiaonan Chen, Yijia Chen, Weiming Feng, Ofer Grossman, Tiancheng He, Zhongtian He, Max Hopkins, Hao Huang, Nathan Ju, Tanuj Khattar, Seri Khoury, Yuqing Kong, Jiatu Li, Tongyang Li, Zhihan Li, Jyun-Jie Liao, Mingmou Liu, Sihan Liu, Siqi Liu, Tianyi Liu, Yunchao Liu, Fermi Ma, Jie Ma, Sidhanth Mohanty, Dana Moshkovitz, Chinmay Nirkhe, Anurudh Peduri, Anup Rao, Luyao Ren, Joseph Slote, Longke Tang, Alexander Volberg, Yixuan Wang, Ryan Williams, Hongxun Wu, Ke Wu, Penghui Yao, Guanghao Ye, Huacheng Yu, Fred Zhang, Ye Zhang, Zhijun Zhang, Ziyi Zhang.

Apologize in advance if you don't find your name above. I should probably list more people with whom I had a fun time in tennis, pool, badminton, table tennis, cooking, hiking, ..., but this would become endless. So let me just focus on fishing (of course).

Yes, fishing is like my second career now. I have fishing photos all over my webpage; I talk about fishing all the time with everyone; I even talk about fishing on my job talk. All these started in 2021 when I came to the US after COVID: in a social gathering after Nathan Ju mentioned that he went fishing last weekend, I was completely blown away by the possibility of fishing in the wild here. I typed so many stories until I realized that here was not a fishing column. So let me constrain myself by sending out thanks, from which I am sure you can infer my spectacular fisherman's life. I thank Nathan Ju for introducing me to fish and taking me on shore fishing. I thank James Kao for teaching me lure fishing basics. I thank Jiashu Chen, Yuerou Tang, and Taize Yu for going after California (and all western) native trouts with me; I couldn't have completed those long-distance driving and planning without you. I thank Taize Yu for fishing with me in Slovenia for Marble Trout when we drove from Switzerland across Italy to Slovenia and back; and thank Jure Cotič for being our fishing guide in Slovenia. I thank Nick Buckmaster, Paul Divine, and Lee Duckwall for advice on Eagle Lake Rainbow Trout, Little Kern Golden Trout, Kern River Rainbow Trout, Paiute Cutthroat Trout, and Warner Lakes Redband Trout in California. I thank Stephan Charette, Matthew Falk, and Jerry George for advice on Bull Trout and Westslope Cutthroat Trout in Oregon. I thank Bryan Ferguson, Daniel Trujillo, and Jane Trujillo for advice on Gila Trout and Rio Grande Cutthroat Trout in New Mexico. I thank Tyler Coleman and Josh Nehring for advice on Greenback Cutthroat Trout in Colorado. I thank Sam Simmons for advice on Gila Trout and Apache Trout in Arizona. I thank Brian Stephens for advice on Westslope Cutthroat Trout in Montana. I thank Corey Brown for advice on targetting Alaksa species. I also get help from various Facebook fishing channels.

Last but most importantly, I want to thank my family. It feels weird to me that I want to expand on this but surprisingly I don't know where to start. My parents and grandparents have been supporting me in pursuit of whatever I want since, I don't know, maybe I could tell good and bad. My wife and I met in our undergrad and our relationship has been stable since its establishment in 2018. It is almost like granted nowadays for me to say "I am supported by my family". Nevertheless, let me say one more time that I am genuinely grateful for all your support and understanding!

Chapter 1

Introduction

Quantum computing has revolutionized computation by solving some classically intractable problems. This raises a fascinating and fundamental question:

What is the relative power of quantum versus classical computation?

Most known quantum advantages are conditioned on heuristics or assumptions. A famous example here is the story of *random circuit sampling*. In this task, we aim to draw samples from a distribution induced by shallow random quantum circuits. Formulated directly in terms of quantum circuits, the problem is easy for quantum devices and considered hard for classical computers. In 2019, Google [AAB+19] proposed it as a viable demonstration of quantum supremacy on near-term quantum devices, and they estimated it to take 10,000 years if run on a classical supercomputer. However shortly after their notice, IBM [Cho19] discussed a better classical simulation algorithm that only takes a few days on a classical supercomputer. In 2022, Pan, Chen, and Zhang [PCZ22] were even able to execute the experiment on a conventional computer within a few hours. The key ingredient behind these improved classical algorithms is the unavoidable physical noise in the quantum circuits. The culmination of this line of classical attacks is the work by Aharonov, Gao, Landau, Liu, and Vazirani [AGL+23] in 2023, which developed a classical algorithm that is provably polynomial time as long as the noise rate is constant. Though very recently Google [MVM+24] reclaimed quantum advantages for random circuit sampling in the scenario of diminishing noise, this story shows how fragile our heuristics can be.

Another example comes from post-quantum cryptography. There, *Learning With Error (LWE)* is a widely used primitive that is considered hard against quantum computers. However the work by Chen, Liu, and Zhandry [CLZ22] showed that some natural variants of LWE are actually vulnerable under some cleverly designed quantum attack. Hence we must be careful in formulating these assumptions: sometimes even a slight tweak would make the statement just false.

Though it is generally believed that quantum computers outperform classical ones, such debates are still disturbing and distracting. This motivates the questions for

proving rigorous quantum-classical separations, not relying on heuristics or assumptions.

Furthermore, understanding the reasons behind these quantum-classical separations brings out another question for

characterizing properties of computation tasks that allow for quantum speedups.

This thesis aims to address the above questions through a unified framework of Fourier growth analysis of Boolean functions.

1.1 Analysis of Boolean Functions

To explain what Fourier growth is, we start with a brief introduction to the analysis of Boolean functions.

Boolean functions map binary inputs to bounded outputs and capture most computational tasks and mathematical objects. The analysis of these functions involves studying their structural and analytic properties to understand their performance, efficiency, and robustness, which has been an ongoing success and delivered numerous breakthroughs in complexity theory, social choice theory, learning theory, cryptography, quantum computing, combinatorics, probability theory, and more. See [O'D14] for a comprehensive introduction to the analysis of Boolean functions.

A common theme in the analysis of Boolean functions is proving structural results on special classes of Boolean devices (e.g., decision trees, bounded-depth circuits) and then exploiting the structure to (1) devise pseudorandom generators fooling these devices [NN93], (2) prove lower bounds, showing that some explicit function cannot be computed by such Boolean devices of certain size [RT22], or (3) design learning algorithms for the class of Boolean devices in either the membership query model or the random sample model [KM93]. Such structural results can involve properties of the Fourier spectrum of Boolean functions associated with Boolean devices, like concentration on low-degree terms or concentration on a few terms.

To give a concrete example pertinent to this thesis, we formally define Boolean function $f\colon \{\pm 1\}^n \to [-1,1]$ and recall that it 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 each $\widehat{f}(S) \in \mathbb{R}$ is the Fourier coefficient of f, computed by

$$\widehat{f}(S) = \mathbb{E}\left[f(\boldsymbol{x}) \cdot \prod_{i\in S} \boldsymbol{x}_i\right] \quad \text{with } \boldsymbol{x} \text{ uniform over } \{\pm 1\}^n.$$

The seminal paper of Kushilevitz and Mansour [KM93] presents a learning algorithm in the membership query model, running in time $\mathsf{poly}(t, n)$ that can learn the unknown f assuming

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

This learning algorithm works by approximating f with a sparse polynomial and then learn each non-zero coefficient using the membership query oracle.

While having small $\left\|\widehat{f}\right\|_1$ implies learning algorithms and also simple pseudorandom generators fooling f [NN93], this property can be quite restrictive. In particular, very simple functions (e.g., the tribes function) have $\left\|\widehat{f}\right\|_1$ exponential in n. Such examples motivate a more refined metric for the Fourier spectrum which we now discuss.

1.2 Fourier Growth

Dating back to Mansour [Man95], the notion of *Fourier growth* captures the scaling of the Fourier spectrum with respect to different Fourier levels. 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.

Formally, the level-ℓ Fourier growth of f is the sum of absolute values of its level-ℓ Fourier coefficients

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

The idea behind this more refined notion is that Fourier coefficients of different levels behave differently under standard manipulations to the function like random restrictions or noise operators. For example, under a noise operator with parameter γ, level-ℓ coefficients are multiplied by γ^ℓ.

Indeed, upper bounds on the Fourier growth, even for the first few Fourier levels, have interesting applications.

- A bound on the level-1 Fourier growth is sufficient to control the advantage of distinguishing biased coins from unbiased ones [Agr20].
- A bound on the level-2 Fourier growth already gives non-trivial pseudorandom generators [CHLT19], oracle separations between BQP and PH [RT22, Wu22], and separations between efficient quantum communication and randomized classical communication [GRT22].
- A bound on higher levels leads to better constructions of pseudorandom generators [CHHL19, CHRT18, CGL+21], improved quantum-classical separations [BS21], pseudorandomness regarding expander random walks [CPTS21], efficient learing algorithms [Man95], and more.

Motivated by above, Fourier growth bounds have been extensively studied and established for different computation 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, SSW23], functions related to small-cost communication protocols [GRZ21, GRT22], low-degree $\mathbb{F}_2$ polynomials [CHHL19, CHLT19, BIJ+21], product tests [Lee19], small-depth parity decision trees [BTW15], low-degree bounded functions [IRR+21], and more.

We remark that, by Parseval's identity and Cauchy-Schwarz inequality, we always have $L_\ell(f) \le \sqrt{\binom{n}{\ell}}$. 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 $\mathbb{F}_2$ polynomials (which well-approximate $\mathsf{AC}^0[\oplus]$ when we set $d = \mathsf{polylog}(n)$ [Raz87, Smo87]), while we know a level-one bound of $L_1(f) \le O(d)$ due to [CHLT19], the current best bound for levels $\ell \ge 2$ is roughly $2^{O(d\ell)}$ [CHHL19], whereas the conjectured bound is $d^{O(\ell)}$. Validating such a bound, even for the second level $\ell = 2$, will imply unconditional pseudorandom generators of polylogarithmic seed length for $\mathsf{AC}^0[\oplus]$ [CHLT19], a longstanding open problem in circuit complexity and pseudorandomness.

1.3 Quantum Advantages and Forrelation Problem

As hinted in Section 1.2, Fourier growth is closely related to quantum advantages. Intuitively, quantum computations can aggregate signals on the Fourier basis, whereas classical algorithms cannot. The notion of Fourier growth helps to justify this intuition and *prove* quantum advantages in the query model [RT22, Tal20, BS21] and beyond.

The quantum query model, also known as the *black-box* or *oracle* model, has been a successful test bed to develop quantum algorithms and to give provable guarantees on speedups over classical algorithms. In this model, a quantum algorithm has "black-box access" to the input and is only charged for quantum queries to the input, while any intermediate computation is considered free. In other words, this model is an abstraction of the cloud computing setting where the communication with the remote server is the dominating cost. Most well-known quantum algorithms, such as Grover's search [Gro96], Deutsch-Josza's algorithm [DJ92], Bernstein-Vazirani's algorithm [BV97], Simon's Algorithm [Sim97], and Shor's period-finding algorithm [Sho99], are all captured by this black-box access model.

The focus in the query model has been to compare quantum algorithms with classical ones. The culmination of this line of work led to the resolution of the following speedup question:

What is the largest possible quantum speedup over classical algorithms?

The motivation for this question stems from an attempt to pinpoint the exact limit of quantum speedups, and it has helped us develop a better understanding of the fundamental nature of quantum speedups.

Towards this question, Aaronson [Aar10] introduced the Forrelation problem, which measures the correlation between a Boolean function and the Fourier spectrum of another. To give some detail, on input $x = (x_1, x_2) \in \{\pm 1\}^{2n}$ where n is a power of 2, we define

$$\mathrm{forr}(x) = \frac{1}{n} \cdot x_1^\top H x_2,$$

where H denotes the $n \times n$ Hadamard matrix. The Forrelation problem asks to distinguish the case when $|\mathrm{forr}(x)|$ is large from the case where $\mathrm{forr}(x)$ is close to zero. For this problem, [Aar10] showed that we only need to make 1 quantum query, but would require $\Omega(\sqrt{n})$ classical queries.

This problem was later generalized by Aaronson and Ambainis [AA18] as the k-fold Forrelation problem (see Definition 2.0.10 for a precise definition). They showed that the k-fold Forrelation problem can be solved with $r := \lceil k/2 \rceil$ quantum queries, or a classical algorithm with $O\left(n^{1-1/2r}\right)$ queries. Not only they conjectured that this classical simulation is optimal, they also conjectured that one should be able to simulate *any* r-query quantum algorithm with $O(n^{1-1/2r})$ classical queries, making the k-fold Forrelation problem a witness for *maximal* quantum advantage. The latter conjecture was resolved by Bravyi, Gosset, Grier, and Schaeffer [BGGS22]. Up to low-order terms, the first conjecture was also proved by Sherstov, Storozhenko, and Wu [SSW23] and Bansal and Sinha [BS21], building on the work of Raz and Tal [RT22] and Tal [Tal20].

The framework developed in [RT22, BS21, Tal20, SSW23] is rather general. Informally, they show that if the Fourier growth of f scales slower than $(\sqrt{n})^{(1-1/k)\ell}$, which is the level-ℓ Fourier growth of the k-fold Forrelation function, then f cannot approximately compute the k-fold Forrelation problem and its variants. Given this and to separate computation model $\mathcal{C}_1$ and $\mathcal{C}_2$, it suffices to show (1) $\mathcal{C}_1$ can compute the k-fold Forrelation problem or its variants, which is usually easy due to [AA18], and (2) every Boolean function in $\mathcal{C}_2$ has small Fourier growth, which is usually the main effort. Indeed, the results in [RT22, Tal20, SSW23] are upper bounds on the Fourier growth of classical query algorithms that make limited number of queries (aka decision trees of small depth).

1.4 Our Contribution

The main contribution of this thesis is to extend the Fourier growth bounds in the standard classical query model to other meaningful computation models. Along the way and as illustrated in Section 1.2, we obtain unconditional exponential quantum advantages and various applications in pseudorandomness and communication complexity:

- Chapter 3 is based on the joint work [GTW21] with Uma Girish and Avishay Tal. We extend the query model results in [RT22, BS21, Tal20, SSW23] by allowing classical

algorithms to perform parity queries, a natural generalization of the standard bit query. We establish near optimal Fourier growth bounds for parity query classical algorithms, implying that quantum algorithms have significant advantages even when classical query algorithms have the power to perform parity queries.

- Chapter 4 is based on the joint work [GSTW23] with Uma Girish, Makrand Sinha, and Avishay Tal. We go beyond the query model and ask for quantum advantages in the more challenging communication model. We extend the Fourier growth arguments in [GTW21] and give exponential separation between efficient quantum simultaneous communication and randomized two-way classical communication, providing quantitative improvements over the prior work [GRT22].
- Chapter 5 is based on the joint work [GSTW24] with Uma Girish, Makrand Sinha, and Avishay Tal. To get a better understanding of quantum query algorithms, we study them in term of the number of adaptive query rounds, which is an abstraction of shallow quantum circuits from near-term quantum devices. We use Fourier growth to separate deep versus shallow quantum algorithms, demonstrating the necessity of adaptivity in the query model.

Finally in Chapter 6, I briefly discuss other projects that I was involved in and completed during my PhD study at UC Berkeley.

Chapter 2

Preliminaries

Here we define notation and quote standard results that are shared by the later chapters. Special definitions that are not commonly used will be provided in the corresponding chapter afterwards.

Standard Notation

We use $\mathbb{N} = \{0, 1, 2, \ldots\}$ to denote the set of natural numbers; and use $\mathbb{R}, \mathbb{C}$ to denote real numbers and complex numbers respectively. We use $\log(\cdot)$ and $\ln(\cdot)$ to denote logarithms with base 2 and e respectively.

For a complex number $x \neq 0$, define $\mathsf{Phase}(x) = x/|x|$ as its phase; and we additionally define $\mathsf{Phase}(0) = 1$. For a real number x, define $\mathsf{sgn}(x) \in \{-1, 0, 1\}$ to be the sign of x, i.e., $\mathsf{sgn}(x)$ equals -1 if $x < 0$; 1 if $x > 0$; and 0 if $x = 0$.

For positive integer n, we use $[n]$ to denote $\{1, 2, \ldots, n\}$; and $\binom{[n]}{k}$ (resp., $\binom{[n]}{\leq k}$) to denote the set of all size-k (resp., size-$(\leq k)$) sets from $[n]$.

For sets S_1, S_2, we use $S_1 \oplus S_2$ to denote their symmetric difference, i.e., $S_1 \oplus S_2 = (S_1 \setminus S_2) \cup (S_2 \setminus S_1)$. If $S_1, S_2 \subseteq \mathbb{R}^n$, then we define $S_1 + S_2 = \{x + y \colon x \in S_1, y \in S_2\}$.

For a finite set S, we use 2^S to denote the set of all its subsets. If S is a set from universe U clear from the context, then we write $\overline{S}$ for its complement $U \setminus S$. If $S \subseteq \mathbb{R}^n$, then for any $t \in \mathbb{R}$ we define $tS = \{t \cdot x \colon x \in S\}$.

Asymptotics

We use $O(\cdot), \Omega(\cdot), \Theta(\cdot)$ to hide universal multiplicative constants that do not depend on any parameter. $\widetilde{O}(\cdot)$ and $\widetilde{\Omega}(\cdot)$ hide polylogarithmic factors, i.e., $\widetilde{O}(f) = O(f \cdot \mathsf{polylog} f)$ and $\widetilde{\Omega}(f) = \Omega(f \cdot \mathsf{polylog} f)$. We also use subscript to hide dependence on minor parameters, e.g., $O_{r,d}(f) = O(f \cdot K(r, d))$ for an implicit function K.

For convenience, we will use $\lesssim, \gtrsim, \ll, \gg$ to hide minor factors in bounds; these will only be used in informal contexts and not in actual proofs.

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 $\boldsymbol{x} \sim \xi$ to denote a random sample distributed according to ξ and $\mathbb{E}_{\boldsymbol{x}\sim\xi}[f(\boldsymbol{x})]$ to denote the expectation of a function f under the measure ξ. For a finite set $\mathcal{X}$ we use $\boldsymbol{x} \sim \mathcal{X}$ to denote that x is a random variable sampled uniformly from $\mathcal{X}$.

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 σ-algebra $\mathcal{F}$ and the σ-algebra generated by $\mathcal{E}$.

Linear Algebra

We use $\mathbb{F}_2 = \{0, 1\}$ to denote the binary field, $\mathsf{Span}\,\langle\mathsf{vectors}\rangle$ to denote the subspace spanned by $\mathsf{vectors}$ over $\mathbb{F}_2$. For a (complex) vector, we use $\|\cdot\|$ to denote its ℓ_2-norm; for a (complex) matrix, we use $\|\cdot\|$ and $\|\cdot\|_{\mathrm{F}}$ to denote its operator and Frobenius norm respectively. We say a vector u is a *unit vector* if $\|u\| = 1$. We use I_m to denote the m by m identity matrix, and, when m is clear from the context, we will simply use I. We use $\mathbb{C}^{[n]\times[m]}$ and $\mathbb{C}^{n\times m}$ to denote the space of complex n by m matrices. For nonzero vector x or matrix X, we define $\mathrm{unit}(x)$ or $\mathrm{unit}(X)$ as the unit vector or matrix along direction x and X respectively: $\mathrm{unit}(x) = x/\|x\|$ and $\mathrm{unit}(X) = X/\|X\|_{\mathrm{F}}$.

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 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 Frobenius norm) in $\mathbb{R}^{n\times n}$ where additionally the diagonal entries of the $n \times n$ matrices are zero. We use $\mathcal{B}^n = \{x \in \mathbb{R}^n \mid \|x\| \le 1\}$ to denote the unit Euclidean ball in $\mathbb{R}^n$. 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.

Fact 2.0.1. *Let $M \in \mathbb{C}^{[n]\times[m]}$ be an n by m matrix. Then for any $S \subseteq [n]$ and $T \subseteq [m]$, we have $\|M[S,T]\| \le \|M\|$, where $M[S,T]$ is the sub-matrix of M formed by rows in S and columns in T.*

Fact 2.0.2. *Assume $M = \mathsf{diag}(M_1, \ldots, M_t)$ is a block diagonal matrix. Then $\|M\| = \max_{i\in[t]} \|M_i\|$.*

Fact 2.0.3 (Hölder's Inequality)**.** *$\|M\| \le \sqrt{\|M\|_1 \|M\|_\infty}$ holds for any matrix $M \in \mathbb{C}^{[n]\times[m]}$, where*

$$\|M\|_1 = \max_{1\le j\le m} \sum_{i=1}^{n} |M[i,j]| \quad \textit{and} \quad \|M\|_\infty = \max_{1\le i\le n} \sum_{j=1}^{m} |M[i,j]|.$$

Gaussian

We use γ_n to denote the n-dimensional standard Gaussian measure in $\mathbb{R}^n$. We say a random variable $\boldsymbol{x} \in \mathbb{R}^n$ 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 $\gamma_{\mathrm{rel}}(S)$ is then defined to be the Gaussian measure of the set $S - t$ under γ_V.

Let $\Phi\colon \mathbb{R} \to [0,1]$ be the cumulative distribution function of the standard Gaussian distribution, i.e., $\Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-u^2/2} \,\mathrm{d}\, u$.

Theorem 2.0.4 (Gaussian Isoperimetric Inequality [Bor75, ST78])**.** *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 $a = 0$ and have

$$\gamma_n(A + t\mathcal{B}^n) \geq \Phi(t) \geq 1 - e^{-t^2/2}. \tag{2.1}$$

Fact 2.0.5 (See e.g., [Ver18])**.** *For any $m \in \mathbb{N}$ and $r \geq 2m$, we have*

$$\Pr_{\boldsymbol{x} \sim \gamma_m}\left[\sum_{i=1}^{m} \boldsymbol{x}_i^2 \geq r\right] \leq e^{-r/4}.$$

Theorem 2.0.6 (Level-k Inequality, see e.g., [EM22, Lemma 2.2])**.** *Let $k \in \{1, 2\}$. Assume $A \subseteq \mathbb{R}^n$ is measurable. Let $1_A\colon \mathbb{R}^n \to \{0,1\}$ be the indicator function of A and define $\mu := \mathbb{E}_{\boldsymbol{x} \sim \gamma}[1_A(\boldsymbol{x})]$. Then, we have*[1]

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

In particular, if $\mu > 0$, we have

$$\sum_{|S|=k} \left(\mathop{\mathbb{E}}_{\boldsymbol{x} \sim \gamma_n}[\boldsymbol{x}_S \,|\, \boldsymbol{x} \in A]\right)^2 \leq 2e^2 \cdot \ln^k(e/\mu).$$

[1] Our Theorem 2.0.6 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.

Martingale

Given a sequence of real-valued random variables $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ in a probability space $(\Omega, \mathcal{F}, \xi)$ and a function $f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ satisfying $\mathbb{E}\left[|f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)|\right] < \infty$, the sequence of random variables $\boldsymbol{z}^{(t)} = \mathbb{E}\left[f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \,\middle|\, \mathcal{F}^{(t-1)}\right]$ is called the *Doob martingale* where $\mathcal{F}^{(t-1)}$ is the σ-algebra generated by $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1}$ 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 $(\boldsymbol{z}^{(t)})_t$ is called *predictable* (or *adapted*) with respect to $\mathcal{F}^{(t)}$ if $\boldsymbol{z}^{(t)}$ is $\mathcal{F}^{(t)}$-measurable for every t, meaning that it is determined by the randomness in $\mathcal{F}^{(t)}$.

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

Let $\Delta\boldsymbol{z}^{(t)} := \boldsymbol{z}^{(t)} - \boldsymbol{z}^{(t-1)}$ be the martingale differences. Note that $\mathbb{E}\left[\Delta\boldsymbol{z}^{(t)} \,\middle|\, \mathcal{F}^{(t-1)}\right] = 0$ and thus

$$\mathbb{E}\left[\left(\boldsymbol{z}^{(t)}\right)^2\right] = \mathbb{E}\left[\left(\sum_{t=1}^{n} \Delta\boldsymbol{z}^{(t)}\right)^2\right] = \mathbb{E}\left[\sum_{t=1}^{n} \left(\Delta\boldsymbol{z}^{(t)}\right)^2\right], \tag{2.2}$$

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 $\left(\boldsymbol{z}^{(t)}\right)_t$. If the sequence $(\boldsymbol{z}^{(t)})_t$ is vector-valued (resp., matrix-valued) and satisfies $\mathbb{E}\left[\Delta\boldsymbol{z}^{(t)} \,\middle|\, \mathcal{F}^{(t-1)}\right] = 0$ 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:

$$\mathbb{E}\left[\left\|\boldsymbol{z}^{(t)}\right\|_{\mathrm{F}}^2\right] = \mathbb{E}\left[\left\|\sum_{t=1}^{n} \Delta\boldsymbol{z}^{(t)}\right\|_{\mathrm{F}}^2\right] = \mathbb{E}\left[\sum_{t=1}^{n} \left\|\Delta\boldsymbol{z}^{(t)}\right\|_{\mathrm{F}}^2\right]. \tag{2.3}$$

Boolean Function

Here we recall definitions in the analysis of Boolean functions (see [O'D14] for a detailed introduction). Let $\mathcal{U}_n$ be the uniform probability measure over $\{\pm 1\}^n$.

Let $f\colon \{\pm 1\}^n \to \mathbb{R}$ be any Boolean function. The restriction of f refers to fixing some of its input bits to ± 1. For any $p > 0$, the p-norm of f is defined as

$$\|f\|_p = \left(\mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mathcal{U}_n} \left[|f(\boldsymbol{x})|^p\right]\right)^{1/p}.$$

For any subset $S \subseteq [n]$, x_S denotes $\prod_{i \in S} x_i$ (in particular, $x_\emptyset = 1$). It is a well-known fact that we can uniquely represent f as a linear combination of $\{x_S\}_{S \subseteq [n]}$:

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S) x_S,$$

where the coefficients $\left\{\widehat{f}(S)\right\}_{S \subseteq [n]}$ are referred to as the *Fourier coefficients* of f and are given by $\widehat{f}(S) = \mathbb{E}_{\boldsymbol{x} \sim \mathcal{U}_n}[f(\boldsymbol{x})\boldsymbol{x}_S]$. The above representation expresses f as a multilinear polynomial and is called the Fourier representation of f. We say that f is of degree at most d if its Fourier representation is a polynomial of degree at most d, i.e., if $\widehat{f}(S) = 0$ for all $S \subseteq [n], |S| > d$.

The level-ℓ Fourier growth of f is denoted by $L_\ell(f)$ and defined as the sum of absolute values of its level-ℓ Fourier coefficients

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

Theorem 2.0.7 ([Bon70], see also [O'D14, $(2,q)$-hypercontractivity])**.** *Let $f \colon \{\pm 1\}^n \to \mathbb{R}$ be a degree-d polynomial. Then for any $q \geq 2$, we have $\|f\|_q \leq (q-1)^{d/2} \|f\|_2$.*

We use the standard notion of k-wise independence and some relevant concentration bounds.

Definition 2.0.8 (k-Wise Independence)**.** *A distribution $\mathcal{D}$ over $\{\pm 1\}^n$ is k-wise independent if for $\boldsymbol{x} \sim \mathcal{D}$ and any k-indices $1 \leq i_1 < i_2 < \cdots < i_k \leq n$, the vector $(\boldsymbol{x}_{i_1}, \ldots, \boldsymbol{x}_{i_k})$ has distribution $\mathcal{U}_k$.*

Lemma 2.0.9. *Let $f \colon \{\pm 1\}^n \to \mathbb{R}$ be a degree-d polynomial. Let $\mathcal{D}$ be a $2k$-wise independent distribution over $\{\pm 1\}^n$, where $k \geq d$. Let $\mu = \mathbb{E}_{\boldsymbol{x} \sim \mathcal{D}}[f(\boldsymbol{x})]$ and $\sigma^2 = \mathbb{E}_{\boldsymbol{x} \sim \mathcal{D}}[(f(\boldsymbol{x}) - \mu)^2]$. Then for any $\alpha > 0$ and any integer $1 \leq \ell \leq k/d$, we have*

$$\mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mathcal{D}}\left[(f(\boldsymbol{x}) - \mu)^{2\ell}\right] \leq \sigma^{2\ell} \cdot (2\ell - 1)^{d \cdot \ell}.$$

In particular we have

$$\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \mathcal{D}}[|f(\boldsymbol{x}) - \mu| \geq \alpha \cdot \sigma] \leq \alpha^2 \cdot \left(\frac{2k}{d \cdot \alpha^{2/d}}\right)^k.$$

Proof. Since $(f(x) - \mu)^{2\ell}$ is a polynomial of degree at most $2\ell \cdot d \leq 2k$, its expectation under $\mathcal{D}$ is the same as its expectation under $\mathcal{U}_n$. By Theorem 2.0.7, we have

$$\|f - \mu\|_{2\ell} \leq (2\ell - 1)^{d/2} \|f - \mu\|_2 = \sigma \cdot (2\ell - 1)^{d/2}.$$

Hence we obtain the first bound by Markov's inequality. For the second bound, the RHS is trivial if $\alpha \leq 1$ and follows from the first bound by setting $\ell = \lfloor k/d \rfloor$ if $\alpha \geq 1$. □

(Parallel) Quantum Query Algorithm

We use the following standard model for quantum query algorithms with parallel queries [Mon10]. Let O_x be the standard quantum query oracle of input $x = x_1 \cdots x_n \in \{\pm 1\}^n$. That is, O_x acts on an $(n+1)$-dimensional space indexed by basis states $|0\rangle, |1\rangle, \ldots, |n\rangle$, and performs the operation $O_x |0\rangle = |0\rangle$ and $O_x |i\rangle = x_i |i\rangle$ for each $i \in [n]$.

Let $\mathcal{A}$ be a quantum query algorithm that makes r rounds of adaptive queries with t parallel queries per round. Assume it uses w auxiliary qubits, and computes a Boolean function $f\colon \{\pm 1\}^n \to \{\pm 1\}$ with probability at least $1 - \varepsilon$, then it is equivalent to the existence of

- a unit state $|\psi\rangle \in \mathbb{C}^{\{0,\ldots,n\}^t \times [2^w]}$,
- $r-1$ unitary matrices $U_1, \ldots, U_{r-1}$,
- a measurement matrix M that $\|M\| \le 1$ and M is positive semi-definite,

such that

$$\left\| \sqrt{M}(O_x^{\otimes t} \otimes I_{2^w}) U_{r-1} \cdots U_2 (O_x^{\otimes t} \otimes I_{2^w}) U_1 (O_x^{\otimes t} \otimes I_{2^w}) |\psi\rangle \right\|^2 \begin{cases} \ge 1-\varepsilon & \text{for all } x \in f^{-1}(1), \\ \le \varepsilon & \text{for all } x \in f^{-1}(0). \end{cases}$$

Note that standard quantum query algorithm will have $t = 1$ and r being its quantum query complexity.

We remark that another natural way of describing the quantum query is through an oracle O'_x, which acts on a $2(n+1)$-dimensional space indexed by basis states $\{|i\rangle |b\rangle\}_{i \in \{0,\ldots,n\}, b \in \{\pm 1\}}$, and performs the operation $O'_x |0\rangle |b\rangle = |0\rangle |b\rangle$ and $O'_x |i\rangle |b\rangle = |i\rangle |b \cdot x_i\rangle$ for each $i \in [n], b \in \{\pm 1\}$. It turns out that the two models are equivalent [HS05, Mon10] in the sense that

$$O'_x = V_1 (O_x \otimes I_2) V_2$$

for some unitary matrices V_1, V_2. We will use the standard model with the O_x oracle, which is more convenient for our purposes.

Folded Forrelation

We give a formal definition of the folded Forrelation problem, introduced by Aaronson and Ambainis [AA18].

Definition 2.0.10 (k-fold Forrelation Problem)**.** *For an integer $k \ge 2$, the k-fold Forrelation problem is a partial Boolean function on n bits. Let $H = H_n$ denote the $n \times n$ (orthonormal) Hadamard matrix where $n = 2^m$ for $m \in \mathbb{N}$. Let $x_1, \ldots, x_k \in \{\pm 1\}^n$ denote truth tables of k different Boolean functions. Define the degree-k polynomial* $\mathrm{forr}_k : \{\pm 1\}^{kn} \to \mathbb{R}$ *as follows*

$$\mathrm{forr}_k(x) = \frac{1}{n} \sum_{(i_1,\ldots,i_k) \in [n]^k} x_1(i_1) \cdot H_{i_1,i_2} \cdot x_2(i_2) \cdot H_{i_2,i_3} \cdot \cdots \cdot x_{k-1}(i_{k-1}) \cdot H_{i_{k-1},i_k} \cdot x_k(i_k).$$

The k-fold Forrelation problem is to decide whether $|\mathrm{forr}_k(x)| \leq \frac{\delta}{2}$ or $\mathrm{forr}_k(x) \geq \delta$ for a parameter δ. For the applications in this paper, we take $\delta = 2^{-5k}$.

As shown in [AA18], the folded Forrelation problem has an efficient quantum query algorithm.

Fact 2.0.11 ([AA18])**.** *There exists a quantum circuit Q that makes $\lceil k/2 \rceil$ queries and uses $O(k \log n)$ gates, such that for any input $x \in \{\pm 1\}^{kn}$, it holds that*

$$\mathbf{Pr}\left[Q \text{ accepts } x\right] = \frac{1 + \mathrm{forr}_k(x)}{2}.$$

The following theorem relates the hardness of computing folded Forrelation problem with Fourier growth.

Theorem 2.0.12 ([BS21, Theorem 3.2])**.** *Let $f\colon \{\pm 1\}^n \to [0,1]$ such that f and all its restrictions satisfy $L_\ell(f) \leq t^\ell$ for $\ell = \{k, \ldots, k(k-1)\}$. Let $\delta = 2^{-5k}$. Suppose f is δ-close to the value of k-fold Forrelation of x for all x on which k-fold Forrelation is defined. Then*

$$t \geq \Omega\left(\frac{n^{(1-1/k)/2}}{k^{15}}\right).$$

Chapter 3

Quantum Advantages over Parity Query Algorithms

We prove that for every parity decision tree of depth d on n variables, the sum of absolute values of Fourier coefficients at level ℓ is at most

$$d^{\ell/2} \cdot O(\ell \cdot \log(n))^{\ell}.$$

Our result is nearly tight for small values of ℓ and extends a previous Fourier bound for standard decision trees by Sherstov, Storozhenko, and Wu (STOC, 2021).

As an application of our Fourier bounds, using the results of Bansal and Sinha (STOC, 2021), we show that the k-fold Forrelation problem has (randomized) parity decision tree complexity $\widetilde{\Omega}\left(n^{1-1/k}\right)$, while having quantum query complexity $\lceil k/2 \rceil$.

Our proof follows a random-walk approach, analyzing the contribution of a random path in the decision tree to the level-ℓ Fourier expression. To carry the argument, we apply a careful cleanup procedure to the parity decision tree, ensuring that the value of the random walk is bounded with high probability. We observe that step sizes for the level-ℓ walks can be computed by the intermediate values of level $\leq \ell - 1$ walks, which calls for an inductive argument. Our approach differs from previous proofs of Tal (FOCS, 2020) and Sherstov, Storozhenko, and Wu (STOC, 2021) that relied on decompositions of the tree. In particular, for the special case of standard decision trees we view our proof as slightly simpler and more intuitive.

In addition, we prove a similar bound for noisy decision trees of cost at most d – a model that was introduced by Ben-David and Blais (FOCS, 2020).

Organization. In Section 3.1, we give a brief introduction on the literature of parity decision tree. Then in Section 3.2, we discuss our main results and applications. In Section 3.3, we give an overview of our analysis. In Section 3.4, we state and prove an adaptive version of Azuma's inequality for later referencing. In Section 3.5, we present the cleanup process for parity decision trees. Finally we prove the Fourier growth bounds for parity decision trees in Section 3.6 and for noisy decision trees in Section 3.7.

3.1 Introduction

A parity decision tree (PDT) is an extension of the standard decision tree model. On input $x = (x_1, \ldots, x_n) \in \{\pm 1\}^n$, a PDT is a binary tree where each internal node is marked by a linear function (i.e., a product of coordinates), with two outgoing edges marked with ± 1, and each leaf is marked with either 0 or 1. A PDT naturally describes a computation model: on input x, start at the root and at each step query the linear function specified by the current node on the input x and continue on the edge marked with the value of the linear function evaluated on x. Finally, when reaching a leaf, output the value specified in the leaf. PDTs naturally generalize standard decision trees that can only query the value of a single input bit in each internal node. See Definition 3.1.1 for a formal definition.

PDTs were introduced in the seminal paper of Kushilevitz and Mansour [KM93], which proved a structural result for PDTs and used it to design learning algorithms for PDTs. They showed that every PDT of size s computing a Boolean function $f\colon \{\pm 1\}^n \to \{0,1\}$ has

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

Then, they gave a learning algorithm in the membership query model, running in time $\mathsf{poly}(t, n)$ that can learn any function f with $\left\|\widehat{f}\right\|_1 \leq t$. Combining the two results together, they obtained a $\mathsf{poly}(s, n)$-time algorithm for learning PDTs of size s.

Parity decision trees were also studied in relation to communication complexity and the log-rank conjecture [MO09, ZS09, ZS10, TWXZ13, STlV17, OWZ+14, CS16, KQS15, HHL18, San19, MS20]. Suppose Alice gets input $x \in \{\pm 1\}^n$, Bob gets input $y \in \{\pm 1\}^n$ and they want to compute some function $f(x, y)$. When f is an XOR function, namely $f(x, y) = g(x \odot y)$ for some $g : \{\pm 1\}^n \to \{\pm 1\}$. Then any PDT for g of depth d can be translated into a communication protocol for f at cost $2d$: Alice and Bob simply traverse the PDT together, both exchanging the parity of their part of the input to simulate each query in the PDT. With this view, parity decision trees can be thought of as special cases of communication protocols for XOR functions. A surprising result by Hatami, Hosseini, and Lovett [HHL18], shows that this is not far from the optimal strategy for XOR functions. Namely, if the communication cost for computing f is c, then the parity decision tree complexity of g is at most $\mathsf{poly}(c)$. Due to this connection, the log-rank conjecture for XOR functions reduces to the question of whether Boolean functions with at most s non-zero Fourier coefficients can be computed by PDTs of depth $\mathsf{polylog}(s)$ [MO09, ZS09]. The best known upper bound is that such functions can be computed by PDTs of depth $O(\sqrt{s})$ [TWXZ13] (or even non-adaptive PDTs of depth $\widetilde{O}(\sqrt{s})$ [San19]).

The most relevant result to our Fourier growth analysis is the tight Fourier growth bounds for decision trees of depth d. Sherstov, Storozhenko and Wu [SSW23] proved that for any randomized decision tree of depth d computing a function f, it holds that $L_\ell(f) \leq \sqrt{\binom{d}{\ell} \cdot O(\log(n))^{\ell-1}}$. Their bound is nearly tight (see [Tal20, Section 7] and [O'D14, Chapter 5.3] for tightness examples). As mentioned in Section 1.3, one motivation for show-

ing such a bound for decision trees is that it demonstrates a stark difference between quantum algorithms making few queries and randomized algorithms making many queries. Based on that difference, both [SSW23] and [BS21] showed that there are partial functions, relevant to k-fold Forrelation, that can be correctly computed with probability at least $1/2+\Omega(1)$ by quantum algorithms making $\lceil k/2 \rceil$ queries, but require $\widetilde{\Omega}\left(n^{1-1/k}\right)$ queries for any randomized algorithm. Moreover, due to the result of Aaronson and Ambainis [AA18] this is the largest possible separation between the two models.

For parity decision trees, the work of Blais, Tan, and Wan [BTW15] established a tight bound of $O\left(\sqrt{d}\right)$ on the first level $\ell = 1$. To the best of our knowledge, bounds on higher levels were not considered previously in the literature (in fact, even for standard decision trees, such bounds were not considered prior to [Tal20]).

Here we formally define parity decision trees.

Definition 3.1.1 (Parity Decision Tree)**.** *A* parity decision tree $\mathcal{T}$ *is a representation of a Boolean function* $f\colon \{\pm 1\}^n \to \{0,1\}$*. It consists of a rooted binary tree in which each internal node* v *is labeled by a non-empty set* $Q_v \subseteq [n]$*, the outgoing edges of each internal node are labeled by* $+1$ *and* -1*, and the leaves are labeled by* 0 *and* 1*.*

On input $x \in \{\pm 1\}^n$*, the tree* $\mathcal{T}$ *constructs a* computation path $\mathcal{P}$ *from the root to a leaf. Specifically, when* $\mathcal{P}$ *reaches an internal node* v *we say that* $\mathcal{T}$ queries Q_v*; then* $\mathcal{P}$ *follows the outgoing edge labeled by* $\prod_{i \in Q_v} x_i$*. We require that* Q_v *is not implied by its ancestors' queries. The output of* $\mathcal{T}$ *(and hence* f*) on input* x *is the label of the leaf reached by the computation path. Conversely, we say* x *is* consistent with *the path* $\mathcal{P}$ *if* $\mathcal{P}$ *is the computation path (possibly ending before reaching a leaf) for* x*.*

We make a few more remarks on a parity decision tree $\mathcal{T}\colon \{\pm 1\}^n \to \{0,1\}$.

- A node v in $\mathcal{T}$ can be either an internal node or a leaf, and we use $\mathcal{T}(v) \in \{0,1\}$ to denote the label on v when v is a leaf. Meanwhile, we use $\mathcal{T}_v$ to denote the sub parity decision tree starting with node v.
- The *depth* of a node is the number of its ancestors (e.g., the root has depth 0) and the depth of $\mathcal{T}$ is the maximum depth over all its leaves.
- We say that two parity decision trees $\mathcal{T}$ and $\mathcal{T}'$ are *equivalent* (denoted by $\mathcal{T} \equiv \mathcal{T}'$) if they compute the same function.

3.2 Our Results

We prove level-ℓ bounds for any parity decision tree of depth d.

Theorem 3.2.1 (Informal, see Theorem 3.6.5 and Theorem 3.6.12)**.** *Let* $\mathcal{T}$ *be a depth-*d *parity decision tree on* n *variables. Then*

$$L_\ell(\mathcal{T}) \leq d^{\ell/2} \cdot O(\ell \cdot \log(n))^\ell .$$

Theorem 3.2.1 extends the result of [SSW23] from standard decision trees to parity decision trees at the cost of an $(\ell \cdot \log(n))^{O(\ell)}$ multiplicative factor. We remark that even for standard decision tree there is a lower bound of $L_\ell(\mathcal{T}) \geq \sqrt{\binom{d}{\ell} \cdot (\log(n))^{\ell-1}}$ [Tal20, Section 7] for constant ℓ and $L_\ell(\mathcal{T}) \geq \frac{1}{\mathsf{poly}(\ell)} \cdot \sqrt{\binom{d}{\ell}}$ for all ℓ [O'D14, Chapter 5.3]. Thus, our bounds are tight up to $\mathsf{polylog}(n)$ factors for constant ℓ, and they deteriorate as ℓ grows. Nevertheless, our main application relies on the bounds for small values of ℓ from Theorem 2.0.12.

Noisy Decision Trees

We also investigate the Fourier spectrum of noisy decision trees. Noisy decision trees are a different generalization of the standard model; here in each internal node v we query a noisy version of an input bit, that equals the true bit with probability $(1+\gamma_v)/2$. Any such query costs γ_v^2. We say that a noisy decision tree has cost at most d if the total cost in any root-to-leaf path is at most d. See Definition 3.7.2 for a formal definition.

Recent work studied this model and established connections to the question of how randomized decision tree complexity behaves under composition [BB20]. We prove level-ℓ bounds for any noisy decision tree of cost at most d.

Theorem 3.2.2 (Informal, see Theorem 3.7.5)**.** *Let $\mathcal{T}$ be a noisy decision tree of cost at most d on n variables. Then*

$$L_\ell(\mathcal{T}) \leq O(d)^{\ell/2} \cdot (\ell \cdot \log(n))^{(\ell-1)/2}.$$

Extension to Randomized Query Models

It is simple to verify that if f is a convex combination of Boolean functions $f_1, \ldots, f_m$ each with $L_\ell(f_i) \leq t_\ell$ then also f satisfy $L_\ell(f) \leq t_\ell$. Thus, if we take a distribution over PDTs of depth d (resp., noisy decision trees of cost d) we get the same bounds on their $L_{1,\ell}$ as those in Theorem 3.2.1 (resp., Theorem 3.2.2). This is captured in the following corollary.

Corollary 3.2.3. *Let $\mathcal{T}$ be a randomized parity decision tree of depth at most d on n variables. Then,*

$$\forall \ell \in [n] : L_\ell(\mathcal{T}) \leq d^{\ell/2} \cdot O(\ell \cdot \log(n))^\ell.$$

Let $\mathcal{T}'$ be a randomized noisy decision tree of cost at most d on n variables. Then,

$$\forall \ell \in [n] : L_\ell(\mathcal{T}') \leq O(d)^{\ell/2} \cdot (\ell \cdot \log(n))^{(\ell-1)/2}.$$

Quantum versus Randomized Query Complexity

Let $k \leq \log(n)$. Bansal and Sinha [BS21] gave a $\lceil k/2 \rceil$ versus $\widetilde{\Omega}\left(n^{1-1/k}\right)$ separation between the quantum and randomized query complexity of k-fold Forrelation. Our main application is an extension of Bansal and Sinha's lower bound for the model of randomized parity decision trees. This follows from their main technical result Theorem 2.0.12 and Theorem 3.2.1.

Corollary 3.2.4. *If $\mathcal{T}$ is a randomized parity decision tree of depth d computing k-fold Forrelation with success probability $\frac{1}{2}+\gamma$, then*

$$d \geq \gamma^2 \cdot \frac{n^{1-1/k}}{\mathsf{poly}(k)\log^2 n}.$$

Proof. We can amplify the success probability of the randomized parity decision tree from $1/2+\gamma$ to $1-2^{-5k}$ by repeating the query algorithm $O(k/\gamma^2)$ times independently and taking majority. This results in a randomized parity decision tree $\mathcal{T}'$ of depth $d' = O(d \cdot k/\gamma^2)$. Now, Corollary 3.2.3 gives $L_\ell(\mathcal{T}') \leq (d')^{\ell/2} \cdot O(\ell \cdot \log(n))^\ell$ for all ℓ. In particular, $L_\ell(\mathcal{T}') \leq t^\ell$ for all $\ell \leq k(k-1)$ where $t = O\left(\sqrt{d'} \cdot k(k-1) \cdot \log(n)\right)$.

This is also true for any restriction of $\mathcal{T}'$, since fixing variables to constants yields another randomized parity decision tree of depth at most d'. Combining the bounds on $L_\ell(\mathcal{T}')$ for $\ell \in \{k, \ldots, k(k-1)\}$ with Theorem 2.0.12 gives

$$d' \geq \frac{n^{1-1/k}}{O(k^{34}) \cdot \log^2(n)}$$

and thus the claimed bound. □

For constant k and $\gamma = 2^{-O(k)}$, we get a $\lceil k/2 \rceil$ versus $\widetilde{\Omega}\left(n^{1-1/k}\right)$ separation between the quantum query complexity and the randomized parity query complexity of k-fold Forrelation. We remark that separations in the reverse direction are also known: for the n-bit parity function, the (randomized) parity query complexity is 1 whereas the quantum query complexity is $\Omega(n)$ [MNR11].

Similarly, we can obtain the following corollary for noisy decision trees.

Corollary 3.2.5. *If $\mathcal{T}$ is a randomized noisy decision tree of cost at most d computing k-fold Forrelation with success probability $\frac{1}{2}+\gamma$, then*

$$d \geq \gamma^2 \cdot \frac{n^{1-1/k}}{\mathsf{poly}(k)\log(n)}.$$

Towards Communication Complexity Lower Bounds

We recall an open question from [GRT22], which, if true, would demonstrate that the randomized communication complexity of the Forrelation problem composed with the XOR gadget is $\widetilde{\Omega}(n^{1/2})$. The *simultaneous* quantum communication complexity of this problem is $O(\mathsf{polylog}(n))$ and the best known randomized lower bound is $\widetilde{\Omega}(n^{1/4})$ due to [GRT22]. See [GRT22] for more references on this problem.

Conjecture 3.2.6. *Let $f\colon \{\pm 1\}^n \times \{\pm 1\}^n \to \{0,1\}$ computed by a deterministic communication protocol of cost at most c. Let $h\colon \{\pm 1\}^n \to [0,1]$ defined by $h(z) = \mathbb{E}_{\boldsymbol{x}}[f(\boldsymbol{x}, \boldsymbol{x} \odot z)]$. Then $L_2(h) \leq c \cdot \mathsf{polylog}(n)$.*

We view Theorem 3.2.1 as a first step towards this conjecture. Indeed, for communication protocols that follow a parity decision tree strategy according to some tree $\mathcal{T}$, it is simple to verify that $h = \mathcal{T}$ (as functions), and thus $L_2(h) = L_2(\mathcal{T}) \leq c \cdot \mathsf{polylog}(n)$.

Application to Expander Random Walk

The work by Cohen, Peri, and Ta-Shma [CPTS21] showed that expander random walks fool symmetric functions and also general functions with bounded Fourier growth. To be more precise, assume $L_\ell(f) \leq t^\ell$ for all $\ell \geq 1$. Let G be an expander, with second eigenvalue $\lambda \ll 1/t^4$, where half of G's vertices are labeled by $+1$ and the rest are labeled by -1. Then the expected value of f on bits sampled by an $(m-1)$-step random walk on G is approximately the value it would get on a uniformly random string in $\{\pm 1\}^m$. Combined with our Theorem 3.2.1, this shows that if f can be computed by low-depth parity decision trees then f can be fooled by the expander random walk.

Fourier Bounds for Small-Size Parity Decision Trees

By a simple size-to-depth reduction, we also obtain Fourier growth bounds for parity decision trees of bounded size.

Corollary 3.2.7. *Let $\mathcal{T}$ be a parity decision tree of size at most $s > 1$ on n variables. Then*

$$\forall \ell \in [n] : L_{1,\ell}(f) \leq (\log(s))^{\ell/2} \cdot O(\ell \cdot \log(n))^{1.5\ell}.$$

Proof. We approximate $\mathcal{T}$ with error $\varepsilon = 1/n^\ell$ by another parity decision tree $\mathcal{T}'$ of depth $d = \lceil \log\left(s \cdot n^\ell\right) \rceil$, where we simply replace all nodes of depth d in $\mathcal{T}$ with leaves that return 0. Since there are at most s nodes in $\mathcal{T}$, the probability that a random input would reach one of the nodes of depth d is at most $2^{-d} \cdot s \leq 1/n^\ell$. Hence $\mathbf{Pr}_{\boldsymbol{x}}\left[\mathcal{T}(\boldsymbol{x}) \neq \mathcal{T}'(\boldsymbol{x})\right] \leq \varepsilon$. This implies that $\left|\widehat{\mathcal{T}}(S) - \widehat{\mathcal{T}'}(S)\right| \leq \varepsilon$ for any subset $S \subseteq [n]$. Thus,

$$L_\ell(\mathcal{T}) = \sum_{S:|S|=\ell} \left|\widehat{\mathcal{T}}(S)\right| \leq \sum_{S:|S|=\ell} \left(\left|\widehat{\mathcal{T}'}(S)\right| + \varepsilon\right) \leq L_\ell(\mathcal{T}') + 1.$$

Since $\mathcal{T}'$ is of depth at most $d = \lceil \log(s) + \ell \cdot \log(n) \rceil = O\left(\log(s) \cdot \ell \cdot \log(n)\right)$, we obtain our bound using Theorem 3.2.1. □

3.3 Proof Overview

Let $\ell \geq 1$ and $\varepsilon \in (0, 1/2]$. We will use $\lesssim, \gtrsim$ to hide minor factors like $\mathsf{polylog}(n/\varepsilon)$ and dependence on ℓ. We first describe the proof for standard decision trees and then show how to generalize to parity decision trees.

Standard Decision Trees

Let $\mathcal{T}$ be a decision tree and for simplicity, assume that every leaf is of depth d. Let $\boldsymbol{v}_0, \ldots, \boldsymbol{v}_d$ be a random root-to-leaf path in $\mathcal{T}$ and $\boldsymbol{v}^{(0)}, \ldots, \boldsymbol{v}^{(d)} \in \{-1, 0, 1\}^n$ denote the sequence of partial assignments, i.e., for $j \in [n]$ and $i \in \{0, \ldots, d\}$, let

$$\boldsymbol{v}_j^{(i)} = \begin{cases} 1 & \text{if } x_j \text{ is fixed to 1 before reaching } \boldsymbol{v}_i, \\ -1 & \text{if } x_j \text{ is fixed to } -1 \text{ before reaching } \boldsymbol{v}_i, \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

For $u \in \mathbb{R}^n$, we use u_S to denote $\prod_{j \in S} u_j$. Let $a_S = \mathsf{sgn}\left(\widehat{\mathcal{T}}(S)\right)$ for $|S| = \ell$ and 0 otherwise. Note that

$$\begin{aligned} \sum_{S:|S|=\ell} \left|\widehat{\mathcal{T}}(S)\right| &= \sum_{S:|S|=\ell} a_S \cdot \widehat{\mathcal{T}}(S) = \sum_{S:|S|=\ell} a_S \cdot \mathop{\mathbb{E}}_{\boldsymbol{v}_d}\left[\mathcal{T}(\boldsymbol{v}_d)\boldsymbol{v}_S^{(d)}\right] \\ &= \mathop{\mathbb{E}}_{\boldsymbol{v}_d}\left[\mathcal{T}(\boldsymbol{v}_d) \cdot \left(\sum_{S:|S|=\ell} a_S \cdot \boldsymbol{v}_S^{(d)}\right)\right]. \end{aligned} \tag{3.2}$$

Thus, to bound $\sum_{S:|S|=\ell} |\widehat{\mathcal{T}}(S)|$ it suffices to show that $\left|\sum_{S:|S|=\ell} a_S \cdot \boldsymbol{v}_S^{(d)}\right| \lesssim d^{\ell/2}$ in expectation.

Denote by $\boldsymbol{X}^{(i)} := \sum_{S:|S|=\ell} a_S \cdot \boldsymbol{v}_S^{(i)}$ for $i = 0, 1, \ldots, d$. We write $\boldsymbol{X}^{(d)}$ as a telescoping sum $\boldsymbol{X}^{(d)} = \sum_{i=1}^{d} \left(\boldsymbol{X}^{(i)} - \boldsymbol{X}^{(i-1)}\right)$. To analyze the difference sequence, observe that in the expression

$$\boldsymbol{X}^{(i)} - \boldsymbol{X}^{(i-1)} = \sum_{S:|S|=\ell} a_S \cdot \left(\boldsymbol{v}_S^{(i)} - \boldsymbol{v}_S^{(i-1)}\right),$$

if set S contributes to the sum, then S must include the bit queried at the $(i-1)$-th step of the path. Conditioning on $\boldsymbol{v}_0, \ldots, \boldsymbol{v}_{i-1}$, let $\boldsymbol{x}_j$ be the variable queried in $\boldsymbol{v}_{i-1}$, then we have

$$\boldsymbol{X}^{(i)} - \boldsymbol{X}^{(i-1)} = \sum_{S:|S|=\ell, j \in S} a_S \cdot \boldsymbol{v}_S^{(i)} = \boldsymbol{x}_j \cdot \left(\sum_{S:|S|=\ell, j \in S} a_S \cdot \boldsymbol{v}_{S\setminus\{j\}}^{(i-1)}\right).$$

Furthermore, we observe that the sum $\sum_{S:|S|=\ell, j \in S} a_S \cdot \boldsymbol{v}_{S\setminus\{j\}}^{(i-1)}$ is determined by $\boldsymbol{v}_{i-1}$; thus conditioning on $\boldsymbol{v}_0, \ldots, \boldsymbol{v}_{i-1}$ the value of $X^{(i)} - X^{(i-1)}$ is a random coin in $\{\pm 1\}$ multiplied by some fixed integer. In other words, $\boldsymbol{X}^{(0)}, \ldots, \boldsymbol{X}^{(d)}$ is a martingale with varying step sizes.

Recall that Azuma's inequality provides concentration bounds for martingales with step sizes bounded, thus now we need to bound $\left|\sum_{S:|S|=\ell, j \in S} a_S \cdot \boldsymbol{v}_{S\setminus\{j\}}^{(i-1)}\right|$, which is similar to our initial goal. Put differently, we wish to analyze the sum

$$\sum_{S' \subseteq [n]\setminus\{j\}:|S'|=\ell-1} a_{S'\cup\{j\}} \cdot \boldsymbol{v}_{S'}^{(i-1)},$$

which calls for an inductive argument on ℓ. In addition, since we eventually apply a union bound on all steps, we need to show that $\left|\sum_{S'} a_{S'\cup\{j\}} \boldsymbol{v}_{S'}^{(i-1)}\right|$ is bounded with high probability (and not just in expectation).

More generally, to carry an inductive argument we define for any set $T \subseteq [n], |T| \le \ell$ and any $i \in \{0,\dots,d\}$, the random variable

$$\boldsymbol{X}_T^{(i)} := \sum_{S\supseteq T:|S|=\ell} a_S \cdot \boldsymbol{v}_{S\setminus T}^{(i)} = \sum_{S'\subseteq \overline{T}:|S'|=\ell-|T|} a_{S'\cup T} \cdot \boldsymbol{v}_{S'}^{(i)}.$$

Note that our initial goal was to bound $\left|\boldsymbol{X}_\emptyset^{(d)}\right| = \left|\boldsymbol{X}^{(d)}\right|$, which is analyzed by (reverse) induction on $|T|$ going from larger sets to smaller sets as Lemma 3.3.1.

Lemma 3.3.1. *For all $t \in \{0,\dots,\ell\}$ and $\varepsilon > 0$, the probability that there exist $i \in \{0,\dots,d\}$ and $T \subseteq [n]$ of size at least t such that $\left|\boldsymbol{X}_T^{(i)}\right| \gtrsim d^{(\ell-t)/2}$ is at most $\varepsilon \cdot (\ell - t)$.*

The main observation for the proof is that $\boldsymbol{X}_T^{(0)}, \boldsymbol{X}_T^{(1)}, \dots, \boldsymbol{X}_T^{(d)}$ is a martingale whose difference sequence consists of terms of the form $\boldsymbol{X}_{T'}^{(i-1)}$ where $T \subsetneq T'$. To see this, if we are querying $\boldsymbol{x}_j$ at v_{i-1}, then

$$\boldsymbol{X}_T^{(i)} - \boldsymbol{X}_T^{(i-1)} = \begin{cases} 0 & j \in T, \\ \boldsymbol{x}_j \cdot \left(\sum_{j\notin S\subseteq\overline{T}} a_{S\cup T\cup\{j\}} \cdot \boldsymbol{v}_S^{(i-1)}\right) = \boldsymbol{x}_j \cdot \boldsymbol{X}_{T\cup j}^{(i-1)} & j \notin T. \end{cases}$$

Note that $\boldsymbol{X}_{T\cup j}^{(i-1)}$ depends only on the history until $\boldsymbol{v}_{i-1}$, and $\boldsymbol{x}_j$ is a uniformly random bit independent of this history, thus $\boldsymbol{X}_T^{(i)}$ is a martingale. The inductive hypothesis implies that with at least $1 - \varepsilon \cdot (\ell - t - 1)$ probability, $\left|\boldsymbol{X}_{T\cup j}^{(i-1)}\right| \lesssim d^{(\ell-t-1)/2}$ for all T of size t and $j \in [n] \setminus T$. Whenever this happens, Azuma's inequality implies that[1] with probability at least $1 - \varepsilon/(d \cdot n^t)$, we have

$$\left|\boldsymbol{X}_T^{(i)}\right| \lesssim 2\sqrt{\log(d \cdot n^t/\varepsilon)} \cdot \sqrt{\sum_{i=1}^{d} d^{\ell-t-1}} \lesssim d^{(\ell-t)/2}.$$

This, along with a union bound over T of size t and $i \in \{0,\dots,d\}$ completes the inductive step. The Fourier growth bound for noisy decision trees can be proved using a similar approach.

[1] Technically this is not true, since a martingale after conditioning may not still be a martingale. We handle this by truncating the martingale when a bad event happens instead of conditioning on the good event.

Parity Decision Trees

The basic approach is as before. Let $\mathcal{T}$ be a parity decision tree. As in (3.1), we use $\boldsymbol{v}_i$ and $\boldsymbol{v}^{(i)}$ to denote the random walk and the partial assignments to the variables respectively. We say $\boldsymbol{v}_i$ is k-*clean* if

$$\forall S \subseteq [n], |S| \leq k, \quad \boldsymbol{v}_S^{(i)} = \begin{cases} 1 & \text{if } \boldsymbol{x}_S \text{ is fixed to 1 before reaching } \boldsymbol{v}_i, \\ -1 & \text{if } \boldsymbol{x}_S \text{ is fixed to } -1 \text{ before reaching } \boldsymbol{v}_i, \\ 0 & \text{otherwise.} \end{cases} \tag{3.3}$$

For (3.2) to be true, we need that at least $\boldsymbol{v}_d$ is ℓ-clean. Note that this is not always true,[2] but it is useful as it simplifies the study of high-level Fourier coefficients. To address this issue, we define a *cleanup* process for parity decision trees in which we make additional queries to ensure that certain key nodes are k-clean. We do this by recursively cleaning nodes in a top-down fashion so that for every node v in the original tree $\mathcal{T}$, any node v' in the new tree $\mathcal{T}'$ obtained at the end of the cleanup step for v is k-clean.

The cleanup process is simple to describe: let $v_1, \ldots, v_d$ be any root-to-leaf path in $\mathcal{T}$. Assume we have completed the cleanup process for $v_1, \ldots, v_{i-1}$. We then query the parity at v_i. While there exists a (minimal) set S violating (3.3), we pick and query an arbitrary coordinate in S. Once (3.3) is satisfied, we proceed to the cleanup process for v_{i+1}. This process increases the depth by a factor of at most k. We set $k = \Theta(\ell \cdot \log(n))$ and work with the new tree $\mathcal{T}'$ of depth $D \leq k \cdot d$.

Let $\boldsymbol{v}_0, \ldots, \boldsymbol{v}_D$ be a random root-to-leaf path in $\mathcal{T}'$ and $\boldsymbol{I}_i, i \in [D]$ be the set of coordinates fixed due to the query at $\boldsymbol{v}_{i-1}$. Note that this set might be of size larger than 1.[3] It follows from simple linear algebra that $\sum_{i=1}^{D} |\boldsymbol{I}_i| \leq D$. Since $\boldsymbol{v}_D$ is k-clean, (3.2) holds. Defining $\boldsymbol{X}_T^{(i)}$ exactly as before, our goal is to prove Lemma 3.3.1 with D instead of d. The proof is still by induction on $\ell - t$. It turns out that $\boldsymbol{X}_T^{(0)}, \boldsymbol{X}_T^{(1)}, \ldots, \boldsymbol{X}_T^{(D)}$ is no longer a martingale; instead, $\boldsymbol{X}_T^{(i)} - \boldsymbol{X}_T^{(i-1)} = \boldsymbol{Y}_i + \boldsymbol{Z}_i$ where

$$\boldsymbol{Y}_i := \sum_{\substack{\emptyset \neq J \subseteq \boldsymbol{I}_i \cap \overline{T} \\ |J| \text{ is even}}} \boldsymbol{x}_J \cdot \boldsymbol{X}_{J \cup T}^{(i-1)} \quad \text{and} \quad \boldsymbol{Z}_i := \sum_{\substack{\emptyset \neq J \subseteq \boldsymbol{I}_i \cap \overline{T} \\ |J| \text{ is odd}}} \boldsymbol{x}_J \cdot \boldsymbol{X}_{J \cup T}^{(i-1)}. \tag{3.4}$$

and $\boldsymbol{Z}_i$ (resp., $\boldsymbol{Y}_i$) is an odd (resp., even) polynomial of degree at most ℓ over the newly fixed variables $\{\boldsymbol{x}_j \mid j \in \boldsymbol{I}_i\}$. Conditioning on $\boldsymbol{v}_{i-1}$, every pair of random bits $(\boldsymbol{x}_j, \boldsymbol{x}_{j'})$ from $\{\boldsymbol{x}_j \mid j \in \boldsymbol{I}_i\}$ is either identical ($\boldsymbol{x}_j \equiv \boldsymbol{x}_{j'}$) or opposite ($\boldsymbol{x}_j \equiv -\boldsymbol{x}_{j'}$), which means $\boldsymbol{Y}_i$ is a constant and $\boldsymbol{Z}_i$ can be written as $\boldsymbol{z}_i \cdot |\boldsymbol{Z}_i|$ where $|\boldsymbol{Z}_i|$ is a constant and $\boldsymbol{z}_i \sim \{\pm 1\}$.

For now, let us ignore $\boldsymbol{Y}_i$ and assume that we have a martingale $\boldsymbol{X}_T^{(i)}$ such that $\boldsymbol{X}_T^{(i)} - \boldsymbol{X}_T^{(i-1)} = \boldsymbol{z}_i \cdot |\boldsymbol{Z}_i|$, where $\boldsymbol{z}_i \sim \{\pm 1\}$ is a uniformly random bit independent of $\boldsymbol{z}_0, \ldots, \boldsymbol{z}_{i-1}$ and

[2] For example, let $S = \{1, 2\}$ and consider the parity decision tree whose only query is $x_1 x_2$. At any leaf, the value of $x_1 x_2$ is fixed, however, the values of x_1 and x_2 are free, hence S violates (3.3).

[3] For example, suppose we query $x_1 x_2$, $x_1 x_3$, $x_1 x_4$ and finally x_1. Then, the last query reveals 4 coordinates.

$|\boldsymbol{Z}_i|$ depends only on $\boldsymbol{v}_{i-1}$. Combined with an adaptive version of Azuma's inequality, we only need to show the sum of squares of step sizes $\sum_{i=1}^{D} |\boldsymbol{Z}_i|^2 \lesssim D^{\ell-t}$ to prove $\left|\boldsymbol{X}_T^{(i)}\right| \lesssim D^{(\ell-t)/2}$. By the induction hypothesis, with probability at least $1 - \varepsilon \cdot (\ell - t - 1)$ the coefficients of $\boldsymbol{Z}_i$ are bounded appropriately. Since $\sum_{i=1}^{D} |\boldsymbol{I}_i| \le D$ and in particular $|\boldsymbol{I}_i| \le D$, we have

$$|\boldsymbol{Z}_i| \le \sum_{\text{odd } j \ge 1} \binom{|\boldsymbol{I}_i|}{j} \cdot \max_{|T'|=j+t} \left|\boldsymbol{X}_{T'}^{(i-1)}\right| \lesssim \sum_{j \ge 1}^{\ell-t} \binom{|\boldsymbol{I}_i|}{j} \cdot D^{(\ell-j-t)/2} \lesssim |\boldsymbol{I}_i| \cdot D^{(\ell-t-1)/2}$$

and thus $\sum_{i=1}^{D} |\boldsymbol{Z}_i|^2 \lesssim D^2 \cdot D^{\ell-t-1}$. This is too loose for our purpose.

We instead try to bound the sum of squares of step sizes *with high probability*. Imagine for now that $\boldsymbol{v}_{i-1}$ is 2-clean.[4] Then, the variables $\{\boldsymbol{x}_j \mid j \in I_i\}$ are 2-wise independent conditioning on $\boldsymbol{v}_{i-1}$. This gives

$$\begin{aligned}
\mathbb{E}\left[|\boldsymbol{Z}_i|^2 \,\middle|\, \boldsymbol{v}_{i-1}\right] &\le \sum_{\text{odd } j \ge 1} \binom{|\boldsymbol{I}_i|}{j} \cdot \max_{|T'|=j+t} \left|\boldsymbol{X}_{T'}^{(i-1)}\right|^2 \\
&\lesssim \sum_{j \ge 1}^{\ell-t} \binom{|\boldsymbol{I}_i|}{j} \cdot D^{\ell-j-t} \\
&\lesssim |\boldsymbol{I}_i| \cdot D^{\ell-t-1}
\end{aligned}$$

and thus $\mathbb{E}\left[\sum_{i=1}^{D} |\boldsymbol{Z}_i|^2\right] \lesssim D^{\ell-t}$. To show this bound holds with high probability, we use concentration properties of degree-ℓ polynomials under k-wise independent distributions for $k \gg \ell$.

In the actual proof, we proceed by conditioning on $C(\boldsymbol{v}_{i-1})$, the nearest ancestor of $\boldsymbol{v}_{i-1}$ that is k-clean, instead of conditioning on $\boldsymbol{v}_{i-1}$, which allows to remove the assumption that $\boldsymbol{v}_{i-1}$ is 2-clean. This is because the queries within a cleanup step are non-adaptive, thus $\boldsymbol{Z}_i$ depends only on $C(\boldsymbol{v}_{i-1})$ and not on $\boldsymbol{v}_{i-1}$.

Meanwhile, although $\boldsymbol{X}_T^{(i)}$ is not quite a martingale sequence (due to $\boldsymbol{Y}_i$) and the step sizes (i.e., $|\boldsymbol{Z}_i|$) are adaptive and not always bounded, we are nonetheless able to prove an adaptive version of Azuma's inequality of the form $\mathbf{Pr}\left[\max_{i \in [D]} \left|\boldsymbol{X}_T^{(i)}\right| \ge \mu + t \cdot \sigma\right] \le e^{-\Omega\left(t^2\right)} + \varepsilon$ provided $\mathbf{Pr}\left[\left(\sum_{i=1}^{D} |\boldsymbol{Y}_i| \le \mu\right) \wedge \left(\sum_{i=1}^{D} |\boldsymbol{Z}_i|^2 \le \sigma^2\right)\right] \ge 1 - \varepsilon$. Then it suffices to bound $\sum_{i=1}^{D} |\boldsymbol{Y}_i|$ similarly to $\sum_{i=1}^{D} |\boldsymbol{Z}_i|^2$ above.

Related Work

We remark that our proof for level-ℓ Fourier growth (even when specialized to the case of standard decision trees) differs from the proofs appearing in [Tal20] and [SSW23]. There,

[4] This assumption immediately implies that $|\boldsymbol{I}_i| \le 1$ and trivially proves our inequality, however, this type of reasoning doesn't generalize to the case when $\boldsymbol{v}_{i-1}$ is not 2-clean.

the results were based on decompositions of decision trees. We view our martingale approach as natural and intuitive. We wonder if one can obtain the tight results from [SSW23] using this approach. It seems that the main bottleneck is a union bound on events related to all sets $T \subseteq [n]$ of size at most ℓ.

Our bounds for level-1 improve those obtained by [BTW15]. They prove that $L_1(\mathcal{T}) \leq O(\sqrt{p \cdot d})$ when $p = \mathbf{Pr}_{\boldsymbol{x}}[\mathcal{T}(\boldsymbol{x}) = 1]$, whereas we obtain a bound of

$$L_1(\mathcal{T}) \leq O\left(p\sqrt{d} \cdot \log(1/p)\right).$$

In particular, our bound is almost quadratically better for small values of p. It remains open whether the bound can be further improved to $O\left(p\sqrt{d \cdot \log(1/p)}\right)$, which is the optimal bound for standard decision trees.

Our cleanup technique is inspired by [BTW15], which used cleanup to prove their level-1 bound. However, our proof strategies and the way we use the cleanup procedure is quite different than that of [BTW15].

3.4 Adaptive Azuma's Inequality

We show an adaptive version of Azuma's inequality for martingales.

Lemma 3.4.1 (Adaptive Azuma's inequality)**.** *Let $\boldsymbol{X}^{(0)}, \ldots, \boldsymbol{X}^{(D)}$ be a martingale and $\boldsymbol{\Delta}^{(1)}, \ldots, \boldsymbol{\Delta}^{(D)}$ be a sequence of magnitudes such that $\boldsymbol{X}^{(0)} = 0$ and $\boldsymbol{X}^{(i)} = \boldsymbol{X}^{(i-1)} + \boldsymbol{\Delta}^{(i)} \cdot \boldsymbol{z}^{(i)}$ for $i \in [D]$, where if conditioning on $\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(i-1)}$,*

(1) $\boldsymbol{z}^{(i)}$ is a mean-zero random variable and $\left|\boldsymbol{z}^{(i)}\right| \leq 1$ always holds;

(2) $\boldsymbol{\Delta}^{(i)}$ is a fixed value.

If there exists some constant $U \geq 0$ such that $\sum_{i=1}^{D} \left|\boldsymbol{\Delta}^{(i)}\right|^2 \leq U$ always holds, then for any $\beta \geq 0$ we have

$$\mathbf{Pr}\left[\max_{i=0,1,\ldots,D} \left|\boldsymbol{X}^{(i)}\right| \geq \beta \cdot \sqrt{2U}\right] \leq 2 \cdot e^{-\beta^2/2}.$$

We will use the definition of sub-Gaussian random variables.

Definition 3.4.2 (Sub-Gaussian)**.** *A random variable $\boldsymbol{x}$ is Δ-sub-Gaussian if $\mathbb{E}\left[e^{t \cdot \boldsymbol{x}}\right] \leq e^{t^2 \Delta^2}$ holds for all $t \in \mathbb{R}$.*

Now we prove the following sub-Gaussian adaptive Azuma's inequality, which generalizes Lemma 3.4.1

Lemma 3.4.3 (Sub-Gaussian adaptive Azuma's inequality)**.** *Let $\boldsymbol{X}^{(0)}, \ldots, \boldsymbol{X}^{(D)}$ be a martingale with respect to a filtration $\left(\mathcal{F}^{(i)}\right)_{i=0}^{D}$ and $\boldsymbol{\Delta}^{(1)}, \ldots, \boldsymbol{\Delta}^{(D)}$ be a sequence of magnitudes*

such that $\boldsymbol{X}^{(0)} = 0$ *and* $\boldsymbol{X}^{(i)} = \boldsymbol{X}^{(i-1)} + \boldsymbol{\delta}^{(i)}$ *for* $i \in [D]$*, where if conditioning on* $\mathcal{F}^{(i-1)}$*,* $\boldsymbol{\delta}^{(i)}$ *is a* $\boldsymbol{\Delta}^{(i)}$*-sub-Gaussian random variable and* $\boldsymbol{\Delta}^{(i)}$ *is a fixed value.*

If there exists some constant $U \geq 0$ *such that* $\sum_{i=1}^{D} \left|\boldsymbol{\Delta}^{(i)}\right|^2 \leq U$ *always holds, then for any* $\beta \geq 0$ *we have*

$$\mathbf{Pr}\left[\max_{i=0,1,\ldots,D} \left|\boldsymbol{X}^{(i)}\right| \geq \beta \cdot \sqrt{2U}\right] \leq 2 \cdot e^{-\beta^2/2}.$$

Proof. The bound holds trivially when $\beta = 0$, hence we assume $\beta > 0$ from now on. We construct another martingale $\widehat{\boldsymbol{X}}^{(0)}, \ldots, \widehat{\boldsymbol{X}}^{(D)}$ as follows:

$$\widehat{\boldsymbol{X}}^{(i)} = \begin{cases} \boldsymbol{X}^{(i)} & 0 \leq i \leq d, \\ \boldsymbol{X}^{(d)} & i > d, \end{cases} \quad \text{where} \quad d = \min\{D\} \cup \left\{ i \in \{0, 1 \ldots, D\} \,\middle|\, \left|\boldsymbol{X}^{(i)}\right| \geq \beta \cdot \sqrt{2U} \right\}.$$

We write $\widehat{\boldsymbol{\delta}}^{(i)} = \widehat{\boldsymbol{X}}^{(i)} - \widehat{\boldsymbol{X}}^{(i-1)}$, then $\widehat{\boldsymbol{\delta}}^{(i)} = \boldsymbol{\delta}^{(i)}$ for all $i \leq d$; and $\widehat{\boldsymbol{\delta}}^{(i)} \equiv 0$ for all $i > d$. Let $\widehat{\Delta}^{(i)} = \boldsymbol{\Delta}^{(i)}$ for all $i \leq d$; and $\widehat{\boldsymbol{\Delta}}^{(i)} \equiv 0$ for all $i > d$. Thus $\widehat{\boldsymbol{\delta}}^{(i)}$ is $\widehat{\boldsymbol{\Delta}}^{(i)}$-sub-Gaussian given $\mathcal{F}^{(i-1)}$; and

$$\sum_{i=1}^{D} \left|\widehat{\boldsymbol{\Delta}}^{(i)}\right|^2 = \sum_{i=1}^{d} \left|\boldsymbol{\Delta}^{(i)}\right|^2 \leq U.$$

Moreover, we have

$$\mathbf{Pr}\left[\max_{i=0,1,\ldots,D} \left|\boldsymbol{X}^{(i)}\right| \geq \beta \cdot \sqrt{2U}\right] = \mathbf{Pr}\left[\left|\widehat{\boldsymbol{X}}^{(D)}\right| \geq \beta \cdot \sqrt{2U}\right].$$

Let $t > 0$ be a parameter and we bound $\mathbb{E}\left[e^{t \cdot \widehat{\boldsymbol{X}}^{(D)}}\right]$ as follows

$$\begin{aligned}
\mathbb{E}\left[e^{t \cdot \widehat{\boldsymbol{X}}^{(D)}}\right] &= \mathop{\mathbb{E}}_{\mathcal{F}^{(D-1)}}\left[e^{t \cdot \widehat{\boldsymbol{X}}^{(D-1)}} \cdot \mathop{\mathbb{E}}_{\mathcal{F}^{(D)}}\left[e^{t \cdot \left(\widehat{\boldsymbol{X}}^{(D)} - \widehat{\boldsymbol{X}}^{(D-1)}\right)} \,\middle|\, \mathcal{F}^{(D-1)}\right]\right] && (3.5) \\
&= \mathop{\mathbb{E}}_{\mathcal{F}^{(D-1)}}\left[e^{t \cdot \widehat{\boldsymbol{X}}^{(D-1)}} \cdot \mathop{\mathbb{E}}_{\mathcal{F}^{(D)}}\left[e^{t \cdot \widehat{\boldsymbol{\delta}}^{(D)}} \,\middle|\, \mathcal{F}^{(D-1)}\right]\right] && (3.6) \\
&\leq \mathop{\mathbb{E}}_{\mathcal{F}^{(D-1)}}\left[e^{t \cdot \widehat{\boldsymbol{X}}^{(D-1)}} \cdot e^{t^2 \left(\widehat{\boldsymbol{\Delta}}^{(D)}\right)^2}\right] && \text{(since } \widehat{\boldsymbol{\delta}}^{(D)} \text{ is } \widehat{\boldsymbol{\Delta}}^{(D)}\text{-sub-Gaussian)} \\
&\leq \mathop{\mathbb{E}}_{\mathcal{F}^{(D-1)}}\left[e^{t \cdot \widehat{\boldsymbol{X}}^{(D-1)}} \cdot e^{t^2 \left(U - \left(\widehat{\boldsymbol{\Delta}}^{(1)}\right)^2 - \cdots - \left(\widehat{\boldsymbol{\Delta}}^{(D-1)}\right)^2\right)}\right] \\
&\leq \mathop{\mathbb{E}}_{\mathcal{F}^{(D-2)}}\left[e^{t \cdot \widehat{\boldsymbol{X}}^{(D-2)}} \cdot e^{t^2 \left(U - \left(\widehat{\boldsymbol{\Delta}}^{(1)}\right)^2 - \cdots - \left(\widehat{\boldsymbol{\Delta}}^{(D-1)}\right)^2\right)} e^{t^2 \left(\widehat{\boldsymbol{\Delta}}^{(D-1)}\right)^2}\right] \\
& && \text{(similar to (3.5) and (3.6))} \\
&= \mathop{\mathbb{E}}_{\mathcal{F}^{(D-2)}}\left[e^{t \cdot \widehat{\boldsymbol{X}}^{(D-2)}} \cdot e^{t^2 \left(U - \left(\widehat{\boldsymbol{\Delta}}^{(1)}\right)^2 - \cdots - \left(\widehat{\boldsymbol{\Delta}}^{(D-2)}\right)^2\right)}\right] \\
&\leq \cdots \leq \mathop{\mathbb{E}}_{\mathcal{F}^{(D-k)}}\left[e^{t \cdot \widehat{\boldsymbol{X}}^{(D-k)}} \cdot e^{t^2 \left(U - \left(\widehat{\boldsymbol{\Delta}}^{(1)}\right)^2 - \cdots - \left(\widehat{\boldsymbol{\Delta}}^{(D-k)}\right)^2\right)}\right] \leq \cdots
\end{aligned}$$

$$\leq e^{t^2 U}. \tag{3.7}$$

Setting $t = \beta/\sqrt{2U}$ implies that

$$\mathbf{Pr}\left[\widehat{\boldsymbol{X}}^{(D)} \geq \beta \cdot \sqrt{2U}\right] \leq \frac{\mathbb{E}\left[e^{t \cdot \widehat{\boldsymbol{X}}^{(D)}}\right]}{e^{t \cdot \beta \cdot \sqrt{2U}}} \leq \frac{e^{t^2 U}}{e^{\beta^2}} = e^{-\beta^2/2}.$$

Similarly we can show $\mathbf{Pr}\left[\widehat{\boldsymbol{X}}^{(D)} \leq -\beta \cdot \sqrt{2U}\right] \leq e^{-\beta^2/2}$, which completes the proof by a union bound. □

For our Lemma 3.4.1, we need the following fact.

Fact 3.4.4. *Let $\boldsymbol{x}$ be a mean-zero random variable and assume $|\boldsymbol{x}| \leq \Delta$ always holds. Then $\boldsymbol{x}$ is Δ-sub-Gaussian.*

Proof. Note that $e^{t \cdot x}$ is convex for all $t \in \mathbb{R}$. By Jensen's inequality, we have

$$\mathbb{E}\left[e^{t \cdot \boldsymbol{x}}\right] \leq \frac{1}{2}\left(e^{-t\Delta} + e^{t\Delta}\right) = \sum_{i=0}^{+\infty} \frac{(t\Delta)^{2i}}{(2i)!} \leq \sum_{i=0}^{+\infty} \frac{(t\Delta)^{2i}}{i!} = e^{t^2\Delta^2}.$$ □

As a corollary of Lemma 3.4.3 and Fact 3.4.4, we obtain Lemma 3.4.1. Next, we generalize Lemma 3.4.1 as Lemma 3.4.5, which will be frequently used later.

Lemma 3.4.5. *Let $m \geq 1$ be an integer. For each $t \in [m]$, let $\boldsymbol{X}_t^{(0)}, \ldots, \boldsymbol{X}_t^{(D)}$ be a sequence of random variables and $\boldsymbol{\Delta}_t^{(1)}, \ldots, \boldsymbol{\Delta}_t^{(D)}$ be a sequence of magnitudes such that $\boldsymbol{X}_t^{(0)} = 0$ and $\boldsymbol{X}_t^{(i)} = \boldsymbol{X}_t^{(i-1)} + \boldsymbol{\Delta}_t^{(i)} \cdot \boldsymbol{z}_t^{(i)} + \boldsymbol{\mu}_t^{(i)}$ for $i \in [D]$, where if conditioning on $\boldsymbol{z}_t^{(1)}, \ldots, \boldsymbol{z}_t^{(i-1)}$,*

1. *$\boldsymbol{z}_t^{(i)}$ is a mean-zero random variable and $\left|\boldsymbol{z}_t^{(i)}\right| \leq 1$ always holds;*
2. *$\boldsymbol{\Delta}_t^{(i)}$ is a fixed value and $\boldsymbol{\mu}_t^{(i)}$ is a random variable.*

If there exist some constants $U, V \geq 0$ and $\eta \in [0, 1]$ such that

$$\mathbf{Pr}\left[\exists t \in [m],\ \left(\sum_{i=1}^{D} \left|\boldsymbol{\Delta}_t^{(i)}\right|^2 > U\right) \vee \left(\sum_{i=1}^{D} \left|\boldsymbol{\mu}_t^{(i)}\right| > V\right)\right] \leq \eta,$$

then for any $\beta \geq 0$ we have

$$\mathbf{Pr}\left[\exists t \in [m],\ \max_{i=0,1,\ldots,D} \left|\boldsymbol{X}_t^{(i)}\right| \geq V + \beta \cdot \sqrt{2U}\right] \leq \eta + 2m \cdot e^{-\beta^2/2}.$$

Proof. We divide the proof into the following two cases.

The $\eta = 0$ Case. Let $\widehat{\boldsymbol{X}}_t^{(i)} = \boldsymbol{X}_t^{(i)} - \sum_{j=1}^{i} \boldsymbol{\mu}_t^{(j)}$ for each t and i. Then

$$\left|\boldsymbol{X}_t^{(i)}\right| = \left|\widehat{\boldsymbol{X}}_t^{(i)} + \sum_{j=1}^{i} \boldsymbol{\mu}_t^{(j)}\right| \le V + \left|\widehat{\boldsymbol{X}}_t^{(i)}\right|.$$

By a union bound, it suffices to show for any fixed t, we have

$$\mathbf{Pr}\left[\max_{i=0,1,\ldots,D} \left|\widehat{\boldsymbol{X}}_t^{(i)}\right| \ge \beta \cdot \sqrt{2U}\right] \le 2 \cdot e^{-\beta^2/2},$$

which follows from Lemma 3.4.1.

The $\eta \ge 0$ Case. Consider $\widetilde{\boldsymbol{X}}_t^{(0)}, \ldots, \widetilde{\boldsymbol{X}}_t^{(D)}$ defined by setting $\widetilde{\boldsymbol{X}}_t^{(0)} = 0$ and $\widetilde{\boldsymbol{X}}_t^{(i)} = \widetilde{\boldsymbol{X}}_t^{(i-1)} + \widetilde{\boldsymbol{\Delta}}_t^{(i)} \cdot \boldsymbol{z}_t^{(i)} + \widetilde{\boldsymbol{\mu}}_t^{(i)}$, where

$$\widetilde{\boldsymbol{\Delta}}_t^{(i)} = \begin{cases} \boldsymbol{\Delta}_t^{(i)} & \sum_{j=1}^{i} \left|\boldsymbol{\Delta}_t^{(j)}\right|^2 \le U, \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad \widetilde{\boldsymbol{\mu}}_t^{(i)} = \begin{cases} \boldsymbol{\mu}_t^{(i)} & \sum_{j=1}^{i} \left|\boldsymbol{\mu}_t^{(j)}\right| \le V, \\ 0 & \text{otherwise.} \end{cases}$$

Then Item 1 and Item 2 hold for $\left(\widetilde{\boldsymbol{X}}_t^{(i)}\right)_{t,i}$ and $\left(\widetilde{\boldsymbol{\Delta}}_t^{(i)}\right)_{t,i}, \left(\widetilde{\boldsymbol{\mu}}_t^{(i)}\right)_{t,i}$. Note that we always have

$$\mathbf{Pr}\left[\exists t \in [m], i \in \{0, 1 \ldots, D\}, \widetilde{\boldsymbol{X}}_t^{(i)} \ne \boldsymbol{X}_t^{(i)}\right] \le \eta$$

and

$$\sum_{i=1}^{D} \left|\widetilde{\boldsymbol{\Delta}}_t^{(i)}\right|^2 \le U, \sum_{i=1}^{D} \left|\widetilde{\boldsymbol{\mu}}_t^{(i)}\right| \le V.$$

Hence from the previous case, we have

$$\begin{aligned} &\mathbf{Pr}\left[\exists t \in [m], \max_{i=0,1,\ldots,D} \left|\boldsymbol{X}_t^{(i)}\right| \ge V + \beta \cdot \sqrt{2U}\right] \\ \le\ &\mathbf{Pr}\left[\exists t \in [m], i \in \{0, 1 \ldots, D\}, \widetilde{\boldsymbol{X}}_t^{(i)} \ne \boldsymbol{X}_t^{(i)}\right] \\ &\quad + \mathbf{Pr}\left[\exists t \in [m], \max_{i=0,1,\ldots,D} \left|\widetilde{\boldsymbol{X}}_t^{(i)}\right| \ge V + \beta \cdot \sqrt{2U}\right] \\ \le\ &\eta + 2m \cdot e^{-\beta^2/2}. \end{aligned}$$

□

3.5 How to Clean Up Parity Decision Trees

In this section we show how to *clean up* a given parity decision tree to make it easier to analyze.

It will be useful to identify $\mathbb{F}_2^n$ with $\{\pm 1\}^n$ by $\mathsf{Enc}\colon (x_1, \ldots, x_n) \mapsto ((-1)^{x_1}, \ldots, (-1)^{x_n})$. For a subset $X \subseteq \mathbb{F}_2^n$ we will denote $\mathsf{Enc}(X) = \{\mathsf{Enc}(x) : x \in X\}$. Thus, we may think of Boolean functions also as $f\colon \mathbb{F}_2^n \to \{0,1\}$. We observe that under this representation of the input, a parity decision tree $\mathcal{T} : \mathbb{F}_2^n \to \{0,1\}$ indeed queries parity functions (i.e., linear functions over $\mathbb{F}_2$) of the input bits $x \in \mathbb{F}_2^n$ and decides whether to go left or right based on their outcome. Thus, the set of all possible inputs in $\mathbb{F}_2^n$ that reach a given node in a parity decision tree is an affine subspace of $\mathbb{F}_2^n$.

Notation 3.5.1. *Let $\mathcal{T}\colon \{\pm 1\}^n \to \{0,1\}$ be a parity decision tree and let v be a node in it.*

- *We use $\mathcal{P}_v \subseteq \{\pm 1\}^n$ to denote the set of all points reaching node v. Note that $\mathcal{P}_v = \mathsf{Enc}(H_v + a)$ where H_v is a linear subspace of $\mathbb{F}_2^n$ of dimension $n - \mathsf{depth}(v)$ and $a \in \mathbb{F}_2^n$.*
- *For any $S \subseteq [n]$, we define $\widehat{\mathcal{P}_v}(S) = \mathbb{E}_{\boldsymbol{x} \sim \mathcal{P}_v}[\boldsymbol{x}_S]$.*
- *We use $\mathcal{S}_v$ to denote all fully correlated sets with $\mathcal{P}_v$, i.e.,*

$$\mathcal{S}_v = \left\{ S \subseteq [n] \,\middle|\, \widehat{\mathcal{P}_v}(S) \in \{\pm 1\} \right\}.$$

  We observe that if $\mathcal{P}_v = \mathsf{Enc}(H_v + a)$, then $\mathcal{S}_v = H_v^{\perp}$. Additionally, if the queries on the path from root to v are $Q_{v_0}, \ldots, Q_{v_{i-1}}$, then $\mathcal{S}_v = \mathsf{Span}\langle\{Q_{v_0}, \ldots, Q_{v_{i-1}}\}\rangle$.
- *If v is an internal node, then define $J(v)$ as the set of newly fixed coordinates after querying Q_v, i.e., $i \in J(v)$ iff $\{i\} \notin \mathcal{S}_v$ but $\{i\} \in \mathsf{Span}\,\langle \mathcal{S}_v \cup \{Q_v\}\rangle$.*

The following simple fact shows that there is no weakly correlated set.

Fact 3.5.2. *For any parity decision tree $\mathcal{T}$ and any node v in $\mathcal{T}$, $\widehat{\mathcal{P}_v}(S) \in \{+1, 0, -1\}$ holds for any set S.*

Proof. Since $\mathcal{P}_v = \mathsf{Enc}(H_v + a)$ where $H_v + a$ is an affine subspace, $\mathcal{P}_v$ falls into one of the following 3 cases: (a) all points in $\mathcal{P}_v$ satisfy $\chi_S(x) = 1$, (b) all points satisfy $\chi_S(x) = -1$, (c) exactly half of the points satisfy $\chi_S(x) = 1$. □

Let $\mathcal{S} \subseteq \mathbb{F}_2^n$ be a subspace and $S \subseteq [n]$. For simplicity, we write $S \in \mathcal{S}$ iff the indicator vector of S is contained in $\mathcal{S}$. Now we describe the desired property: *k-clean*.

Definition 3.5.3 (k-Clean Subspace and Mess-Witness)**.** *Let k be a positive integer. A subspace $\mathcal{S}$ is* k-clean *if for any set $S \in \mathcal{S}$ such that $|S| \le k$, we have that $\{i\} \in \mathcal{S}$ holds for any $i \in S$.*

Moreover, when $\mathcal{S}$ is not k-clean, we say i is a mess-witness *if there exists some $S \ni i, |S| \le k$ such that $S \in \mathcal{S}$ but $\{i\} \notin \mathcal{S}$.*

Definition 3.5.4 (k-Clean Parity Decision Tree)**.** *A parity decision tree $\mathcal{T}$ is* k-clean *if the following holds:*

- *For any internal node v, either (a) $\mathcal{S}_v$ is k-clean, or (b) $Q_v = \{i\}$ where i is a mess-witness for $\mathcal{S}_v$. Moreover, we say v is* k-clean *if (a) holds; and we say v is* cleaning *if (b) holds.*
- *For any leaf v, $\mathcal{S}_v$ is k-clean (in such a case, we say that v is* k-clean*).*
- *For any k-clean internal node v, $\mathcal{T}_v$ starts with $\ell(v)$ non-adaptive queries[5] where $\ell(v) \geq 1$. In addition, for any $i \in \{1, \ldots, \ell(v) - 1\}$, any node of depth i in $\mathcal{T}_v$ is cleaning; and all node of depth $\ell(v)$ are k-clean.*[6]

Example 3.5.5. *If $\mathcal{T}$ is a decision tree (i.e., $|Q_v| \equiv 1$ for any internal node v) then it is k-clean for any k, where each internal node is k-clean.*

If $\mathcal{T}$ is the depth-1 parity decision tree for $\mathcal{T}(x) = x_1x_2x_3$ (i.e., $\mathcal{T}$ only has a root v_0 querying $Q_{v_0} = \{1, 2, 3\}$), then it is 2-clean but not 3-clean, since for either leaf v we have $\{1, 2, 3\} \in \mathcal{S}_v$ but $\{1\} \notin \mathcal{S}_v$.

The benefit of having a k-clean parity decision tree is that it makes the expression of Fourier coefficients simpler.

Lemma 3.5.6. *Let $\mathcal{T} \colon \{\pm 1\}^n \to \{0, 1\}$ be a k-clean parity decision tree and let S be a set of size $\ell \leq k$. Let $\boldsymbol{v}_0, \ldots, \boldsymbol{v}_d$ be a random root-to-leaf path. Define $\boldsymbol{v}^{(0)}, \ldots, \boldsymbol{v}^{(d)} \in \{-1, 0, +1\}^n$ by setting $\boldsymbol{v}_j^{(i)} = \widehat{\mathcal{P}_{\boldsymbol{v}_i}}(j)$ for each i, j. Recall that $\boldsymbol{v}_S^{(d)} = \prod_{j \in S} \boldsymbol{v}_j^{(d)}$. Then we have*

$$\widehat{\mathcal{T}}(S) = \mathop{\mathbb{E}}_{\boldsymbol{v}_0, \ldots, \boldsymbol{v}_d} \left[\mathcal{T}(\boldsymbol{v}_d) \cdot \boldsymbol{v}_S^{(d)} \right].$$

Proof. Observe that for any $j \in J(\boldsymbol{v}_i) \subseteq J$, the j-th coordinate is fixed after querying $Q_{\boldsymbol{v}_i}$. Therefore we have

$$\widehat{\mathcal{T}}(S) = \mathop{\mathbb{E}}_{\boldsymbol{y} \sim \{\pm 1\}^n} [\mathcal{T}(\boldsymbol{y}) \cdot \boldsymbol{y}_S] = \mathbb{E} [\mathcal{T}(\boldsymbol{v}_d) \cdot \mathbb{E} [\boldsymbol{y}_S]] = \mathbb{E} \left[\mathcal{T}(\boldsymbol{v}_d) \cdot \widehat{\mathcal{P}}_{\boldsymbol{v}_d}(S) \right]$$

By Fact 3.5.2, $\widehat{\mathcal{P}}_{\boldsymbol{v}_d}(S) \neq 0$ iff $S \in \mathcal{S}_{\boldsymbol{v}_d}$, which, due to $\ell \leq k$ and $\boldsymbol{v}_d$ being a k-clean leaf, is equivalent to all coordinates in S being fixed along this path. Hence $\widehat{\mathcal{P}}_{\boldsymbol{v}_d}(S) = \prod_{j \in S} \boldsymbol{v}_j^{(d)}$. □

Cleanup Process

We first analyze the cleanup process for a subspace.[7]

[5] This means for any $i \in \{0, 1 \ldots, \ell(v) - 1\}$, all nodes of depth i in $\mathcal{T}_v$ make the same query.

[6] This "leveled adaptive" condition is required just for convenience of proofs. In fact, one can show that the first few queries in $\mathcal{T}_v$ can be rearranged to make sure they are non-adaptive until we reach a k-clean node. See Lemma 3.5.7.

[7] The $k = 2$ case of Lemma 3.5.7 is essentially [BTW15, Proposition 3.5]. However there is a gap in their proof. For example, if the parity decision tree non-adaptively queries $x_1x_2x_3x_4, x_1x_5, x_2x_6$ in order, then their analysis fails.

Lemma 3.5.7 (Clean Subspace)**.** *Let $k \geq 2$ be an integer and $\mathcal{S}$ be a subspace of rank at most d. We construct a new subspace $\mathcal{S}'$ (initialized as $\mathcal{S}$) as follows: while $\mathcal{S}'$ is not k-clean, we continue to update $\mathcal{S}' \leftarrow \mathsf{Span}\left\langle \mathcal{S}' \cup \{\{i\}\}\right\rangle$ with some mess-witness i. Then $\mathsf{rank}(\mathcal{S}') \leq d \cdot k$ and any update choice of mess-witnesses will result in the same final subspace $\mathcal{S}'$.*

Proof. Assume $\mathcal{S}$ is a subspace of $\mathbb{F}_2^n$. Then first note that the number of updates is finite, since we can update for at most n times.

Next we show that the number of updates and the final $\mathcal{S}'$ does not depend on the choice of mess-witnesses. We do so by an exchange argument. Let $i_1, \ldots, i_r$ and $i'_1, \ldots, i'_{r'}$ be two rounds of execution using different mess-witnesses. Then there exists some $t < \min\{r, r'\}$ such that $i_j = i'_j$ for all $j \leq t$, but $i_{t+1} \neq i'_{t+1}$. Let $\mathcal{S}_t = \mathsf{Span}\left\langle \mathcal{S} \cup \{\{i_1\}, \ldots, \{i_t\}\}\right\rangle$. Then there exist $S \ni i_{t+1}$ and $S' \ni i'_{t+1}$ (possibly $S = S'$) such that $S, S' \in \mathcal{S}_t$ but $\{i_{t+1}\}, \{i'_{t+1}\} \notin \mathcal{S}_t$. Since the final subspace is k-clean, we know there exists some $T \geq t$ such that

$$\{i_{t+1}\} \notin \mathsf{Span}\left\langle \mathcal{S} \cup \{\{i'_1\}, \ldots, \{i'_T\}\}\right\rangle \quad \text{but} \quad \{i_{t+1}\} \in \mathsf{Span}\left\langle \mathcal{S} \cup \left\{\{i'_1\}, \ldots, \left\{i'_{T+1}\right\}\right\}\right\rangle,$$

which means $\left\{i'_{T+1}, i_{t+1}\right\} \in \mathsf{Span}\left\langle \mathcal{S} \cup \{\{i'_1\}, \ldots, \{i'_T\}\}\right\rangle$. Hence we can safely replace i'_{T+1} with i_{t+1}, and then swap i_{t+1} with i'_{t+1}. We can perform this process as long as $(i_1, \ldots, i_r) \neq (i'_1, \ldots, i'_{r'})$, which means $r = r'$ and the final $\mathcal{S}'$ is always the same.

For any subspace $\mathcal{H}$, we define $\mathsf{rank}_1(\mathcal{H}) = |\{i \mid \{i\} \in \mathcal{H}\}|$ and thus $\mathsf{rank}(\mathcal{H}) - \mathsf{rank}_1(\mathcal{H}) \geq 0$. Now we analyze the following particular way to construct $\mathcal{S}'$: We initialize $\mathcal{S}'$ as $\mathcal{S}$. While $\mathcal{S}'$ is not k-clean, we find a minimal $S = \{i_1, \ldots, i_s\} \in \mathcal{S}'$ such that i_1 is a mess-witness; then we update $\mathcal{S}' \leftarrow \mathsf{Span}\left\langle \mathcal{S}' \cup \{\{i_1\}, \ldots, \{i_{s-1}\}\}\right\rangle$. Note that before the update, $1 < s \leq k$ and $\{i_j\} \notin \mathcal{S}'$ holds for each $j \in [s]$, since S is minimal and $\mathcal{S}'$ is not k-clean. Thus after the update, $\mathsf{rank}(\mathcal{S}')$ grows by $s - 1 \leq k - 1$ and $\mathsf{rank}_1(\mathcal{S}')$ grows by s, which means $\mathsf{rank}(\mathcal{S}') - \mathsf{rank}_1(\mathcal{S}')$ shrinks by 1. Hence we have at most $\mathsf{rank}(\mathcal{S}) - \mathsf{rank}_1(\mathcal{S}) \leq d$ updates before $\mathcal{S}'$ is k-clean; and the final $\mathcal{S}'$ has rank at most $\mathsf{rank}(\mathcal{S}) + (k-1) \cdot d \leq d \cdot k$. □

We now show how to convert an arbitrary parity decision tree into a k-clean parity decision tree which still has a small depth and fixes a small number of variables along each path. The latter quantity is in fact bounded by the depth as shown in Fact 3.5.8.

Fact 3.5.8. *Let $\mathcal{T}$ be a depth-d parity decision tree. Let $v_0, \ldots, v_{d'}$ be any root-to-leaf path. Then we have $\sum_{i=0}^{d'-1} |J(v_i)| \leq d'$.*

Proof. Observe that $\sum_{i=0}^{d'-1} |J(v_i)| = \left|\left\{i \mid \{i\} \in \mathsf{Span}\left\langle Q_{v_0}, \ldots, Q_{v_{d'-1}}\right\rangle\right\}\right| \leq d'$. □

Corollary 3.5.9. *Let $\mathcal{T}$ be a depth-D k-clean parity decision tree. Let $v_0, \ldots, v_{D'}$ be any root-to-leaf path where at most d of $v_0, \ldots, v_{D'-1}$ are k-clean. Then $\sum_{i:|J(v_{i-1})|>1} |J(v_i)| \leq 2d$.*

Proof. By Fact 3.5.8 we have $\sum_{i=0}^{D'-1} |J(v_i)| - 1 \leq 0$. Since any v_i with $J(v_i) = \emptyset$ is not cleaning and therefore must be k-clean. Thus

$$\sum_{i:|J(v_i)|>1} |J(v_i)| - 1 \leq |\{i : J(v_i) = \emptyset\}| \leq d.$$

For $|J(v_i)| > 1$, we have $|J(v_i)| - 1 \geq |J(v_i)|/2$ and thus $\sum_{i:|J(v_i)|>1} |J(v_i)| \leq 2d$. □

Lemma 3.5.10 (Clean Parity Decision Tree)**.** *Let $k \geq 2$ be an integer. Let $\mathcal{T}$ be an arbitrary depth-d parity decision tree. Then there exists a k-clean parity decision tree $\mathcal{T}'$ of depth at most $d \cdot k$ equivalent to $\mathcal{T}$. Moreover, any root-to-leaf path in $\mathcal{T}'$ has at most d nodes that are k-clean.*

Proof. We build $\mathcal{T}'$ by the following recursive algorithm. An example of the algorithm is provided in Figure 3.1.

Algorithm 1: Clean parity decision tree: build $\mathcal{T}'$ from $\mathcal{T}$

Input: an arbitrary depth-d parity decision tree $\mathcal{T}$
Output: a parity decision tree $\mathcal{T}'$ with desired properties

$r \leftarrow$ root of $\mathcal{T}$
Initialize the root of $\mathcal{T}'$ as r'
`Build`$(r, r', 1)$
Procedure `Build`(v, v', ℓ)
  `/* (`v, v'`) are current nodes on (`$\mathcal{T}, \mathcal{T}'$`); `$\ell$` is the recursion depth. */`
  if *v is a leaf* **then** Label v' with the label of v
  else
    $(v_-, v_+) \leftarrow$ the left and right child of v
    if $\widehat{\mathcal{P}_{v'}}(Q_v) = -1$ **then** `Build`$(v_-, v', \ell + 1)$
    else if $\widehat{\mathcal{P}_{v'}}(Q_v) = +1$ **then** `Build`$(v_+, v', \ell + 1)$
    else `/*` $\widehat{\mathcal{P}_{v'}}(Q_v) = 0$ `due to Fact 3.5.2 */`
      $Q_{v'} \leftarrow Q_v$
      $(v'_-, v'_+) \leftarrow$ the left and right child of v'
      Initialize $O \leftarrow \emptyset$
      while Span $\langle \mathcal{S}_{v'} \cup \{Q_{v'}\} \cup O \rangle$ *is not k-clean* **do**
        Update $O \leftarrow O \cup \{\{i\}\}$, where i is a *mess-witness*
      end
      $\mathcal{T}'$ non-adaptively queries every set (which is a singleton) in O under v' in arbitrary order
      foreach *leaf $\widehat{v}$ under v'_-* **do** `Build`$(v_-, \widehat{v}, \ell + 1)$
      foreach *leaf $\widehat{v}$ under v'_+* **do** `Build`$(v_+, \widehat{v}, \ell + 1)$
    end
  end

Then the correctness of Algorithm 1 follows from the following claims.

Claim 3.5.11. *For any internal node $v' \in \mathcal{T}'$, $Q_{v'}$ is not implied by its ancestors' queries.*

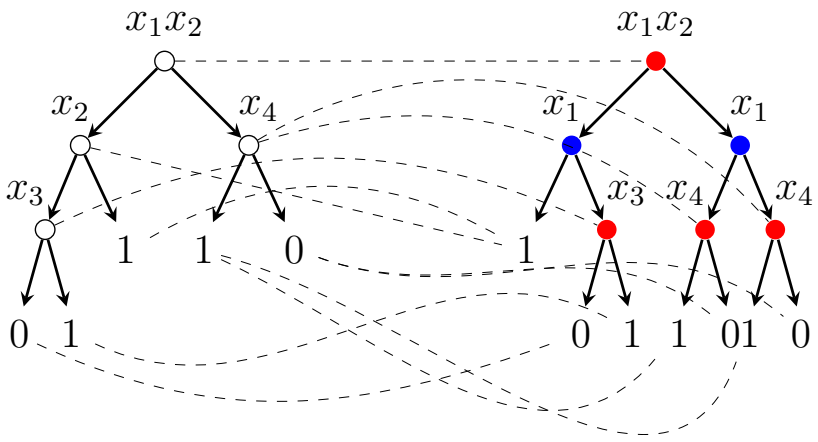


Figure 3.1: An example of the cleanup process with $k = 2$ where the LHS is $\mathcal{T}$ and the RHS is $\mathcal{T}'$. All the left (resp., right) outgoing edges are labeled with -1 (resp., $+1$). Red nodes and leaves are k-clean, and blue nodes are cleaning (i.e., non-adaptive queries). Nodes connected with dashed curves are invoked by Build.

Proof. We apply Fact 3.5.2. This is equivalent to $Q_{v'} \notin \mathcal{S}_{v'}$, which follows from the conditions in Line 8/9/13. □

Claim 3.5.12. *The depth of $\mathcal{T}'$ is at most $d \cdot k$.*

Proof. Let $v_0, \ldots, v_{d'}$ be any root-to-leaf path of $\mathcal{T}$ and let $\mathcal{P}'$ be its corresponding path in $\mathcal{T}'$. The construction process of $\mathcal{P}'$ corresponds to the cleanup process for $\mathsf{Span}\left\langle Q_{v_0}, \ldots, Q_{v_{d'-1}}\right\rangle$ in Lemma 3.5.7; hence the depth of $\mathcal{T}'$ equals $\mathsf{rank}(\mathcal{S}') \le d' \cdot k \le d \cdot k$ where $\mathcal{S}'$ is the k-clean subspace produced by applying Lemma 3.5.7. □

Claim 3.5.13. *$\mathcal{T} \equiv \mathcal{T}'$ and any root-to-leaf path in $\mathcal{T}'$ has at most d k-clean nodes.*

Proof. This is evident, as $\mathcal{T}'$ only refines $\mathcal{T}$ by inserting cleaning nodes. □

Claim 3.5.14. *Whenever we call* `Build(·, v′, ·)`*, v' is k-clean.*

Proof. We prove by induction on ℓ. The base case Line 3 is obvious. For Line 8/9, we recurse on the same v', which is k-clean by induction. For Line 17/18, note that $\mathcal{S}_{\widehat{v}} = \mathsf{Span}\left\langle \mathcal{S}_{v'} \cup \{Q_{v'}\} \cup O\right\rangle$; hence from the condition in Line 13, it is k-clean. □

Claim 3.5.15. *Nodes created in Line 16 are cleaning.*

Proof. Let $o = |O|$ and let $i_1, i_2, \ldots, i_o$ be the query order. For any $j \in [o]$, let v'_j be any one of the nodes created for i_j, then $\mathcal{S}_{v'_j} = \mathsf{Span}\left\langle \mathcal{S}_{v'} \cup \{Q_{v'}\} \cup \{\{i_1\}, \ldots, \{i_{j-1}\}\}\right\rangle$, which is not k-clean by Line 13; hence v'_j is cleaning by the condition in Line 13. □

This completes the proof of Lemma 3.5.10. □

3.6 Fourier Growth Bounds for Parity Decision Trees

Our goal in this section is to prove Theorem 3.2.1 with detailed bounds provided.

Level-1 Bound

We first prove the concentration result for level-1. We start with the following simple bound for general parity decision trees.

Lemma 3.6.1. *Let $\mathcal{T}\colon \{\pm 1\}^n \to \{0,1\}$ be a depth-D parity decision tree. Let $v_0, \dots, v_{D'}$ be any root-to-leaf path. Define $v^{(0)}, \dots, v^{(D')} \in \{-1,0,+1\}^n$ by setting $v_j^{(i)} = \widehat{\mathcal{P}_{v_i}}(j)$ for each $0 \le i \le D'$ and $j \in [n]$. Then for any $a_1, \dots, a_n \in \{-1,0,1\}$, we have $\left|\sum_{j=1}^n a_j \cdot v_j^{(D')}\right| \le D' \le D$.*

Proof. Note that the set of non-zero coordinates in $v^{(D')}$ is exactly $\bigcup_{i=0}^{D'-1} J(v_i)$. Hence by Fact 3.5.8, we have

$$\left|\sum_{j=1}^n a_j \cdot v_j^{(D')}\right| \le \sum_{j=1}^n \left|v_j^{(D')}\right| = \sum_{i=0}^{D'-1} |J(v_i)| \le D' \le D. \qquad \square$$

Now we give an improved bound for k-clean parity decision trees. To do so, we need one more notation which will be crucial in our analysis.

Notation 3.6.2. *Let $\mathcal{T}$ be a k-clean parity decision tree. For any node v, we define $C(v)$ as the nearest ancestor of v (including itself) that is k-clean.*

Lemma 3.6.3. *There exists a universal constant $\kappa \ge 1$ such that the following holds. Let $\mathcal{T}\colon \{\pm 1\}^n \to \{0,1\}$ be a depth-D $2k$-clean parity decision tree where $k \ge 1$ and any root-to-leaf path has at most d nodes that are $2k$-clean.*

Let $\boldsymbol{v}_0, \dots, \boldsymbol{v}_{\boldsymbol{D'}}$ be a random root-to-leaf path. Define $\boldsymbol{v}^{(0)}, \dots, \boldsymbol{v}^{(\boldsymbol{D'})} \in \{-1,0,+1\}^n$ by setting $\boldsymbol{v}_j^{(i)} = \widehat{\mathcal{P}_{\boldsymbol{v}_i}}(j)$ for each $0 \le i \le \boldsymbol{D'}$ and $j \in [n]$. Then for any $a_1, \dots, a_n \in \{-1,0,1\}$ and any $\varepsilon \le 1/2$, we have $\mathbf{Pr}\left[\left|\sum_{j=1}^n a_j \cdot \boldsymbol{v}_j^{(\boldsymbol{D'})}\right| \ge R(D,d,k,\varepsilon)\right] \le \varepsilon$, where

$$R(D,d,k,\varepsilon) = \kappa \cdot \sqrt{\left(D + dk\left(\frac{1}{\varepsilon}\right)^{\frac{1}{k}}\right)\log\left(\frac{1}{\varepsilon}\right)}.$$

In the proof of Lemma 3.6.3 we will use the following simple claim.

Fact 3.6.4. *Let $p_1, \dots, p_n$ be a sub-probability distribution, i.e., $p_i \ge 0$ and $\sum_{i=1}^n p_i \le 1$. Let $a_1, \dots, a_n \in \mathbb{R}$. Then for any $k \in \mathbb{N}$, we have $\sum_{i=1}^n p_i a_i^{2k} \ge \left(\sum_{i=1}^n p_i a_i^2\right)^k$.*

Proof. We add $p_{n+1} = 1-(\sum_{i=1}^{n} p_i)$ and $a_{n+1} = 0$ so p is a probability distribution. Then the claim follows from $\mathbb{E}[\boldsymbol{X}^k] \geq \mathbb{E}[\boldsymbol{X}]^k$, where random variable $\boldsymbol{X}$ gets value a_i^2 with probability p_i. □

Proof of Lemma 3.6.3. Extend $\boldsymbol{v}^{(\boldsymbol{D}'+1)} = \cdots = \boldsymbol{v}^{(D)}$ to equal $\boldsymbol{v}^{(\boldsymbol{D}')}$. For each $0 \leq i \leq D$, let $\boldsymbol{X}^{(i)} = \sum_{j=1}^{n} a_j \cdot \boldsymbol{v}_j^{(i)}$. We define $\boldsymbol{\delta}^{(i)} = 0$ for $\boldsymbol{D}' < i \leq D$. For $1 \leq i \leq \boldsymbol{D}'$, we let

$$\boldsymbol{\delta}^{(i)} = \boldsymbol{X}^{(i)} - \boldsymbol{X}^{(i-1)} = \sum_{j=1}^{n} a_j \cdot \left(\boldsymbol{v}_j^{(i)} - \boldsymbol{v}_j^{(i-1)}\right) = \sum_{j \in J(\boldsymbol{v}_{i-1})} a_j \cdot \boldsymbol{v}_j^{(i)},$$

where $J(\boldsymbol{v}_{i-1})$ depends only on $C(\boldsymbol{v}_{i-1})$ since $\mathcal{T}_{C(\boldsymbol{v}_{i-1})}$ performs non-adaptive queries before (and possibly even after) reaching $\boldsymbol{v}_i$. Note that for the two possible outcomes of querying $Q_{\boldsymbol{v}_i}$, $\boldsymbol{v}_j^{(i)}$ is fixed to ± 1 respectively for each $j \in J(\boldsymbol{v}_{i-1})$. Thus $\boldsymbol{\delta}^{(i)} = \boldsymbol{\Delta}^{(i)} \cdot \boldsymbol{z}^{(i)}$ where $\boldsymbol{\Delta}^{(i)}$ is a fixed value given $\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(i-1)}$ and $\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(\boldsymbol{D}')}$ are independent unbiased coins in $\{\pm 1\}$.

Since $C(\boldsymbol{v}_{i-1})$ is $2k$-clean, the collection of random variables $\left\{\boldsymbol{v}_j^{(i)} \,\middle|\, j \in J(\boldsymbol{v}_{i-1})\right\}$ is $2k$-wise independent conditioning on $C(\boldsymbol{v}_{i-1})$. Note that $\boldsymbol{\delta}^{(i)}$ is a linear function and

$$\mathbb{E}\left[\boldsymbol{\delta}^{(i)} \,\middle|\, C(\boldsymbol{v}_{i-1})\right] = 0 \quad \text{and} \quad \mathbb{E}\left[\left(\boldsymbol{\delta}^{(i)}\right)^2 \,\middle|\, C(\boldsymbol{v}_{i-1})\right] = \sum_{j \in J(\boldsymbol{v}_{i-1})} a_j^2 \leq \left|J(\boldsymbol{v}_{i-1})\right|.$$

By the first bound in Lemma 2.0.9, we have

$$\mathbb{E}\left[\left(\boldsymbol{\delta}^{(i)}\right)^{2k} \,\middle|\, C(\boldsymbol{v}_{i-1})\right] \leq (2k-1)^k \cdot \left|J(\boldsymbol{v}_{i-1})\right|^k. \tag{3.8}$$

Meanwhile $\left|\boldsymbol{\delta}^{(i)}\right| \leq \left|J(\boldsymbol{v}_{i-1})\right|$. Our first goal is to bound $\mathbf{Pr}\left[\sum_{i=1}^{D}\left(\boldsymbol{\delta}^{(i)}\right)^2 > D + 2\alpha^2 d\right]$. Observe that whenever the event $\sum_{i=1}^{D}\left(\boldsymbol{\delta}^{(i)}\right)^2 > D + 2\alpha^2 d$ happens, it must be the case that $\sum_{i:|J(\boldsymbol{v}_{i-1})|>1}\left(\boldsymbol{\delta}^{(i)}\right)^2 > 2\alpha^2 d$. Thus,

$$\begin{aligned}
\mathbf{Pr}\left[\sum_{i=1}^{D}\left(\boldsymbol{\delta}^{(i)}\right)^2 > D + 2\alpha^2 d\right] &\leq \mathbf{Pr}\left[\sum_{i:|J(\boldsymbol{v}_{i-1})|>1}\left(\boldsymbol{\delta}^{(i)}\right)^2 > 2\alpha^2 d\right] \\
&= \mathbf{Pr}\left[\sum_{i:|J(\boldsymbol{v}_{i-1})|>1} \frac{\left|J(\boldsymbol{v}_{i-1})\right|}{2d} \cdot \frac{\left(\boldsymbol{\delta}^{(i)}\right)^2}{\left|J(\boldsymbol{v}_{i-1})\right|} > \alpha^2\right] \\
&\leq \mathbf{Pr}\left[\sum_{i:|J(\boldsymbol{v}_{i-1})|>1} \frac{\left|J(\boldsymbol{v}_{i-1})\right|}{2d} \cdot \frac{\left(\boldsymbol{\delta}^{(i)}\right)^{2k}}{\left|J(\boldsymbol{v}_{i-1})\right|^k} > \alpha^{2k}\right] \\
&\qquad \text{(by Fact 3.6.4 and Corollary 3.5.9)}
\end{aligned}$$

$$= \mathbf{Pr}\left[\sum_{i:|J(\boldsymbol{v}_{i-1})|>1} \frac{\left(\boldsymbol{\delta}^{(i)}\right)^{2k}}{|J(\boldsymbol{v}_{i-1})|^{k-1}} > 2d \cdot \alpha^{2k}\right]$$
$$\leq \mathbb{E}\left[\sum_{i:|J(\boldsymbol{v}_{i-1})|>1} \frac{\left(\boldsymbol{\delta}^{(i)}\right)^{2k}}{|J(\boldsymbol{v}_{i-1})|^{k-1}}\right] \cdot \frac{1}{2d \cdot \alpha^{2k}}.$$
(by Markov's inequality)

On the other hand,

$$\mathbb{E}\left[\sum_{i:|J(\boldsymbol{v}_{i-1})|>1} \frac{\left(\boldsymbol{\delta}^{(i)}\right)^{2k}}{|J(\boldsymbol{v}_{i-1})|^{k-1}}\right] = \sum_{i=1}^{D} \mathop{\mathbb{E}}_{C(\boldsymbol{v}_{i-1})}\left[\frac{1_{|J(\boldsymbol{v}_{i-1})|>1}}{|J(\boldsymbol{v}_{i-1})|^{k-1}} \cdot \mathbb{E}\left[\left(\boldsymbol{\delta}^{(i)}\right)^{2k} \middle| C(\boldsymbol{v}_{i-1})\right]\right]$$
$$\leq \sum_{i=1}^{D} \mathop{\mathbb{E}}_{C(\boldsymbol{v}_{i-1})}\left[1_{|J(\boldsymbol{v}_{i-1})|>1} \cdot (2k-1)^k \cdot |J(\boldsymbol{v}_{i-1})|\right] \quad \text{(by (3.8))}$$
$$= (2k-1)^k \cdot \mathbb{E}\left[\sum_{i:|J(\boldsymbol{v}_{i-1}|>1} |J(\boldsymbol{v}_{i-1})|\right]$$
$$\leq (2k-1)^k \cdot 2d. \quad \text{(by Corollary 3.5.9)}$$

Overall, we have

$$\mathbf{Pr}\left[\sum_{i=1}^{D} \left(\boldsymbol{\delta}^{(i)}\right)^2 > D + 2\alpha^2 d\right] \leq \frac{(2k-1)^k}{\alpha^{2k}}.$$

Then by Lemma 3.4.5 with $m = 1$, we have

$$\mathbf{Pr}\left[\left|\boldsymbol{X}^{(D)}\right| = \left|\sum_{j=1}^{n} a_j \cdot \boldsymbol{v}_j^{(D)}\right| \geq \beta\sqrt{2 \cdot (D + 2\alpha^2 d)}\right] \leq 2 \cdot e^{-\beta^2/2} + \frac{(2k-1)^k}{\alpha^{2k}}.$$

The desired bound follows from setting

$$\alpha = \left(\frac{2}{\varepsilon}\right)^{\frac{1}{2k}} \sqrt{2k-1}, \quad \text{and} \quad \beta = \Theta\left(\sqrt{\log\left(\frac{1}{\varepsilon}\right)}\right). \qquad \square$$

Now we prove the complete level-1 bound for parity decision trees.

Theorem 3.6.5. *Let $\mathcal{T}\colon \{\pm 1\}^n \to \{0,1\}$ be a depth-d parity decision tree. Let $p = \mathbf{Pr}[\mathcal{T}(\boldsymbol{x}) = 1] \in [2^{-d}, 1/2]$.*[8] *Then we have*

$$\sum_{j=1}^{n} \left|\widehat{\mathcal{T}}(j)\right| \leq p \cdot \min\left\{d, O\left(\sqrt{d} \cdot \log\left(\frac{1}{p}\right)\right)\right\} = O\left(\sqrt{d}\right).$$

Proof. For any $i \in [n]$, let $a_i = \mathsf{sgn}\left(\widehat{\mathcal{T}}(i)\right)$. Now we prove the two bounds separately.

[8] If $p < 2^{-d}$, then $p = 0$ and $\mathcal{T} \equiv 0$. If $p > 1/2$, we can consider $\widetilde{\mathcal{T}} = 1 - \mathcal{T}$ by symmetry.

The First Bound. Let $\boldsymbol{v}_0, \ldots, \boldsymbol{v}_{\boldsymbol{d}'}$ be a random root-to-leaf path in $\mathcal{T}$. Then define $\boldsymbol{v}^{(0)}, \ldots, \boldsymbol{v}^{(\boldsymbol{d}')} \in \{-1, 0, +1\}^n$ by setting $\boldsymbol{v}_j^{(i)} = \widehat{\mathcal{P}_{\boldsymbol{v}_i}}(j)$ for each $0 \le i \le \boldsymbol{d}'$ and $j \in [n]$. Since $\mathcal{T}$ is 1-clean in itself, by Lemma 3.5.6 we have

$$\sum_{j=1}^{n} \left|\widehat{\mathcal{T}}(j)\right| = \sum_{j=1}^{n} a_i \cdot \widehat{\mathcal{T}}(j) = \mathop{\mathbb{E}}_{\boldsymbol{v}_0,\ldots,\boldsymbol{v}_{\boldsymbol{d}'}} \left[\mathcal{T}(\boldsymbol{v}_{\boldsymbol{d}'}) \cdot \sum_{j=1}^{n} a_j \cdot \boldsymbol{v}_j^{(\boldsymbol{d}')}\right] \le \mathop{\mathbb{E}}_{\boldsymbol{v}_0,\ldots,\boldsymbol{v}_{\boldsymbol{d}'}} \left[\mathcal{T}(\boldsymbol{v}_{\boldsymbol{d}'}) \cdot |\boldsymbol{V}|\right], \quad (3.9)$$

where $\boldsymbol{V} = \sum_{j=1}^{n} a_j \cdot \boldsymbol{v}_j^{(\boldsymbol{d}')}$. Hence by Lemma 3.6.1, we have (3.9) $\le d \cdot \mathbb{E}\left[\mathcal{T}(\boldsymbol{v}_{\boldsymbol{d}'})\right] = p \cdot d$.

The Second Bound. By Lemma 3.5.10, we construct a $2k$-clean parity decision tree $\mathcal{T}'$ of depth $D \le 2d \cdot k$ equivalent to $\mathcal{T}$, where $k = \Theta(\log(1/p))$. Let $\boldsymbol{U} = \sum_{j=1}^{n} a_j \cdot \boldsymbol{u}_j^{(\boldsymbol{D}')}$. Then we have

$$\sum_{j=1}^{n} \left|\widehat{\mathcal{T}}(j)\right| = \sum_{j=1}^{n} \left|\widehat{\mathcal{T}'}(j)\right| = \mathop{\mathbb{E}}_{\boldsymbol{u}_0,\ldots,\boldsymbol{u}_{\boldsymbol{D}'}} \left[\mathcal{T}'(\boldsymbol{u}_{\boldsymbol{D}'}) \cdot \sum_{j=1}^{n} a_j \cdot \boldsymbol{u}_j^{(\boldsymbol{D}')}\right] \le \mathop{\mathbb{E}}_{\boldsymbol{u}_0,\ldots,\boldsymbol{u}_{\boldsymbol{D}'}} \left[\mathcal{T}'(\boldsymbol{u}_{\boldsymbol{D}'}) \cdot |\boldsymbol{U}|\right]. \quad (3.10)$$

Lemma 3.6.3 implies that for all $\varepsilon > 0$, $\mathbf{Pr}\left[|\boldsymbol{U}| \ge R(\varepsilon)\right] \le \varepsilon$ where

$$R(\varepsilon) = R(D, d, k, \varepsilon) = O\left(\sqrt{dk \cdot \left(\frac{1}{\varepsilon}\right)^{\frac{1}{k}} \cdot \log\left(\frac{1}{\varepsilon}\right)}\right).$$

For integer $i \ge 1$, let $I_i = \left[R\left(p/2^i\right), R\left(p/2^{i+1}\right)\right]$ and $I_0 = [0, R(p/2)]$ be intervals. Then for each $i \ge 1$, $\mathbf{Pr}\left[|\boldsymbol{U}| \in I_i\right] \le p/2^i$. We also know that $\mathbb{E}_{\boldsymbol{u}_0,\ldots,\boldsymbol{u}_{\boldsymbol{D}'}}\left[\mathcal{T}'(\boldsymbol{u}_{\boldsymbol{D}'})\right] \le p$. Thus,

$$\begin{aligned}
(3.10) &= \mathop{\mathbb{E}}_{\boldsymbol{u}_0,\ldots,\boldsymbol{u}_{\boldsymbol{D}'}} \left[\mathcal{T}'(\boldsymbol{u}_{\boldsymbol{D}'}) \cdot |\boldsymbol{U}| \cdot \sum_{i=0}^{+\infty} 1_{|\boldsymbol{U}| \in I_i}\right] \\
&\le R\left(\frac{p}{2}\right) \cdot \mathop{\mathbb{E}}_{\boldsymbol{u}_0,\ldots,\boldsymbol{u}_{\boldsymbol{D}'}} \left[\mathcal{T}'(\boldsymbol{u}_{\boldsymbol{D}'})\right] + \sum_{i=1}^{+\infty} R\left(\frac{p}{2^{i+1}}\right) \cdot \mathop{\mathbb{E}}_{\boldsymbol{u}_0,\ldots,\boldsymbol{u}_{\boldsymbol{D}'}} \left[1_{|\boldsymbol{U}| \in I_i}\right] \\
&\le \sum_{i=0}^{+\infty} R\left(\frac{p}{2^{i+1}}\right) \cdot \frac{p}{2^i} \\
&= \sum_{i=0}^{+\infty} O\left(p \cdot \sqrt{dk \cdot \left(\frac{2^{i+1}}{p}\right)^{\frac{1}{k}} \cdot \left(\log\left(\frac{1}{p}\right) + i + 1\right)}\right) \cdot \frac{1}{2^i} \\
&= O\left(p \cdot \sqrt{dk \cdot \log\left(\frac{1}{p}\right)}\right) = O\left(p \cdot \sqrt{d} \cdot \log\left(\frac{1}{p}\right)\right).
\end{aligned}$$

□

Level-ℓ Bound

Now we turn to the general levels.

Lemma 3.6.6. *There exists a universal constant $\tau \geq 1$ such that the following holds. Let $\ell \geq 1$ be an integer. Let $\mathcal{T}\colon \{\pm 1\}^n \to \{0,1\}$ be a depth-D $2k$-clean parity decision tree where $k \geq 4 \cdot \ell$ and $n \geq \max\{\tau, k, D\}$ and any root-to-leaf path has at most d nodes that are $2k$-clean.*

Let $\boldsymbol{v}_0, \ldots, \boldsymbol{v}_{\boldsymbol{D}'}$ be a random root-to-leaf path. Define $\boldsymbol{v}^{(0)}, \ldots, \boldsymbol{v}^{(\boldsymbol{D}')} \in \{-1, 0, +1\}^n$ by setting $\boldsymbol{v}_j^{(i)} = \widehat{\mathcal{P}_{\boldsymbol{v}_i}}(j)$ for each $0 \leq i \leq \boldsymbol{D}'$ and $j \in [n]$. Extend $\boldsymbol{v}^{(\boldsymbol{D}'+1)} = \cdots = \boldsymbol{v}^{(D)}$ to equal $\boldsymbol{v}^{(\boldsymbol{D}')}$. Then for any sequence $a_S \in \{-1, 0, 1\}, S \in \binom{[n]}{\ell}$, any $\varepsilon \leq 1/2$ and $t \in \{0, \ldots, \ell\}$, we have

$$\mathbf{Pr}\left[\exists t' \in \{0,\ldots,t\}, \exists T \in \binom{[n]}{\ell - t'}, \exists i \in [D], \left|\sum_{S \subseteq \overline{T}, |S| = t'} a_{S \cup T} \cdot \boldsymbol{v}_S^{(i)}\right| \geq M(D,d,k,\ell,t',\varepsilon)\right] \leq \varepsilon \cdot t,$$

where we recall that $\boldsymbol{v}_S^{(i)} = \prod_{j \in S} \boldsymbol{v}_j^{(i)}$ and where

$$M(D,d,k,\ell,t',\varepsilon) = \left(\tau \cdot (D + dk) \cdot \left(\frac{n^\ell}{\varepsilon}\right)^{\frac{6}{k}} \log\left(\frac{n^\ell}{\varepsilon}\right)\right)^{t'/2}.$$

Proof. We prove the bound by induction on $t = 0, 1, \ldots, \ell$ and show $\tau = 10^4$ suffices. The base case $t = 0$ is trivial, since for any fixed T and i, we always have $\left|a_T \cdot \boldsymbol{v}_\emptyset^{(i)}\right| \leq 1 = M(D,d,k,\ell,0,\varepsilon)$.

Now we focus on the case where $1 \leq t \leq \ell$. For each $0 \leq i \leq D$ and $T \in \binom{[n]}{\ell - t}$, let

$$\boldsymbol{X}_T^{(i)} = \sum_{S \subseteq \overline{T}, |S| = t} a_{S \cup T} \cdot \boldsymbol{v}_S^{(i)}.$$

For $1 \leq i \leq \boldsymbol{D}'$, we have

$$\begin{aligned}
\boldsymbol{X}_T^{(i)} - \boldsymbol{X}_T^{(i-1)} &= \sum_{S \subseteq \overline{T}, |S| = t, S \cap J(\boldsymbol{v}_{i-1}) \neq \emptyset} a_{S \cup T} \cdot \boldsymbol{v}_S^{(i)} \\
&= \sum_{r=1}^{t} \sum_{\substack{U \subseteq J(\boldsymbol{v}_{i-1}) \cap \overline{T}, \\ |U| = r}} \boldsymbol{v}_U^{(i)} \sum_{\substack{V \subseteq \overline{T \cup J(\boldsymbol{v}_{i-1})}, \\ |U| + |V| = t}} a_{T \cup U \cup V} \cdot \boldsymbol{v}_V^{(i)} \\
&= \sum_{r=1}^{t} \sum_{\substack{U \subseteq J(\boldsymbol{v}_{i-1}) \cap \overline{T}, \\ |U| = r}} \boldsymbol{v}_U^{(i)} \sum_{\substack{V \subseteq \overline{T \cup J(\boldsymbol{v}_{i-1})}, \\ |U| + |V| = t}} a_{T \cup U \cup V} \cdot \boldsymbol{v}_V^{(i-1)} \\
&\qquad \text{(since } \boldsymbol{v}_j^{(i)} = \boldsymbol{v}_j^{(i-1)} \text{ for all } j \notin J(\boldsymbol{v}_{i-1})\text{)}
\end{aligned}$$

$$= \sum_{r=1}^{t} \underbrace{\sum_{\substack{U \subseteq J(\boldsymbol{v}_{i-1}) \cap \overline{T}, \\ |U|=r}} \boldsymbol{v}_U^{(i)} \sum_{\substack{V \subseteq \overline{T \cup U}, \\ |U|+|V|=t}} a_{T \cup U \cup V} \cdot \boldsymbol{v}_V^{(i-1)}}_{\boldsymbol{A}(T,r,i)}.$$

(since $\boldsymbol{v}_j^{(i-1)} = 0$ for all $j \in J(\boldsymbol{v}_{i-1})$)

Observe that conditioning on $\boldsymbol{v}_{i-1}$,

- if r is an even number, then $\boldsymbol{A}(T, r, i)$ is a fixed value independent of $\boldsymbol{v}^{(i)}$;
- if r is an odd number, then $\boldsymbol{A}(T, r, i)$ is an unbiased coin with magnitude independent of $\boldsymbol{v}^{(i)}$.

Therefore, trying to apply Lemma 3.4.5, we write $\boldsymbol{X}_T^{(i)} - \boldsymbol{X}_T^{(i-1)} = \boldsymbol{\mu}_T^{(i)} + \boldsymbol{\Delta}_T^{(i)} \cdot \boldsymbol{z}_T^{(i)}$, where $\boldsymbol{z}_T^{(1)}, \ldots, \boldsymbol{z}_T^{(D)}$ are independent unbiased coins in $\{\pm 1\}$ and $\boldsymbol{\mu}_T^{(i)} = \boldsymbol{\Delta}_T^{(i)} = 0$ for $\boldsymbol{D'} < i \le D$ and

$$\boldsymbol{\mu}_T^{(i)} = \sum_{\substack{r=2, \\ \text{even}}}^{t} \boldsymbol{A}(T, r, i) \quad \text{and} \quad \boldsymbol{\Delta}_T^{(i)} = \left| \sum_{\substack{r=1, \\ \text{odd}}}^{t} \boldsymbol{A}(T, r, i) \right| \quad \text{for } 1 \le i \le \boldsymbol{D'}. \tag{3.11}$$

The First Bound on $\boldsymbol{A}(T, r, i)$. Let $\mathcal{E}_1$ be the following event:

$$\mathcal{E}_1 = \text{“ } \exists \widehat{t} \in \{0, \ldots, t-1\}, \exists T' \in \binom{[n]}{\ell - \widehat{t}}, \exists i' \in [D], \ \left| \boldsymbol{X}_{T'}^{(i')} \right| \ge M\left(D, k, \ell, \widehat{t}, \varepsilon\right) \text{ ”}.$$

By the induction hypothesis, we have

$$\mathbf{Pr}\left[\mathcal{E}_1\right] \le (t-1) \cdot \varepsilon. \tag{3.12}$$

We first derive a simple bound, that will be effective for small values of $|J(\boldsymbol{v}_{i-1})|$.

Claim 3.6.7. *When $\mathcal{E}_1$ does not happen, $|\boldsymbol{A}(T, r, i)| \le |J(\boldsymbol{v}_{i-1})|^r \cdot M(D, d, k, \ell, t-r, \varepsilon)$ holds for all $r \in [t], i \in [D], T \in \binom{[n]}{\ell - t}$.*

Proof. Since $\mathcal{E}_1$ does not happen, by union bound we have

$$|\boldsymbol{A}(T, r, i)| = \left| \sum_{\substack{U \subseteq J(\boldsymbol{v}_{i-1}) \cap \overline{T}, \\ |U|=r}} \boldsymbol{v}_U^{(i)} \sum_{\substack{V \subseteq \overline{T \cup U}, \\ |U|+|V|=t}} a_{T \cup U \cup V} \cdot \boldsymbol{v}_V^{(i-1)} \right| \le |J(\boldsymbol{v}_{i-1})|^r \max_{U \subseteq \overline{T}, |U|=r} \left| \boldsymbol{X}_{T \cup U}^{(i-1)} \right|$$
$$\le |J(\boldsymbol{v}_{i-1})|^r \cdot M(D, d, k, \ell, t-r, \varepsilon). \qquad \square$$

The Second Bound on $\boldsymbol{A}(T,r,i)$. The second bound requires a more refined decomposition on $\boldsymbol{A}(T,r,i)$.

Assume that $\boldsymbol{c}(i-1)$ is the index of $C(\boldsymbol{v}_{i-1})$ in $\boldsymbol{v}_0,\ldots,\boldsymbol{v}_{\boldsymbol{D}'}$, i.e., $\boldsymbol{v}_{\boldsymbol{c}(i-1)} = C(\boldsymbol{v}_{i-1})$. This means that $\boldsymbol{v}_{\boldsymbol{c}(i-1)}$ is the closest ancestor to $\boldsymbol{v}_{i-1}$ that is $2k$-clean. Then define

$$L(\boldsymbol{v}_{i-1}) = \bigcup_{\boldsymbol{c}(i-1)\le i'<i-1} J(\boldsymbol{v}_{i'}).$$

The elements of $L(\boldsymbol{v}_{i-1})$ are precisely the coordinates fixed by the queries from $Q_{\boldsymbol{v}_{\boldsymbol{c}(i-1)}}$ to $Q_{\boldsymbol{v}_{i-1}}$, excluding the latter. Since $\mathcal{T}_{C(\boldsymbol{v}_{i-1})}$ makes non-adaptive queries before (and possibly even after) reaching $\boldsymbol{v}_i$, $L(\boldsymbol{v}_{i-1})$ and $J(\boldsymbol{v}_{i-1})$ depend only on $C(\boldsymbol{v}_{i-1})$ and i. We now expand $\boldsymbol{A}(T,r,i)$ by also grouping terms based on the number of coordinates in $L(\boldsymbol{v}_{i-1})$ as follows:

$$\begin{aligned}
\boldsymbol{A}(T,r,i) &= \sum_{\substack{U\subseteq J(\boldsymbol{v}_{i-1})\cap\overline{T},\\ |U|=r}} \boldsymbol{v}_U^{(i)} \sum_{\substack{V\subseteq\overline{T\cup U},\\ |U|+|V|=t}} a_{T\cup U\cup V}\cdot \boldsymbol{v}_V^{(i-1)} \\
&= \sum_{r'=0}^{t-r} \sum_{\substack{U\subseteq J(\boldsymbol{v}_{i-1})\cap\overline{T},\\ |U|=r}} \boldsymbol{v}_U^{(i)} \sum_{\substack{W\subseteq L(\boldsymbol{v}_{i-1})\cap\overline{T},\\ |W|=r'}} \boldsymbol{v}_W^{(i-1)} \sum_{\substack{W'\subseteq\overline{T\cup U\cup L(\boldsymbol{v}_{i-1})}\\ |W'|=t-r-r'}} a_{T\cup U\cup W\cup W'}\cdot \boldsymbol{v}_{W'}^{(i-1)} \\
&= \sum_{r'=0}^{t-r} \sum_{\substack{U\subseteq J(\boldsymbol{v}_{i-1})\cap\overline{T},\\ |U|=r}} \boldsymbol{v}_U^{(i)} \sum_{\substack{W\subseteq L(\boldsymbol{v}_{i-1})\cap\overline{T},\\ |W|=r'}} \boldsymbol{v}_W^{(i-1)} \sum_{\substack{W'\subseteq\overline{T\cup U\cup L(\boldsymbol{v}_{i-1})}\\ |W'|=t-r-r'}} a_{T\cup U\cup W\cup W'}\cdot \boldsymbol{v}_{W'}^{\boldsymbol{c}(i-1)} \\
&\qquad\qquad \text{(since } \boldsymbol{v}_j^{(i-1)} = \boldsymbol{v}_j^{\boldsymbol{c}(i-1)} \text{ for all } j\notin L(\boldsymbol{v}_{i-1})\text{)} \\
&= \sum_{r'=0}^{t-r} \sum_{\substack{U\subseteq J(\boldsymbol{v}_{i-1})\cap\overline{T},\\ |U|=r}} \boldsymbol{v}_U^{(i)} \sum_{\substack{W\subseteq L(\boldsymbol{v}_{i-1})\cap\overline{T},\\ |W|=r'}} \boldsymbol{v}_W^{(i-1)} \sum_{\substack{W'\subseteq\overline{T\cup U\cup W}\\ |W'|=t-r-r'}} a_{T\cup U\cup W\cup W'}\cdot \boldsymbol{v}_{W'}^{\boldsymbol{c}(i-1)} \\
&\qquad\qquad \text{(since } \boldsymbol{v}_j^{\boldsymbol{c}(i-1)} = 0 \text{ for all } j\in L(\boldsymbol{v}_{i-1})\text{)} \\
&= \sum_{r'=0}^{t-r} \underbrace{\sum_{\substack{U\subseteq J(\boldsymbol{v}_{i-1})\cap\overline{T},\\ |U|=r}} \boldsymbol{v}_U^{(i)} \sum_{\substack{W\subseteq L(\boldsymbol{v}_{i-1})\cap\overline{T},\\ |W|=r'}} \boldsymbol{v}_W^{(i-1)}\cdot \boldsymbol{X}^{\boldsymbol{c}(i-1)}_{T\cup U\cup W}}_{\boldsymbol{\Gamma}_T^{(i)}(r,r')}\cdot
\end{aligned}$$

Since $C(\boldsymbol{v}_{i-1})$ is $2k$-clean, by Fact 3.5.2, the collection of random variables

$$\left\{\boldsymbol{v}_j^{(i)} \,\middle|\, j\in J(\boldsymbol{v}_{i-1})\right\}\cup\left\{\boldsymbol{v}_j^{(i-1)} \,\middle|\, j\in L(\boldsymbol{v}_{i-1})\right\}$$

is $2k$-wise independent conditioning on $C(\boldsymbol{v}_{i-1})$. Note that $\boldsymbol{\Gamma}_T^{(i)}(r,r')$ is a polynomial of

degree at most $r + r' \le \ell < k$, that $\mathbb{E}\left[\boldsymbol{\Gamma}_T^{(i)}(r,r') \,\middle|\, C(\boldsymbol{v}_{i-1})\right] = 0$, and

$$\begin{aligned}
\boldsymbol{\sigma}_T^2(r,r',C(\boldsymbol{v}_{i-1}),i) &:= \mathbb{E}\left[\left(\boldsymbol{\Gamma}_T^{(i)}(r,r')\right)^2 \,\middle|\, C(\boldsymbol{v}_{i-1})\right] = \sum_{\substack{U\subseteq J(\boldsymbol{v}_{i-1})\cap\overline{T},\\ |U|=r}} \sum_{\substack{W\subseteq L(\boldsymbol{v}_{i-1})\cap\overline{T},\\ |W|=r'}} \left(\boldsymbol{X}_{T\cup U\cup W}^{\mathbf{c}(i-1)}\right)^2 \\
&\le \left(|J(\boldsymbol{v}_{i-1})|\right)^r \left(|L(\boldsymbol{v}_{i-1})|\right)^{r'} \left(\max_{|T'|=r+r'+\ell-t, i'\in[D]} \left|\boldsymbol{X}_{T'}^{(i')}\right|\right)^2 \\
&\le \left(|J(\boldsymbol{v}_{i-1})|\right)^r D^{r'} \left(\max_{|T'|=r+r'+\ell-t, i'\in[D]} \left|\boldsymbol{X}_{T'}^{(i')}\right|\right)^2. \\
&\qquad\qquad \text{(since } |L(\boldsymbol{v}_{i-1})| \le D \text{ by Fact 3.5.8)}
\end{aligned}$$

We also have the following claim, the proof of which follows from Lemma 2.0.9 applied to the low degree polynomial $\boldsymbol{\Gamma}_T^{(i)}$. The proof is deferred to the end of this section.

Claim 3.6.8. $\mathbf{Pr}\left[\mathcal{E}_2\right] \le \varepsilon/3$, *where* $\mathcal{E}_2$ *is the following event:*

$$\begin{aligned}
&\text{“}\exists T \in \binom{[n]}{\ell-t}, i, r, r', \\
&\qquad \left|\boldsymbol{\Gamma}_T^{(i)}(r,r')\right| \ge \left(100 \min\left\{k, \log\left(\tfrac{n^\ell}{\varepsilon}\right)\right\} \cdot \left(\tfrac{n^\ell}{\varepsilon}\right)^{\frac{6}{k}}\right)^{\frac{r+r'}{2}} \cdot \boldsymbol{\sigma}_T(r,r',C(\boldsymbol{v}_{i-1}),i)\text{”}.
\end{aligned}$$

On the other hand, when $\mathcal{E}_1 \vee \mathcal{E}_2$ does not happen, the following calculation holds for all $T \in \binom{[n]}{\ell-t}$, $i \in [\boldsymbol{D}']$, $r \in [t]$, $0 \le r' \le t - r$:

$$\begin{aligned}
&\left|\boldsymbol{\Gamma}_T^{(i)}(r,r')\right| \\
&\le M\left(D,k,\ell,t-r-r',\varepsilon\right) \cdot \sqrt{\left(100 \min\left\{k, \log\left(\tfrac{n^\ell}{\varepsilon}\right)\right\} \cdot \left(\tfrac{n^\ell}{\varepsilon}\right)^{\frac{6}{k}}\right)^{r+r'} \left(|J(\boldsymbol{v}_{i-1})|\right)^r \cdot D^{r'}} \\
&\le M\left(D,k,\ell,t-r-r',\varepsilon\right) \cdot \sqrt{\left(100 \cdot \left(\tfrac{n^\ell}{\varepsilon}\right)^{\frac{6}{k}}\right)^{r+r'} \left(|J(\boldsymbol{v}_{i-1})| \cdot k\right)^r \cdot \left(D \cdot \log\left(\tfrac{n^\ell}{\varepsilon}\right)\right)^{r'}} \\
&= \sqrt{\left(\tau(D+dk)\left(\tfrac{n^\ell}{\varepsilon}\right)^{\frac{6}{k}} \log\left(\tfrac{n^\ell}{\varepsilon}\right)\right)^{t-r-r'} \left(100\left(\tfrac{n^\ell}{\varepsilon}\right)^{\frac{6}{k}}\right)^{r+r'} \left(|J(\boldsymbol{v}_{i-1})| \cdot k\right)^r \left(D \cdot \log\left(\tfrac{n^\ell}{\varepsilon}\right)\right)^{r'}} \\
&\le \sqrt{\left(\tau(D+dk)\left(\tfrac{n^\ell}{\varepsilon}\right)^{\frac{6}{k}} \log\left(\tfrac{n^\ell}{\varepsilon}\right)\right)^{t} \left(\tfrac{100}{\tau}\right)^{r+r'} \left(\tfrac{|J(\boldsymbol{v}_{i-1})|}{d \cdot \log\left(n^\ell/\varepsilon\right)}\right)^r} \\
&\le \sqrt{\left(\tau(D+dk)\left(\tfrac{n^\ell}{\varepsilon}\right)^{\frac{6}{k}} \log\left(\tfrac{n^\ell}{\varepsilon}\right)\right)^{t} \left(\tfrac{200}{\tau}\right)^{r+r'} \left(\tfrac{|J(\boldsymbol{v}_{i-1})|}{2d}\right)^r \tfrac{1}{\log\left(n^\ell/\varepsilon\right)}} \\
&= M(D,d,k,\ell,t,\varepsilon) \cdot \sqrt{\left(\tfrac{200}{\tau}\right)^{r+r'} \left(\tfrac{|J(\boldsymbol{v}_{i-1})|}{2d}\right)^r \tfrac{1}{\log\left(n^\ell/\varepsilon\right)}}.
\end{aligned}$$

Hence we have a second bound on $\boldsymbol{A}(T,r,i)$.

Claim 3.6.9. *When $\mathcal{E}_1 \vee \mathcal{E}_2$ does not happen, the following holds for all $r \in [t], i \in [D], T \in \binom{[n]}{\ell-t}$:*

$$|\boldsymbol{A}(T,r,i)| \le \frac{M(D,d,k,\ell,t,\varepsilon)}{\sqrt{\log\left(n^\ell/\varepsilon\right)}} \cdot \sqrt{\left(\frac{800}{\tau}\right)^r \left(\frac{|J(\boldsymbol{v}_{i-1})|}{2d}\right)^r}.$$

Proof. Since $\mathcal{E}_1 \vee \mathcal{E}_2$ does not happen, by union bound and noticing $\tau \ge 800$ we have

$$\begin{aligned}|\boldsymbol{A}(T,r,i)| &\le \sum_{r'=0}^{t-r} \left|\boldsymbol{\Gamma}_T^{(i)}(r,r')\right| \le \frac{M(D,d,k,\ell,t,\varepsilon)}{\sqrt{\log\left(n^\ell/\varepsilon\right)}} \cdot \sqrt{\left(\frac{200}{\tau}\right)^r \left(\frac{|J(\boldsymbol{v}_{i-1})|}{2d}\right)^r} \cdot \sum_{r'=0}^{+\infty} \left(\frac{200}{\tau}\right)^{r'/2} \\ &\le \frac{M(D,d,k,\ell,t,\varepsilon)}{\sqrt{\log\left(n^\ell/\varepsilon\right)}} \cdot \sqrt{\left(\frac{800}{\tau}\right)^r \left(\frac{|J(\boldsymbol{v}_{i-1})|}{2d}\right)^r}.\end{aligned}$$

□

The Final Bound on $\boldsymbol{\mu}_T^{(i)}$ and $\boldsymbol{\delta}_T^{(i)}$. Combining Claim 3.6.7 and Claim 3.6.9, if $\mathcal{E}_1 \vee \mathcal{E}_2$ does not happen we have

$$\begin{aligned}|\boldsymbol{A}(T,r,i)| \le\ & M(D,d,k,\ell,t-r,\varepsilon) \\ & + \frac{M(D,d,k,\ell,t,\varepsilon)}{\sqrt{\log\left(n^\ell/\varepsilon\right)}} \cdot \sqrt{\left(\frac{800}{\tau}\right)^r \left(\frac{|J(\boldsymbol{v}_{i-1})|}{2d}\right)^r} \cdot 1_{|J(\boldsymbol{v}_{i-1})|>1}. \qquad (3.13)\end{aligned}$$

To see this, if $|J(\boldsymbol{v}_{i-1})| \le 1$, we use the bound from Claim 3.6.7 as the first term in (3.13). Otherwise $|J(\boldsymbol{v}_{i-1})| > 1$, in which case we use the bound from Claim 3.6.9 as the second term in (3.13).

By Corollary 3.5.9, we can now bound $\sum_{i=1}^D \left|\boldsymbol{\mu}_T^{(i)}\right|$ and $\sum_{i=1}^D \left|\boldsymbol{\Delta}_T^{(i)}\right|^2$ as Claim 3.6.10. Its proof is deferred to the end of this section.

Claim 3.6.10. *When $\mathcal{E}_1 \vee \mathcal{E}_2$ does not happen, $\sum_{i=1}^D \left|\boldsymbol{\mu}_T^{(i)}\right| \le R$ and $\sum_{i=1}^D \left|\boldsymbol{\Delta}_T^{(i)}\right|^2 \le R^2$ hold for all $T \in \binom{[n]}{\ell-t}$, where*

$$R = \frac{M(D,d,k,\ell,t,\varepsilon)}{5 \cdot \sqrt{\log\left(n^\ell/\varepsilon\right)}}. \qquad (3.14)$$

Complete Induction. Let $\beta = \sqrt{2 \cdot \log\left(n^\ell/\varepsilon\right)} \ge 1$ and observe that

$$\begin{aligned}R + \beta \cdot \sqrt{2} \cdot R &\le \beta \cdot 2\sqrt{2} \cdot R && \text{(due to } \beta \ge 1\text{)} \\ &= \frac{2\sqrt{2} \cdot \sqrt{2 \cdot \log\left(n^\ell/\varepsilon\right)}}{5 \cdot \sqrt{\log\left(n^\ell/\varepsilon\right)}} \cdot M(D,d,k,\ell,t,\varepsilon) && \text{(due to (3.14))} \\ &\le M(D,d,k,\ell,t,\varepsilon).\end{aligned}$$

Then we have

$$\begin{aligned}
&\mathbf{Pr}\left[\exists t' \in \{0,\dots,t\}, \exists T' \in \binom{[n]}{\ell - t'}, \exists i \in [D],\ \left|\boldsymbol{X}_T^{(i)}\right| \geq M\left(D,d,k,\ell,t',\varepsilon\right)\right] \\
&= \mathbf{Pr}\left[\mathcal{E}_1 \bigvee \left(\exists T \in \binom{[n]}{\ell - t}, \exists i \in [D],\ \left|\boldsymbol{X}_T^{(i)}\right| \geq M\left(D,d,k,\ell,t,\varepsilon\right)\right)\right] \\
&\leq \mathbf{Pr}\left[(\mathcal{E}_1 \vee \mathcal{E}_2) \bigvee \left(\exists T \in \binom{[n]}{\ell - t}, \exists i \in [D],\ \left|\boldsymbol{X}_T^{(i)}\right| \geq R + \beta\cdot\sqrt{2}\cdot R\right)\right] \\
&\leq (t-1)\cdot\varepsilon + \frac{\varepsilon}{3} + 2n^{\ell-t}\cdot e^{-\beta^2/2} \\
&\qquad\qquad \text{(due to (3.12), Claim 3.6.8, Lemma 3.4.5, and Claim 3.6.10)} \\
&\leq (t-1)\cdot\varepsilon + \frac{\varepsilon}{3} + \frac{1}{3}\cdot n^{\ell}\cdot e^{-\beta^2/2} \\
&\leq t\cdot\varepsilon. \qquad\qquad \square
\end{aligned}$$

Before we prove the complete level-ℓ bound for parity decision trees, we first prove a simple bound for the number of vectors with a given weight in a subspace.

Lemma 3.6.11. *Let $\ell \geq 1$ be an integer and $\mathcal{S}$ be a subspace of rank at most d. Let $U = \{S\colon |S| = \ell, S \in \mathcal{S}\}$, then $|U| \leq \min\left\{\binom{d\cdot\ell}{\ell}, 2^d - 1\right\}$.*

Proof. Let $\{S_1,\dots,S_{d'}\}$ be a maximal set of independent vectors in U. Then $d' \leq d$ and $|S_i| = \ell$ holds for all $i \in [d']$. Since $U \subseteq \mathsf{Span}\left\langle S_1,\dots,S_{d'}\right\rangle$ and $\emptyset \notin U$, we have

$$|U| \leq \left|\mathsf{Span}\left\langle S_1,\dots,S_{d'}\right\rangle\right| - 1 = 2^{d'} - 1 \leq 2^d - 1.$$

On the other hand, observe that $U \subseteq \binom{S_1\cup\cdots\cup S_{d'}}{\ell}$, hence we also have

$$|U| \leq \left|\binom{S_1\cup\cdots\cup S_{d'}}{\ell}\right| \leq \binom{d'\cdot\ell}{\ell} \leq \binom{d\cdot\ell}{\ell}. \qquad \square$$

We remark that in Lemma 3.6.11, it is conjectured the bound should be $\binom{d+1}{\ell}$ when $d \geq 2\cdot\ell$ [Kra10, BP18].

Theorem 3.6.12. *Let $\ell \geq 1$ be an integer. Let $\mathcal{T}\colon \{\pm1\}^n \to \{0,1\}$ be a depth-d parity decision tree where $n \geq \max\{d,\ell\}$. Let $p = \mathbf{Pr}\left[\mathcal{T}(x) = 1\right] \geq 2^{-d}$.*[9] *Then we have*

$$\sum_{S\subseteq[n]:|S|=\ell}\left|\widehat{\mathcal{T}}(S)\right| \leq p\cdot\min\left\{\binom{d\cdot\ell}{\ell}, 2^d-1,\ O\left(\sqrt{d}\cdot\log\left(\tfrac{n^\ell}{p}\right)\right)^\ell\right\} = O\left(\sqrt{d}\cdot\ell\cdot\log(n)\right)^\ell.$$

Proof. For any $S \in \binom{[n]}{\ell}$, let $a_S = \mathsf{sgn}\left(\widehat{\mathcal{T}}(S)\right)$. Now we prove the bounds separately.

[9] If $p < 2^{-d}$, then $p = 0$ and $\mathcal{T} \equiv 0$.

The First Two Bounds. Let $\boldsymbol{v}_0, \ldots, \boldsymbol{v}_{\boldsymbol{d}'}$ be a random root-to-leaf path. Then by the definition of $\widehat{\mathcal{P}_v}$ and $\mathcal{S}_v$ and Fact 3.5.2, we have

$$\begin{aligned}\sum_S \left|\widehat{\mathcal{T}}(S)\right| = \sum_S a_S \cdot \widehat{\mathcal{T}}(S) &= \mathop{\mathbb{E}}_{\boldsymbol{v}_0,\ldots,\boldsymbol{v}_{\boldsymbol{d}'}}\left[\mathcal{T}(\boldsymbol{v}_{\boldsymbol{d}'}) \cdot \sum_S a_S \cdot \widehat{\mathcal{P}_{\boldsymbol{v}_{\boldsymbol{d}'}}}(S)\right] \\ &\le \mathop{\mathbb{E}}_{\boldsymbol{v}_0,\ldots,\boldsymbol{v}_{\boldsymbol{d}'}}\left[\mathcal{T}(\boldsymbol{v}_{\boldsymbol{d}'}) \cdot \sum_S \left|\widehat{\mathcal{P}_{\boldsymbol{v}_{\boldsymbol{d}'}}}(S)\right|\right] = \mathop{\mathbb{E}}_{\boldsymbol{v}_0,\ldots,\boldsymbol{v}_{\boldsymbol{d}'}}\left[\mathcal{T}(\boldsymbol{v}_{\boldsymbol{d}'}) \cdot |\boldsymbol{V}|\right], \end{aligned} \tag{3.15}$$

where $a_S = \mathsf{sgn}\left(\widehat{\mathcal{T}}(S)\right)$ and $\boldsymbol{V} = \left\{S \in \binom{[n]}{\ell} \;\middle|\; S \in \mathcal{S}_{\boldsymbol{v}_{\boldsymbol{d}'}}\right\}$. Note that

$$\mathsf{rank}\left(\mathcal{S}_{\boldsymbol{v}_{\boldsymbol{d}'}}\right) = \mathsf{rank}\left(\mathsf{Span}\left\langle Q_{\boldsymbol{v}_0}, \ldots, Q_{\boldsymbol{v}_{\boldsymbol{d}'-1}}\right\rangle\right) \le \boldsymbol{d}' \le d.$$

Hence by Lemma 3.6.11, we have

$$(3.15) \le \min\left\{\binom{d\cdot\ell}{\ell}, 2^d - 1\right\} \cdot \mathbb{E}\left[\mathcal{T}(\boldsymbol{v}_{\boldsymbol{d}'})\right] = p \cdot \min\left\{\binom{d\cdot\ell}{\ell}, 2^d - 1\right\}.$$

The Third Bound. By Lemma 3.5.10, we construct a $2k$-clean parity decision tree $\mathcal{T}'$ of depth $D \le 2d \cdot k$ equivalent to $\mathcal{T}$, where $k = \Theta\left(\log\left(n^\ell/p\right)\right) \ge 4 \cdot \ell$. We also add dummy variables to make sure $n' = \max\{\tau, k, 6D, n\}$, where $\mathcal{T}'$ has n' inputs and τ is the universal constant in Lemma 3.6.6.

Let $\boldsymbol{u}_0, \ldots, \boldsymbol{u}_{\boldsymbol{D}'}$ be a random root-to-leaf path in $\mathcal{T}'$. Define $\boldsymbol{u}^{(0)}, \ldots, \boldsymbol{u}^{(\boldsymbol{D}')} \in \{-1, 0, +1\}^n$ by setting $\boldsymbol{u}_j^{(i)} = \widehat{\mathcal{P}_{\boldsymbol{u}_i}}(j)$ for each $0 \le i \le \boldsymbol{D}'$ and $j \in [n]$. Then extend $\boldsymbol{u}^{(\boldsymbol{D}'+1)} = \boldsymbol{u}^{(\boldsymbol{D}'+2)} = \cdots = \boldsymbol{u}^{(D)}$ to equal $\boldsymbol{u}^{(\boldsymbol{D}')}$. By Lemma 3.5.6, we have

$$\sum_S \left|\widehat{\mathcal{T}}(S)\right| = \sum_S \left|\widehat{\mathcal{T}'}(S)\right| = \mathop{\mathbb{E}}_{\boldsymbol{u}_0,\ldots,\boldsymbol{u}_{\boldsymbol{D}'}}\left[\mathcal{T}(\boldsymbol{u}_{\boldsymbol{D}'}) \cdot \sum_S a_S \cdot \boldsymbol{u}_S^{(D)}\right] \le \mathop{\mathbb{E}}_{\boldsymbol{u}_0,\ldots,\boldsymbol{u}_{\boldsymbol{D}'}}\left[\mathcal{T}(\boldsymbol{u}_{\boldsymbol{D}'}) \cdot |\boldsymbol{U}|\right], \tag{3.16}$$

where $\boldsymbol{U} = \sum_S a_S \cdot \boldsymbol{u}_S^{(D)}$.

Now we apply Lemma 3.6.6 with $t = \ell, \varepsilon = \Theta\left(p/d^{\ell/2}\right) \le 1/2$ to obtain the following bound[10]

$$M = M(D, d, k, \ell, \ell, \varepsilon) = \left(O\left(\sqrt{d} \cdot \log\left(\tfrac{n^\ell}{p}\right)\right)\right)^\ell$$

such that $\mathbf{Pr}\left[|\boldsymbol{U}| \ge M\right] \le \ell \cdot \varepsilon$. Then, combining the first bound, we have

$$(3.16) = \mathbb{E}\left[\mathcal{T}(\boldsymbol{u}_{\boldsymbol{D}'}) \cdot |\boldsymbol{U}| \cdot \left(1_{|\boldsymbol{U}|<M} + 1_{|\boldsymbol{U}|\ge M}\right)\right] \le M \cdot \mathbb{E}\left[\mathcal{T}(\boldsymbol{u}_{\boldsymbol{D}'})\right] + \ell \cdot \varepsilon \cdot \binom{d\cdot\ell}{\ell}$$

[10]Since $n \ge \max\{\ell, d\}$, we know $k = \Theta\left(\log\left(n^\ell/p\right)\right) = O(n^2)$ and $D \le 2d \cdot k = O(n^3)$. Hence $n' = \max\{\tau, k, 6D, n\} = O(n^3)$. Also $n^\ell/\varepsilon \le n^{O(\ell)}/p$ and by our choice of $k = \Theta\left(\log(n^\ell/p)\right)$ we have $\left(n^\ell/\varepsilon\right)^{6/k} = O(1)$.

$$= p \cdot \left(O\left(\sqrt{d} \cdot \log\left(\tfrac{n^\ell}{p}\right)\right)\right)^\ell,$$

which is maximized at $p = 1$, hence (3.16) $= O\Big(\sqrt{d} \cdot \ell \cdot \log(n)\Big)^\ell$ as desired. □

Finally we complete the missing proofs.

Proof of Claim 3.6.8

Claim (Claim 3.6.8 restated)**.** $\mathbf{Pr}\left[\mathcal{E}_2\right] \le \varepsilon/3$*, where* $\mathcal{E}_2$ *is the following event:*

$$\text{“}\exists T \in \binom{[n]}{\ell - t}, i, r, r',$$

$$\left|\mathbf{\Gamma}_T^{(i)}(r, r')\right| \ge \left(100 \min\left\{k, \log\left(\tfrac{n^\ell}{\varepsilon}\right)\right\} \cdot \left(\tfrac{n^\ell}{\varepsilon}\right)^{\frac{6}{k}}\right)^{\frac{r+r'}{2}} \cdot \boldsymbol{\sigma}_T(r, r', C(\boldsymbol{v}_{i-1}), i)\text{”}.$$

Proof. Let $k' = \min\left\{k, \left\lceil 6 \log\left(n^\ell/\varepsilon\right)\right\rceil\right\} \le 12 \min\left\{k, \log\left(n^\ell/\varepsilon\right)\right\}$. Then $\mathcal{T}$ is also a depth-D $2k'$-clean parity decision tree. Observe that

$$\begin{aligned}
&\mathbf{Pr}\left[\left|\mathbf{\Gamma}_T^{(i)}(r, r')\right| \ge \left(\frac{4k'}{\eta^{2/k'}}\right)^{(r+r')/2} \cdot \boldsymbol{\sigma}_T(r, r', C(\boldsymbol{v}_{i-1}), i)\right] \\
&\le \max_{C(\boldsymbol{v}_{i-1})} \mathbf{Pr}\left[\left|\mathbf{\Gamma}_T^{(i)}(r, r')\right| \ge \left(\frac{4k'}{\eta^{2/k'}}\right)^{(r+r')/2} \cdot \boldsymbol{\sigma}_T(r, r', C(\boldsymbol{v}_{i-1}), i) \,\middle|\, C(\boldsymbol{v}_{i-1})\right] \\
&\le \underbrace{\frac{(4 \cdot k')^{r+r'}}{(2 \cdot (r + r'))^{k'}}}_{\le 1} \cdot \underbrace{\eta^{2 - \frac{2(r+r')}{k'}}}_{\le \eta} \\
&\qquad\qquad \text{(due to the second bound in Lemma 2.0.9 and } k \ge 4 \cdot \ell \ge 4 \cdot (r + r')) \\
&\le \eta.
\end{aligned}$$

Thus by union bound over all $T \in \binom{[n]}{\ell - t}, i \in [\boldsymbol{D}'], r \in [t], 0 \le r' \le t - r$, we have

$$\begin{aligned}
\mathbf{Pr}\left[\exists T, i, r, r', \ \left|\mathbf{\Gamma}_T^{(i)}(r, r')\right| \ge \left(\tfrac{4k}{\eta^{2/k}}\right)^{(r+r')/2} \cdot \boldsymbol{\sigma}_T(r, r', C(\boldsymbol{v}_{i-1}), i)\right] &\le D t^2 n^{\ell - t} \cdot \eta \\
&\le \tfrac{n^{\ell+2} \cdot \eta}{3} \\
&\le \tfrac{n^{3 \cdot \ell} \cdot \eta}{3},
\end{aligned}$$

where we use the fact $n \ge \max\{D, 3 \cdot t\}$ and $t \ge 1$. By setting $\eta = \varepsilon/n^{3 \cdot \ell}$, we have

$$\frac{4k'}{\eta^{2/k'}} = 4k' \left(\frac{n^{3 \cdot \ell}}{\varepsilon}\right)^{\frac{2}{k'}} \le 4k' \left(\frac{n^\ell}{\varepsilon}\right)^{\frac{6}{k'}} \le 4 \cdot 12 \min\left\{k, \log\left(\frac{n^\ell}{\varepsilon}\right)\right\} \cdot 2 \left(\frac{n^\ell}{\varepsilon}\right)^{\frac{6}{k}},$$

as desired. □

Proof of Claim 3.6.10

We first need the following simple bound on M.

Lemma 3.6.13. *For any integer $s \geq 1$, we have*

$$\sum_{r=s}^{t} M(D,d,k,\ell,t-r,\varepsilon) \leq \frac{2 \cdot M(D,d,k,\ell,t,\varepsilon)}{(\tau D \cdot \log(n^\ell/\varepsilon))^{s/2}}.$$

Proof. We simply expand the formula of M as follows:

$$\begin{aligned}
\frac{\sum_{r=s}^{t} M(D,d,k,\ell,t-r,\varepsilon)}{M(D,d,k,\ell,t,\varepsilon)} &= \sum_{r=s}^{t} \left(\tau \cdot (D+dk) \cdot \left(\tfrac{n^\ell}{\varepsilon}\right)^{6/k} \log\left(\tfrac{n^\ell}{\varepsilon}\right)\right)^{-r/2} \\
&\leq \sum_{r=s}^{+\infty} \left(\tau \cdot (D+dk) \cdot \left(\tfrac{n^\ell}{\varepsilon}\right)^{6/k} \log\left(\tfrac{n^\ell}{\varepsilon}\right)\right)^{-r/2} \\
&\leq 2 \cdot \left(\tau \cdot (D+dk) \cdot \left(\tfrac{n^\ell}{\varepsilon}\right)^{6/k} \log\left(\tfrac{n^\ell}{\varepsilon}\right)\right)^{-s/2} \\
&\qquad \text{(due to } \tau \geq 4 \text{ and } s \geq 1\text{)} \\
&\leq 2 \cdot \left(\tau D \cdot \log\left(n^\ell/\varepsilon\right)\right)^{-s/2}. \qquad \square
\end{aligned}$$

Now we prove Claim 3.6.10.

Claim (Claim 3.6.10 restated)**.** *When $\mathcal{E}_1 \vee \mathcal{E}_2$ does not happen, we have $\sum_{i=1}^{D} \left|\boldsymbol{\mu}_T^{(i)}\right| \leq R$ and $\sum_{i=1}^{D} \left|\boldsymbol{\delta}_T^{(i)}\right|^2 \leq R^2$ hold for all $T \in \binom{[n]}{\ell-t}$, where*

$$R = \frac{M(D,d,k,\ell,t,\varepsilon)}{5 \cdot \sqrt{\log(n^\ell/\varepsilon)}}.$$

Proof. We verify for each $T \in \binom{[n]}{\ell-t}$ as follows:

$$\begin{aligned}
\sum_{i=1}^{D} \left|\boldsymbol{\mu}_T^{(i)}\right| = \sum_{i=1}^{\boldsymbol{D'}} \left|\boldsymbol{\mu}_T^{(i)}\right| &\leq \sum_{i=1}^{\boldsymbol{D'}} \sum_{\substack{r=2,\\ \text{even}}}^{t} |\boldsymbol{A}(T,r,i)| && \text{(due to (3.11))} \\
&\leq \sum_{i=1}^{\boldsymbol{D'}} \sum_{\substack{r=2,\\ \text{even}}}^{t} \left(M(D,d,k,\ell,t-r,\varepsilon) + \tfrac{M(D,d,k,\ell,t,\varepsilon)}{\sqrt{\log(n^\ell/\varepsilon)}} \cdot \sqrt{\left(\tfrac{800}{\tau}\right)^r \left(\tfrac{|J(\boldsymbol{v}_{i-1})|}{2d}\right)^r} \cdot 1_{|J(\boldsymbol{v}_{i-1})|>1} \right) \\
&\qquad \text{(due to (3.13))} \\
&\leq \sum_{i=1}^{\boldsymbol{D'}} \sum_{\substack{r=2,\\ \text{even}}}^{t} \left(M(D,d,k,\ell,t-r,\varepsilon) + \tfrac{M(D,d,k,\ell,t,\varepsilon)}{\sqrt{\log(n^\ell/\varepsilon)}} \cdot \left(\tfrac{|J(\boldsymbol{v}_{i-1})|}{2d}\right) \left(\tfrac{800}{\tau}\right)^{r/2} \cdot 1_{|J(\boldsymbol{v}_{i-1})|>1} \right) \\
&\qquad \text{(Since } |J(\boldsymbol{v}_{i-1})| \leq 2d \text{ from Corollary 3.5.9)}
\end{aligned}$$

$$\leq \frac{2\cdot M(D,d,k,\ell,t,\varepsilon)}{\tau\cdot\log\left(n^\ell/\varepsilon\right)} + \frac{1.1\cdot 800\cdot M(D,d,k,\ell,t,\varepsilon)}{\tau\cdot\sqrt{\log\left(n^\ell/\varepsilon\right)}}$$

(due to Lemma 3.6.13 and Corollary 3.5.9 and $\tau = 10^4$)

$$\leq \frac{M(D,d,k,\ell,t,\varepsilon)}{5\cdot\sqrt{\log\left(n^\ell/\varepsilon\right)}} = R$$

and with similar calculation, we have

$$\sum_{i=1}^{D}\left|\boldsymbol{\delta}_T^{(i)}\right|^2$$

$$\leq \sum_{i=1}^{\boldsymbol{D'}}\left(\sum_{\substack{r=1,\\ \text{odd}}}^{t}\left(M(D,d,k,\ell,t-r,\varepsilon) + \frac{M(D,d,k,\ell,t,\varepsilon)}{\sqrt{\log\left(n^\ell/\varepsilon\right)}}\cdot\sqrt{\frac{|J(\boldsymbol{v}_{i-1})|}{2d}}\left(\frac{800}{\tau}\right)^{r/2}\cdot 1_{|J(\boldsymbol{v}_{i-1})|>1}\right)\right)^2$$

$$\leq \sum_{i=1}^{\boldsymbol{D'}}\left(\frac{2\cdot M(D,d,k,\ell,t,\varepsilon)}{\sqrt{\tau D\cdot\log\left(n^\ell/\varepsilon\right)}} + \frac{1.1\cdot\sqrt{800}\cdot M(D,d,k,\ell,t,\varepsilon)}{\sqrt{\tau}\sqrt{\log\left(n^\ell/\varepsilon\right)}}\cdot\sqrt{\frac{|J(\boldsymbol{v}_{i-1})|}{2d}}\cdot 1_{|J(\boldsymbol{v}_{i-1})|>1}\right)^2 \quad \text{(due to } \tau = 10^4\text{)}$$

$$\leq \left(\frac{M(D,d,k,\ell,t,\varepsilon)}{\sqrt{\log\left(n^\ell/\varepsilon\right)}}\right)^2\sum_{i=1}^{\boldsymbol{D'}} 2\cdot\left(\frac{4}{\tau D} + \frac{968}{\tau}\cdot\frac{|J(\boldsymbol{v}_{i-1})|}{2d}\cdot 1_{|J(\boldsymbol{v}_{i-1})|>1}\right) \quad \text{(due to } (a+b)^2\leq 2(a^2+b^2)\text{)}$$

$$\leq \left(\frac{2000\cdot M(D,d,k,\ell,t,\varepsilon)}{\tau\cdot\sqrt{\log\left(n^\ell/\varepsilon\right)}}\right)^2 = R^2. \qquad \square$$

3.7 Fourier Growth Bounds for Noisy Decision Trees

We establish the Fourier growth bounds for noisy decision trees in this section. We begin by defining the model.

Definition 3.7.1 (Noisy oracle)**.** *A noisy query to a bit $b\in\{\pm1\}$ with correlation $\gamma\in[-1,1]$ returns a bit $\boldsymbol{b}'\in\{\pm1\}$ where*

$$\boldsymbol{b}' = \begin{cases} b & \text{with probability } (1+\gamma)/2,\\ -b & \text{with probability } (1-\gamma)/2.\end{cases}$$

The cost of a noisy query with correlation γ is defined to be γ^2.

Definition 3.7.2 (Noisy decision tree)**.** *A* noisy decision tree $\mathcal{T}$ *is a rooted binary tree in which each internal node v is labeled by an index $q_v\in[n]$ and a correlation $\gamma_v\in[-1,1]$. The outgoing edges are labeled by $+1$ and -1 and the leaves are labeled by 0 and 1.*

On input $x\in\{\pm1\}^n$, the tree $\mathcal{T}$ constructs a computation path $\mathcal{P}$ *from the root to leaf as follows. When $\mathcal{P}$ reaches an internal node v, it makes a noisy query to x_{q_v} with correlation γ_v and follows the edge labeled by the outcome of this noisy query. The output of the tree is defined by sampling a root-to-leaf path and returning the label of the leaf. Since the*

computation path $\mathcal{P}$ is probabilistic, this is an inherently randomized model of computation. We use $\mathcal{T}(x) \in \{0,1\}$ to denote the (probabilistic) output of $\mathcal{T}$ on input x. We also use $\mathcal{T}(v) \in \{0,1\}$ to denote the label on v when v is a leaf. We do not *require that the indices q_v queried along a path $\mathcal{P}$ are distinct. The* cost *of any path is the sum of costs of the noisy queries along that path; and the cost of $\mathcal{T}$ is the maximum cost of any root-to-leaf path.*

We remark that for any noisy decision tree $\mathcal{T}$, its Fourier coefficient $\widehat{\mathcal{T}}(S)$ is given by $\mathbb{E}[\mathcal{T}(x)x_S]$ where the expectation is over the randomness of both $\boldsymbol{x} \sim \mathcal{U}_n$ and $\mathcal{T}$.

Let $\mathcal{T}$ be a noisy decision tree. By adding queries with zero correlation, we assume without loss of generality each root-to-leaf path in the noisy decision tree is of the same length. Let v be any node of $\mathcal{T}$. We use $\mathcal{P}_v$ to denote the uniform distribution over $\{\pm 1\}^n$ *conditioning* on reaching v. Note that $\mathcal{P}_v$ is always a *product distribution*. As before, for any $S \subseteq [n]$ we define $\widehat{\mathcal{P}_v}(S) = \mathbb{E}_{\boldsymbol{x}\sim\mathcal{P}_v}[\boldsymbol{x}_S]$.

Claim 3.7.3. *Let $\mathcal{T}\colon \{\pm 1\}^n \to \{0,1\}$ be a cost-d noisy decision tree. Let $\boldsymbol{v}_0, \dots, \boldsymbol{v}_D$ be any root-to-leaf path in $\mathcal{T}$. Define $\boldsymbol{v}^{(0)}, \dots, \boldsymbol{v}^{(D)} \in [-1,1]^n$ by setting $\boldsymbol{v}_j^{(i)} = \widehat{\mathcal{P}_{\boldsymbol{v}_i}}(j)$ for each $0 \le i \le D$ and $j \in [n]$. Then for any $i \in \{0, \dots, D-1\}$, $\boldsymbol{v}_{q_{\boldsymbol{v}_i}}^{(i+1)} - \boldsymbol{v}_{q_{\boldsymbol{v}_i}}^{(i)}$ is a mean-zero random variable with magnitude bounded by $2 \cdot |\gamma_{\boldsymbol{v}_i}|$.*

Proof. Fix $i \in \{0, \dots, D-1\}$. For convenience, let $j = q_{\boldsymbol{v}_i}$, $\gamma = \gamma_{\boldsymbol{v}_i}$, and $\alpha = \boldsymbol{v}_j^{(i)}$. Suppose $|\gamma| = 1$ then $\left|\boldsymbol{v}_j^{(i+1)} - \boldsymbol{v}_j^{(i)}\right| \le 2 = 2 \cdot |\gamma_{\boldsymbol{v}_i}|$ as desired. Now we turn to the case $|\gamma| < 1$.

Note that for the distribution $\mathcal{P}_{\boldsymbol{v}_i}$, the measure of $x_j = 1$ (resp., $x_j = -1$) inputs is $(1+\alpha)/2$ (resp., $(1-\alpha)/2$). The measure of $x_j = 1$ (resp., $x_j = -1$) inputs that follow the edge labeled 1 is $a := (1+\alpha)(1+\gamma)/4$ (resp., $b := (1-\alpha)(1-\gamma)/4$). The total measure of inputs that take the edge labeled 1 is $a + b$ and the resulting node $\boldsymbol{v}_{i+1}$ satisfies $\boldsymbol{v}_j^{(i+1)} = (a-b)/(a+b)$. This implies that

$$\boldsymbol{v}_j^{(i+1)} = \begin{cases} \frac{\alpha+\gamma}{1+\gamma\cdot\alpha} & \text{with probability } \frac{1+\gamma\cdot\alpha}{2}, \\ \frac{\alpha-\gamma}{1-\gamma\cdot\alpha} & \text{with probability } \frac{1-\gamma\cdot\alpha}{2}. \end{cases}$$

The above calculation implies

$$\boldsymbol{v}_j^{(i+1)} - \boldsymbol{v}_j^{(i)} = \begin{cases} \gamma \cdot \frac{1-\alpha^2}{1+\gamma\cdot\alpha} & \text{with probability } \frac{1+\gamma\cdot\alpha}{2}, \\ -\gamma \cdot \frac{1-\alpha^2}{1-\gamma\cdot\alpha} & \text{with probability } \frac{1-\gamma\cdot\alpha}{2}, \end{cases}$$

and thus $\boldsymbol{v}_j^{(i+1)} - \boldsymbol{v}_j^{(i)}$ is a mean-zero random variable. Since $\alpha \in [-1,1]$ and $\gamma \in (-1,1)$, we have

$$\max\left\{\frac{1-\alpha^2}{1-\gamma\cdot\alpha}, \frac{1-\alpha^2}{1+\gamma\cdot\alpha}\right\} \le \frac{1-\alpha^2}{1-|\alpha|} = 1 + |\alpha| \le 2,$$

which implies $\left|\boldsymbol{v}_j^{(i+1)} - \boldsymbol{v}_j^{(i)}\right| \le 2 \cdot |\gamma|$. □

We now prove the general Fourier bounds. As before, for any $S \subseteq [n]$, let $\boldsymbol{v}_S^{(i)}$ be $\prod_{j \in S} \boldsymbol{v}_j^{(i)}$.

Lemma 3.7.4. *There exists a universal constant τ such that the following holds. Let $\ell \geq 1$ be an integer. Let $\mathcal{T}\colon \{\pm 1\}^n \to \{0,1\}$ be a cost-d noisy decision tree.*

Let $\boldsymbol{v}_0, \dots, \boldsymbol{v}_D$ be a random root-to-leaf path in $\mathcal{T}$. Define $\boldsymbol{v}^{(0)}, \dots, \boldsymbol{v}^{(D)} \in [-1,1]^n$ by setting $\boldsymbol{v}_j^{(i)} = \widehat{\mathcal{P}_{\boldsymbol{v}_i}}(j)$ for each $0 \leq i \leq D$ and $j \in [n]$. Then for any sequence $a_S \in \{-1,0,1\}, S \in \binom{[n]}{\ell}$, any $\varepsilon \leq 1/2$ and $t \in \{0, \dots, \ell\}$, we have

$$\mathbf{Pr}\left[\exists T \in \binom{[n]}{\ell - t}, \exists i \in [D], \left|\sum_{S \subseteq \overline{T}, |S|=t} a_{S \cup T} \cdot \boldsymbol{v}_S^{(i)}\right| \geq S(d,\ell,t,\varepsilon)\right] \leq \varepsilon \cdot t,$$

where $S(d,\ell,0,\varepsilon) = 1$ and

$$S(d,\ell,t,\varepsilon) = \sqrt{(\tau \cdot d)^t \cdot \log\left(\tfrac{n^{\ell-t}}{\varepsilon}\right) \cdots \log\left(\tfrac{n^{\ell-1}}{\varepsilon}\right)} \qquad \text{for } t \in [\ell].$$

Proof. We prove the bound by induction on t and show $\tau = 32$ suffices. The base case $t = 0$ is trivial, since for any T of size ℓ and any i, we have $\left|a_T \cdot \boldsymbol{v}_\emptyset^{(i)}\right| \leq 1 = S(d,\ell,0,\varepsilon)$.

Now we focus on the case $1 \leq t \leq \ell$. For any $T \in \binom{[n]}{\leq \ell}$, define $\boldsymbol{X}_T^{(0)}, \dots, \boldsymbol{X}_T^{(D)}$ by $\boldsymbol{X}_T^{(i)} = \sum_{S \subseteq \overline{T}, |S|+|T|=\ell} a_{S \cup T} \cdot \boldsymbol{v}_S^{(i)}$. Define $\boldsymbol{\delta}_T^{(i)}$ for $i \in [D]$ as follows:

$$\begin{aligned}
\boldsymbol{\delta}_T^{(i)} = \boldsymbol{X}_T^{(i)} - \boldsymbol{X}_T^{(i-1)} &= \sum_{S \subseteq \overline{T}, |S|=t, S \ni q_{\boldsymbol{v}_{i-1}}} a_{S \cup T} \cdot \left(\boldsymbol{v}_S^{(i)} - \boldsymbol{v}_S^{(i-1)}\right) \\
&= \left(\boldsymbol{v}_{q_{\boldsymbol{v}_{i-1}}}^{(i)} - \boldsymbol{v}_{q_{\boldsymbol{v}_{i-1}}}^{(i-1)}\right) \cdot \sum_{S' \subseteq \overline{T \cup \{q_{\boldsymbol{v}_{i-1}}\}}, |S'|=t-1} a_{S' \cup \{q_{\boldsymbol{v}_{i-1}}\} \cup T} \cdot \boldsymbol{v}_S^{(i-1)} \\
&= \left(\boldsymbol{v}_{q_{\boldsymbol{v}_{i-1}}}^{(i)} - \boldsymbol{v}_{q_{\boldsymbol{v}_{i-1}}}^{(i-1)}\right) \cdot \boldsymbol{X}_{T \cup \{q_{\boldsymbol{v}_{i-1}}\}}^{(i-1)}.
\end{aligned}$$

Note that by Claim 3.7.3 and conditioning on $\boldsymbol{v}_{i-1}$, $\boldsymbol{\delta}_T^{(i)}$ is a mean-zero random variable.

The induction hypothesis implies that with all but $\varepsilon \cdot (t-1)$ probability, for all $i \in [D]$ and $T' \in \binom{[n]}{\ell - t + 1}$, we have $\left|\boldsymbol{X}_{T'}^{(i)}\right| \leq S(d,\ell,t-1,\varepsilon)$. By Claim 3.7.3, we have

$$\left|\boldsymbol{\delta}_T^{(i)}\right| = \left|\boldsymbol{v}_{q_{\boldsymbol{v}_{i-1}}}^{(i)} - \boldsymbol{v}_{q_{\boldsymbol{v}_{i-1}}}^{(i-1)}\right| \cdot \left|\boldsymbol{X}_{T \cup \{q_{\boldsymbol{v}_{i-1}}\}}^{(i-1)}\right| \leq 2 \cdot \left|\gamma_{\boldsymbol{v}_{i-1}}\right| \cdot S(d,\ell,t-1,\varepsilon).$$

Denote by $\boldsymbol{\Delta}_T^{(i)} = 2 \cdot \left|\gamma_{\boldsymbol{v}_{i-1}}\right| \cdot S(d,\ell,t-1,\varepsilon)$. We can thus express $\boldsymbol{X}_T^{(i)} = \boldsymbol{X}_T^{(i-1)} + \boldsymbol{\Delta}_T^{(i)} \cdot \boldsymbol{z}_T^{(i)}$ where $\left|\boldsymbol{z}_T^{(i)}\right| \leq 1$. Then we apply Lemma 3.4.5 to the family of martingales $\boldsymbol{X}_T^{(0)}, \dots, \boldsymbol{X}_T^{(D)}, |T| \in \binom{[n]}{\ell - t}$ with difference sequence $\boldsymbol{\delta}_T^{(i)} = \boldsymbol{\Delta}_T^{(i)} \cdot \boldsymbol{z}_T^{(i)}$ satisfying

$$\sum_{i=1}^{D} \left(\boldsymbol{\Delta}_T^{(i)}\right)^2 = 4 \cdot (S(d,\ell,t-1,\varepsilon))^2 \cdot \sum_{i=1}^{D} \left|\gamma_{\boldsymbol{v}_{i-1}}\right|^2 \leq 4d \cdot (S(d,\ell,t-1,\varepsilon))^2.$$

Hence for any $\beta \geq 0$, we have

$$\mathbf{Pr}\left[\exists T \in \binom{[n]}{\ell-t}, \exists i \in [D], \left|\boldsymbol{X}_T^{(i)}\right| \geq 2\beta \cdot \sqrt{2d} \cdot S(d,\ell,t-1,\varepsilon)\right] \leq \varepsilon \cdot (t-1) + 2 \cdot n^{\ell-t} \cdot e^{-\beta^2/2}.$$

Since $\varepsilon \leq 1/2$, we can set $\beta = 2 \cdot \sqrt{\log(n^{\ell-t}/\varepsilon)}$ so that $2 \cdot n^{\ell-t} \cdot e^{-\beta^2/2} \leq \varepsilon$, which completes the induction by noticing

$$2\beta \cdot \sqrt{2d} \cdot S(d,\ell,t-1,\varepsilon) = \sqrt{32 \cdot d \cdot \log\left(\tfrac{n^{\ell-t}}{\varepsilon}\right)} \cdot S(d,\ell,t-1,\varepsilon) \leq S(d,\ell,t,\varepsilon). \qquad \square$$

Theorem 3.7.5. *Let $\ell \geq 1$ and $n \geq \max\{\ell, 2\}$ be integers. Let $\mathcal{T}\colon \{\pm 1\}^n \to \{0,1\}$ be a cost-d noisy decision tree. Let $p = \mathbf{Pr}[\mathcal{T}(\boldsymbol{x}) = 1] \in (0, 1/2]$.*[11] *Then we have*

$$\sum_{S \subseteq [n], |S| = \ell} \left|\widehat{\mathcal{T}}(S)\right| \leq p \cdot O(d)^{\ell/2} \cdot \sqrt{\log\left(\tfrac{1}{p}\right)\left(\log\left(\tfrac{n^\ell}{p}\right)\right)^{\ell-1}} = O(d)^{\ell/2} \cdot \sqrt{1 + (\ell \log(n))^{\ell-1}}.$$

Proof. For any $S \in \binom{[n]}{\ell}$, let $a_S = \mathsf{sgn}\left(\widehat{\mathcal{T}}(S)\right)$. Let $\boldsymbol{v}_0, \ldots, \boldsymbol{v}_D$ be a random root-to-leaf path in $\mathcal{T}$. Note that

$$\sum_S \left|\widehat{\mathcal{T}}(S)\right| = \sum_S a_S \cdot \widehat{\mathcal{T}}(S) = \mathbb{E}\left[\mathcal{T}(\boldsymbol{v}_D) \cdot \sum_S a_S \cdot \boldsymbol{v}_S^{(D)}\right] \leq \mathbb{E}\left[\mathcal{T}(\boldsymbol{v}_D) \cdot |\boldsymbol{V}|\right], \tag{3.17}$$

where $\boldsymbol{V} = \sum_S a_S \cdot_S \boldsymbol{v}_S^{(D)}$. By Lemma 3.7.4, we know $\mathbf{Pr}\left[|\boldsymbol{V}| \geq S(\varepsilon)\right] \leq \varepsilon \cdot \ell$, where

$$S(\varepsilon) = S(d,\ell,\ell,\varepsilon) = \sqrt{O(d)^\ell \cdot \log\left(\tfrac{n^{\ell-1}}{\varepsilon}\right) \cdots \log\left(\tfrac{n^0}{\varepsilon}\right)} \leq \sqrt{O(d)^\ell \cdot \left(\log\left(\tfrac{n^{\ell-1}}{\varepsilon}\right)\right)^{\ell-1} \cdot \log\left(\tfrac{1}{\varepsilon}\right)}.$$

For integer $i \geq 1$, let $I_i = \left[S\left(p/\left(\ell 2^i\right)\right), S\left(p/\left(\ell 2^{i+1}\right)\right)\right]$ and $I_0 = [0, S(p/\ell)]$ be intervals. Then for each $i \geq 1$, $\mathbf{Pr}\left[|\boldsymbol{V}| \in I_i\right] \leq p/2^i$. We also know that $\mathbb{E}_{\boldsymbol{v}_0,\ldots,\boldsymbol{v}_D}\left[\mathcal{T}(\boldsymbol{v}_D)\right] \leq p$. Thus,

$$\begin{aligned}
(3.17) &\leq \mathop{\mathbb{E}}_{\boldsymbol{v}_0,\ldots,\boldsymbol{v}_D}\left[\mathcal{T}(\boldsymbol{v}_D) \cdot |\boldsymbol{V}| \cdot \sum_{i=0}^{+\infty} 1_{|\boldsymbol{V}| \in I_i}\right] \\
&\leq S\left(\frac{p}{\ell}\right) \cdot \mathbb{E}\left[\mathcal{T}(\boldsymbol{v}_D)\right] + \sum_{i=1}^{+\infty} S\left(\frac{p}{\ell \cdot 2^{i+1}}\right) \cdot \mathbb{E}\left[1_{|\boldsymbol{V}| \in I_i}\right] \\
&\leq \sum_{i=0}^{+\infty} S\left(\frac{p}{\ell \cdot 2^{i+1}}\right) \cdot \frac{p}{2^i} \\
&= \sum_{i=0}^{+\infty} p \cdot \sqrt{O(d)^\ell \cdot \left(\log\left(\tfrac{n^{\ell-1} \cdot \ell}{p}\right) + i + 1\right)^{\ell-1} \cdot \left(\log\left(\tfrac{1}{p}\right) + \log(\ell) + i + 1\right)} \cdot \frac{1}{2^i}
\end{aligned}$$

[11] If $p > 1/2$, then we can consider $\widetilde{\mathcal{T}} = 1 - \mathcal{T}$ by symmetry.

$$\leq \sum_{i=0}^{+\infty} p \cdot \sqrt{O(d)^{\ell} \cdot \left(\left(\log\left(\tfrac{n^{\ell}}{p}\right)\right)^{\ell-1} + (i+1)^{\ell-1}\right) \cdot \left(\log\left(\tfrac{1}{p}\right) + i + 1\right)} \cdot \frac{1}{2^i}$$

(since $n \geq \ell$, and $(x+y)^b \leq 2^b \cdot \left(x^b + y^b\right)$ and $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$ for $x, y, b \geq 0$)

$$\leq p \cdot \sqrt{O(d)^{\ell} \cdot \log\left(\tfrac{1}{p}\right)\left(\log\left(\tfrac{n^{\ell}}{p}\right)\right)^{\ell-1}},$$

where the last inequality follows from $p \leq 1/2$, $n \geq 2$ and

$$\sum_{i=0}^{+\infty} (i+1)^{\ell/2} \cdot 2^{-i} = O(\ell)^{\ell/2} \leq O(1)^{\ell} \cdot \ell^{(\ell-1)/2} \leq O(1)^{\ell} \cdot \left(\log\left(n^{\ell}/p\right)\right)^{(\ell-1)/2}.$$

Note that $p \cdot (\log(1/p))^k \leq O(k)^k$ for $p \in (0,1)$ and $k \geq 0$, thus

$$\begin{aligned} p \cdot \sqrt{\log\left(\tfrac{1}{p}\right)\left(\log\left(\tfrac{n^{\ell}}{p}\right)\right)^{\ell-1}} &= p \cdot \sqrt{\log\left(\tfrac{1}{p}\right)\left(\ell\log(n) + \log\left(\tfrac{1}{p}\right)\right)^{\ell-1}} \\ &\leq O(1)^{\ell} \cdot \left(\sqrt{(\ell\log(n))^{\ell-1}} + \ell^{\ell/2}\right) \\ &= O(1)^{\ell} \cdot \sqrt{1 + (\ell\log(n))^{\ell-1}}. \end{aligned}$$

□

Chapter 4

Quantum Advantages over Classical Communication

The level-k 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 $\boldsymbol{x} \odot \boldsymbol{y}$ of the inputs x and y to Alice and Bob). If a protocol $\mathcal{C}$ computes an XOR function, then $\mathcal{C}(x, y)$ depends only on $\boldsymbol{x} \odot \boldsymbol{y}$. This motivates us to analyze the *XOR-fiber* of the communication protocol $\mathcal{C}$, defined as $h(z) := \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}}[\mathcal{C}(\boldsymbol{x}, \boldsymbol{y}) | \boldsymbol{x} \odot \boldsymbol{y} = z]$.

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

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

Organization. In Section 4.1, we give a brief introduction on XOR functions. In Section 4.2, we summarize our main results and applications. An overview of our proofs is given in Section 4.3. In Section 4.4, we quote a useful concentration inequality and give a self-

contained proof. Section 4.5 explains a way to associate the Fourier growth to a martingale process. The proof of level-one bound (Theorem 4.2.1) is given in Section 4.6, and the level-two bound (Theorem 4.2.2) in Section 4.7. The Fourier growth reductions between general gadgets are presented in Section 4.8. The future directions are discussed in Section 4.9.

4.1 Introduction

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 [MO09, 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 bitwise 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 $x \odot y$, 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 4.1.1. *Let $\mathcal{C} : \{\pm 1\}^n \times \{\pm 1\}^n \to \{\pm 1\}$ be any deterministic communication protocol. The XOR-fiber of the communication protocol $\mathcal{C}$ is the function $h \colon \{\pm 1\}^n \to [-1, 1]$ defined at $z \in \{\pm 1\}^n$ as*

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

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 in the Query Model. 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-1 Fourier growth. By inductive tree decompositions, Tal [Tal20] obtained bounds for the higher levels of the form $L_k(f) \leq \sqrt{d^k \cdot O(\log(n))^{k-1}}$. This was later sharpened by Sherstov, Storozhenko, and Wu [SSW23] to the asymptotically tight bound of $L_k(f) \leq \sqrt{\binom{d}{k} \cdot O(\log(n))^{k-1}}$ 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-1 bound of $O(\sqrt{d})$. For higher levels, Girish, Tal, and Wu [GTW21] showed that $L_k(f) \leq \sqrt{d^k \cdot O(k \log(n))^{2k}}$. See Chapter 3 for details. 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.

Related Works in the Communication Model. 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 [GRT22] showed an $O(d^k)$ Fourier growth[1] for level-k. For level-1 and level-2, 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 $d \cdot \mathsf{polylog}(n)$ for level two already requires sophisticated ideas.

The bounds in [GRT22] 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 [GRT22], 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

[1] Technically, [GRT22] only proved a level-2 bound (as it suffices for their analysis), but a level-k bound follows easily from their proof approach, as noted by [GRZ21].

complicated ways that such methods break down. This is one of the key difficulties we face here.

4.2 Our 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 4.2.1. *Let $\mathcal{C} : \{\pm 1\}^n \times \{\pm 1\}^n \to \{\pm 1\}$ be a deterministic communication protocol with at most d bits of communication. Let h be its XOR-fiber as in Definition 4.1.1. Then, $L_1(h) = O\left(\sqrt{d}\right)$.*

Theorem 4.2.2. *Let $\mathcal{C} : \{\pm 1\}^n \times \{\pm 1\}^n \to \{\pm 1\}$ be a deterministic protocol communicating at most d bits. Let h be its XOR-fiber as in Definition 4.1.1. Then, $L_2(h) = O\left(d^{3/2} \log^3(n)\right)$.*

Our bounds in Theorems 4.2.1 and 4.2.2 extend directly to randomized communication protocols. This is because L_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 Theorem 4.2.8, to demonstrate that these bounds apply to general constant-sized gadgets g and the corresponding g-fiber.

Our level-1 and level-2 bounds improve previous bounds in [GRT22] by polynomial factors. Additionally, our level-1 bound is tight since a deterministic protocol with $d + 1$ bits of communication can compute the majority vote of $x_1 \cdot y_1, \ldots, x_d \cdot y_d$, which corresponds to $h(z) = \mathrm{MAJ}(z_1, \ldots, z_d)$ with $L_1(h) = \Theta(\sqrt{d})$.

In terms of techniques, our analysis presents a key new idea that enables us to exploit cancellations between different rectangles induced by the protocol. This idea involves using a novel process to adaptively partition a relatively large set in Gaussian space, which enables us to control its k-wise moments in all directions — this can be thought of as a spectral notion of almost k-wise independence. We achieve this by utilizing martingale arguments and allowing protocols to transmit *real values* rather than just discrete bits. This notion and procedure may be of independent interest. See Section 4.3 for a detailed discussion.

Below, we describe some applications and connections of our main theorems.

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 $\mathsf{AC}^0[\oplus]$ circuits [CGR14, LSS+19], product tests [LV18], and more. Recently, Agrawal [Agr20] showed that the coin problem is closely related to the level-1 Fourier growth of functions in $\mathcal{F}$.

Lemma 4.2.3 ([Agr20, Lemma 3.2])**.** *Assume that $\mathcal{F}$ is closed under restrictions and satisfies $L_1(f) \le t$ for all $f \in \mathcal{F}$. Then, for all $\rho \in (-1,1)$ and $f \in \mathcal{F}$,*

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

Note that communication protocols of small cost are closed under restrictions, so are their XOR-fibers (see [GRT22, 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 Theorem 4.2.8), Theorem 4.2.4 can be established for general gadgets of small size.

Theorem 4.2.4. *Let h be the XOR-fiber of a protocol with total communication d. Then for all ρ,*

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

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 4.2.4 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 $\langle x, y\rangle \le -\sqrt{n}$ or $\langle x, y\rangle \ge \sqrt{n}$.

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 4.2.4, we derive a new proof for the $\Omega(n)$ lower bound on the communication complexity of the Gap-Hamming Problem.

Theorem 4.2.5. *The randomized communication complexity of the Gap-Hamming Problem is $\Omega(n)$.*

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

Proof. 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 4.2.4 on ρ as well as $-\rho$, and apply triangle inequality to conclude that

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

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

$$\left| \mathop{\mathbb{E}}_{(\boldsymbol{x},\boldsymbol{y}) \sim \sigma_\rho} [\mathcal{C}(\boldsymbol{x}, \boldsymbol{y})] - \mathop{\mathbb{E}}_{(\boldsymbol{x},\boldsymbol{y}) \sim \sigma_{-\rho}} [\mathcal{C}(\boldsymbol{x}, \boldsymbol{y})] \right| < 1/9. \tag{4.1}$$

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

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

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

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

and

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

This, along with (4.1) and triangle inequality, implies that

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

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

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-1 Fourier growth, whereas XOR-fibers of small-cost protocols have small Fourier growth.

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, Göö15, 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 $\mathsf{polylog}(n)$. 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,

$$\mathrm{forr}(x) = \frac{2}{n} \langle H x_1, x_2 \rangle ,$$

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

Girish, Raz, and Tal [GRT22] studied the XOR-lift of the Forrelation problem and obtained new separations between quantum and randomized communication protocols. In more detail, they considered the partial function[3] $\mathrm{forr} \circ \mathrm{XOR} \colon \{\pm 1\}^n \times \{\pm 1\}^n \to \{\pm 1\}$ defined as

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

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

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

The lower bound in [GRT22] was obtained from a second level Fourier growth bound (higher levels are not needed) on the XOR-fiber of classical communication protocols. Our level-2 bound strengthens their bound and immediately gives an improved communication lower bound.

Theorem 4.2.6. *The randomized communication complexity of* forr $\circ$ XOR *is* $\widetilde{\Omega}(n^{1/3})$.

Theorem 4.2.6 above gives an $\mathsf{polylog}(n)$ versus $\widetilde{\Omega}(n^{1/3})$ separation between the above quantum communication model and the randomized two-party communication model, improving upon the $\mathsf{polylog}(n)$ versus $\widetilde{\Omega}(n^{1/4})$ separation from [GRT22]. We emphasize that our separations are for players with *efficient quantum* running time, where the only prior separation was shown by the aforementioned work [GRT22]. 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 $\mathsf{polylog}(n)$ versus $\widetilde{\Omega}(\sqrt{n})$ separation is known [Gav20] (see [GRT22, Section 1.1] for a more detailed comparison). Optimal Fourier growth bounds of $d \cdot \mathsf{polylog}(n)$ for level two, which we state later in Conjecture 4.2.7, 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 $\widetilde{\Omega}(n^{1-1/k})$, confirming this conjecture, and a similar separation was proven in [SSW23] 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 $\widetilde{\Omega}(n^{1-1/k})$ separation between two-party quantum and classical communication complexity, where additionally, the number of rounds[4] in the two-party quantum protocol is $2 \cdot \lceil k/2 \rceil$.

Replacing the inner product gadget with the XOR gadget above would yield an improved quantum-classical communication separation where the gadget is simpler and the number of rounds required by the quantum protocol to achieve the same quantitative separation is reduced by half. Bansal and Sinha [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. See Theorem 2.0.12. Thus, in conjunction with their results, to prove the

[4]We remark that for $k = 2$, this is exactly the XOR-lift of the Forrelation problem and can even be computed in the quantum simultaneous model, as shown in [GRT22].

above XOR lifting result for the k-fold Forrelation problem, it suffices to prove the following Fourier growth bounds for XOR-fibers.

Conjecture 4.2.7. *Let $\mathcal{C} : \{\pm 1\}^n \times \{\pm 1\}^n \to \{\pm 1\}$ be a deterministic communication protocol with at most d bits of communication. Let h be its XOR-fiber as in Definition 4.1.1. Then for all $k \in \mathbb{N}$, we have that $L_k(h) \leq (\sqrt{d} \cdot \mathsf{poly}(k, \log(n)))^k$.*

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.

General Gadgets and Fourier Growth from Lifting

Our main results are Fourier growth bounds for XOR-fibers, which corresponds to XOR-lifts of functions. To complement this, we show that similar bounds hold for general lifted functions.

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

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

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are uniform over Σ. We use $L_k(g, d)$ to denote the upper bound of the level-k Fourier growth for the g-fibers of protocols with at most d bits of communication. Using this notation, the XOR-fiber of $\mathcal{C}$ is simply $\mathcal{C}_{\downarrow \mathrm{XOR}}$, and our main results Theorems 4.2.1 and 4.2.2 can be rephrased as

$$L_1(\mathrm{XOR}, d) \leq O\left(\sqrt{d}\right) \quad \text{and} \quad L_2(\mathrm{XOR}, d) \leq O\left(d^{3/2} \log^3(n)\right).$$

In Section 4.8, we relate $L_k(g, d)$ to $L_k(\mathrm{XOR}, d)$, and the main takeaway is, in the study of Fourier growth bounds, constant-sized gadgets are all equivalent.

Theorem 4.2.8 (Informal, see Theorem 4.8.5 and Theorem 4.8.6)**.** *Let $g \colon \Sigma \times \Sigma \to \{\pm 1\}$ be a "balanced" gadget. Then*

$$|\Sigma|^{-k} \cdot L_k(\mathrm{XOR}, d) \leq L_k(g, d) \leq |\Sigma|^k \cdot L_k(\mathrm{XOR}, d).$$

Theorem 4.2.8 also proposes a different approach towards Conjecture 4.2.7: it suffices to establish tight Fourier growth bound for g-fibers for some constant-sized (actually, poly-logarithmic 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, SSW23]. See Section 4.9 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 4.2.7 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.

Pseudorandomness for Communication Protocols

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

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

[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 4.2.7 and the PRG construction from [CHLT19, Theorem 4.5], we would obtain a completely different explicit PRG for this class with error ε and seed length $n + d \cdot \mathsf{polylog}(n/\varepsilon)$.

4.3 Proof Overview

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

- First, we show that the level-1 bound can be characterized as the expected absolute value of a martingale defined as follows: Consider the random walk induced on the protocol tree when Alice and Bob are given inputs $\boldsymbol{x}$ and $\boldsymbol{y}$ uniformly from $\{\pm1\}^n$. Let

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

$\boldsymbol{X}^{(t)} \times \boldsymbol{Y}^{(t)}$ be the rectangle associated with the random walk at time t. The martingale process tracks the inner product $\left\langle \mu(\boldsymbol{X}^{(t)}), \mu(\boldsymbol{Y}^{(t)}) \right\rangle$ where $\mu(\boldsymbol{X}^{(t)}) = \mathbb{E}\left[\boldsymbol{x} \,\middle|\, \boldsymbol{x} \in \boldsymbol{X}^{(t)}\right]$ and $\mu(\boldsymbol{Y}^{(t)}) = \mathbb{E}\left[\boldsymbol{y} \,\middle|\, \boldsymbol{y} \in \boldsymbol{Y}^{(t)}\right]$ are Alice's and Bob's center of masses.

- Second, to bound the value of the martingale, it is necessary to ensure that neither $\boldsymbol{X}^{(t)}$ nor $\boldsymbol{Y}^{(t)}$ become excessively elongated in any direction during the protocol execution. To measure the length of $\boldsymbol{X}^{(t)}$ in a particular direction $\theta \in \mathbb{S}^{n-1}$, we calculate the variance $\mathbb{V}\mathrm{ar}\left[\langle \boldsymbol{x}, \theta \rangle \,\middle|\, \boldsymbol{x} \in \boldsymbol{X}^{(t)}\right]$, i.e. the variance of a uniformly random $\boldsymbol{x} \in \boldsymbol{X}^{(t)}$ in the direction θ. If the set is not elongated in any direction, this can be thought of as a spectral notion of almost pairwise independence. Such a notion also generalizes to almost k-wise independence by considering higher moments.

  To achieve the property that the sets are not elongated, one of the main novel ideas in our paper is to modify the original protocol to a new one that incorporates additional cleanup steps where the parties communicate *real values* $\langle \boldsymbol{x}, \theta \rangle$. Through these communication steps, the sets $\boldsymbol{X}^{(t)}$ and $\boldsymbol{Y}^{(t)}$ are recursively divided into affine slices along problematic directions.

- Last, one needs to show that the number of cleanup steps are small in order to bound the value of the martingale for the new protocol. This is the most involved part of our proof and requires considerable effort because the cleanup steps are real-valued and adaptively depend on the entire history, including the previous real values communicated.

The strategy outlined above also generalizes to level-2 Fourier growth by considering higher moments and sending values of quadratic forms in the inputs. We also remark that since we view the sets $\boldsymbol{X}^{(t)}$ and $\boldsymbol{Y}^{(t)}$ above as embedded in $\mathbb{R}^n$ and allow the protocol to send real values, it is more natural for us to work in Gaussian space by doing a standard transformation. The rotational invariance of the Gaussian space also seems to be essential for us to obtain optimal level-1 bound without losing additional polylogarithmic factors.

We now elaborate on the above components in detail and also highlight the differences between the level-1 and level-2 settings.

Level-One Fourier Growth

The level-1 Fourier growth of the XOR-fiber h is given by

$$L_1(h) = \sum_{i=1}^{n} \left|\widehat{h}(\{i\})\right| = \sum_{i=1}^{n} \left| \mathop{\mathbb{E}}_{\boldsymbol{z} \sim \mathcal{U}_n} [h(\boldsymbol{z})\boldsymbol{z}_i] \right| = \sum_{i=1}^{n} \left| \mathop{\mathbb{E}}_{\boldsymbol{x},\boldsymbol{y} \sim \mathcal{U}_n} [\mathcal{C}(\boldsymbol{x},\boldsymbol{y})\boldsymbol{x}_i\boldsymbol{y}_i] \right|.$$

To bound the above, it suffices to bound $\sum_{i=1}^{n} \eta_i \cdot \mathbb{E}[\mathcal{C}(\boldsymbol{x},\boldsymbol{y})\boldsymbol{x}_i\boldsymbol{y}_i]$ for any sign vector $\eta \in \{\pm 1\}^n$. 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 $\sum_{i=1}^{n} \mathbb{E}[\mathcal{C}(\boldsymbol{x},\boldsymbol{y})\boldsymbol{x}_i\boldsymbol{y}_i]$, consider a random leaf $\boldsymbol{\ell}$ of the protocol and let $\boldsymbol{X}_{\boldsymbol{\ell}} \times \boldsymbol{Y}_{\boldsymbol{\ell}}$ be the corresponding rectangle. Since the leaf determines the answer of the protocol, denoted by $\mathcal{C}(\boldsymbol{\ell})$, the quantity above equals

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

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

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

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

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

where $\Delta\mu(\boldsymbol{X}^{(t+1)}) = \mu(\boldsymbol{X}^{(t+1)}) - \mu(\boldsymbol{X}^{(t)})$ is the change in Alice's center of mass. A similar expression holds if Bob sends a message. Then it suffices to bound the expected quadratic variation (see Chapter 2) since

$$\left(\mathbb{E}\left[\left|\boldsymbol{z}^{(d)}\right|\right]\right)^2 \leq \mathbb{E}\left[\left(\boldsymbol{z}^{(d)}\right)^2\right] = \mathbb{E}\left[\sum_{t=0}^{d-1} \left(\Delta\boldsymbol{z}^{(t+1)}\right)^2\right], \tag{4.2}$$

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

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

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

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

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

It turns out that bounding the quadratic variation of this martingale suffices to give a bound on $L_2(h)$ (see Section 4.5), 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

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

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

  Recall that the martingale $\boldsymbol{z}^{(t)}$ is the inner product of Alice and Bob's center of masses, and Bob's center of mass does not change when Alice speaks. The original communication steps now do not contribute to the martingale and only the steps where the inner products are revealed do. In particular, if $t_{\text{prev}} < t$ are two consecutive times where Alice revealed the inner product, then the change in Alice's center of mass is orthogonal to change in Bob's center of mass between time t_{prev} and t. Thus, conditioned on the rectangle $\boldsymbol{X}^{(t)} \times \boldsymbol{Y}^{(t)}$ fixed by the messages until time t, we have, by Jensen's inequality,

  $$\mathbb{E}\left[(\Delta \boldsymbol{z}^{(t+1)})^2 \,\middle|\, \boldsymbol{X}^{(t)}, \boldsymbol{Y}^{(t)}\right] = \mathbb{E}\left[\left\langle \Delta\mu(\boldsymbol{X}^{(t+1)}), \mu(\boldsymbol{Y}^{(t)}) - \mu(\boldsymbol{Y}^{(t_{\text{prev}})}) \right\rangle^2 \,\middle|\, \boldsymbol{X}^{(t)}, \boldsymbol{Y}^{(t)}\right]$$

$$\leq \mathbb{E}\left[\left\langle \boldsymbol{x} - \mu(\boldsymbol{X}^{(t)}), \mu(\boldsymbol{Y}^{(t)}) - \mu(\boldsymbol{Y}^{(t_{\text{prev}})})\right\rangle^2 \,\middle|\, \boldsymbol{X}^{(t)}, \boldsymbol{Y}^{(t)}\right]. \tag{4.3}$$

Note that the quantity on the right-hand side above is of the form $\langle \boldsymbol{x} - \mathbb{E}[\boldsymbol{x}], v\rangle$. In other words, it is the variance of the random vector $\boldsymbol{x}$ along direction v. To maintain a bound on this quantity, we introduce the notion of "not being elongated in any direction".

- **Not Elongated in any Direction.** We define the following notion to capture the fact that the random vector is not elongated in any direction: we say that a mean-zero random vector $\boldsymbol{x}' = \boldsymbol{x} - \mathbb{E}[\boldsymbol{x}]$ in $\mathbb{R}^n$ is λ-*pairwise clean*, if for every $v \in \mathbb{R}^n$,

$$\mathbb{E}\left[\langle \boldsymbol{x}', v\rangle^2\right] \leq \lambda \cdot \|v\|^2, \tag{4.4}$$

or equivalently, the operator norm of the covariance matrix $\mathbb{E}[\boldsymbol{x}'\boldsymbol{x}'^\top]$ is at most λ. This can be considered a spectral notion of almost pairwise independence, since the pairwise moments are well-behaved in every direction.

If the input distribution conditioned on Alice's set $\boldsymbol{X}^{(t)}$ is $O(1)$-pairwise clean, we say that her set is *pairwise clean*. Based on the above ideas, after Alice sends the initial message, if her set is not yet clean, she partitions it recursively by taking affine slices and transmitting real values. More precisely, while there is direction $\theta \in \mathbb{S}^{n-1}$ violating (4.4), Alice does a cleanup of her set by sending the inner product $\langle x, \theta\rangle$. This direction is known to Bob as it only depends on Alice's current space. In addition, this cleanup does not contribute to the martingale *in the future* because the inner product along this direction is fixed now.

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

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

$$\begin{aligned}\mathbb{E}\left[(\boldsymbol{z}^{(\boldsymbol{d})})^2\right] &\lesssim \mathbb{E}\left[\sum_m \left\|\mu(\boldsymbol{X}^{(\boldsymbol{\tau}_m)}) - \mu(\boldsymbol{X}^{(\boldsymbol{\tau}_{m-1})})\right\|^2 + \left\|\mu(\boldsymbol{Y}^{(\boldsymbol{\tau}'_m)}) - \mu(\boldsymbol{Y}^{(\boldsymbol{\tau}'_m-1)})\right\|^2\right] \\ &= \mathbb{E}\left[\left\|\mu(\boldsymbol{X}^{(\boldsymbol{d})})\right\|^2 + \left\|\mu(\boldsymbol{Y}^{(\boldsymbol{d})})\right\|^2\right],\end{aligned} \tag{4.5}$$

where again, the last equality follows from the martingale property. The right hand side above can be bounded by the expected number of communication rounds $\mathbb{E}[\boldsymbol{d}]$ using the level-1 inequality (see Theorem 2.0.6) — this inequality bounds the Euclidean norm of the center of mass of a set in terms of its Gaussian measure.

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

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 $\boldsymbol{x}$ sampled from X, there are at most $m \lesssim d$ orthogonal directions $\theta_1, \ldots, \theta_m$ such that $\mathbb{E}[\langle \boldsymbol{x}', \theta_i \rangle^2] \gtrsim 1$ where $\boldsymbol{x}' = \boldsymbol{x} - \mathbb{E}[\boldsymbol{x}]$. This is a consequence of the fact that the expectation of $\boldsymbol{q} = \sum_{i=1}^{m} \langle \boldsymbol{x}', \theta_i \rangle^2$ can be bounded by $O(d)$ provided that X has measure $\approx 2^{-d}$.

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

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

We make two remarks in passing: First, we can also prove the optimal level-1 bound using information-theoretic ideas but they do not seem to generalize to the level-2 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 $\sqrt{d} \cdot \mathsf{polylog}(n)$ for the level one while directly working with the uniform distribution on the hypercube.

Level-Two Fourier Growth

We start by noting that the level-2 Fourier growth of the XOR-fiber h is given by

$$L_2(h) = \sum_{i \neq j} \left| \widehat{h}(\{i,j\}) \right| = \sum_{i \neq j} \left| \mathop{\mathbb{E}}_{\boldsymbol{z} \sim \mathcal{U}_n} [h(\boldsymbol{z}) \boldsymbol{z}_i \boldsymbol{z}_j] \right| = \sum_{i \neq j} \left| \mathop{\mathbb{E}}_{\boldsymbol{x}, \boldsymbol{y} \sim \mathcal{U}_n} [\mathcal{C}(\boldsymbol{x}, \boldsymbol{y}) \boldsymbol{x}_i \boldsymbol{x}_j \boldsymbol{y}_i \boldsymbol{y}_j] \right|.$$

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

Martingales and Gram-Schmidt Orthogonalization. Similar to the case of level one, the level-2 Fourier growth also has a martingale formulation. In particular, let $\boldsymbol{X}^{(t)}$ and $\boldsymbol{Y}^{(t)}$ be Alice and Bob's sets at time t as before and define

$$\sigma(\boldsymbol{X}^{(t)}) = \mathbb{E}\left[\boldsymbol{x} \dot{\otimes} \boldsymbol{x} \,\middle|\, \boldsymbol{x} \in \boldsymbol{X}^{(t)}\right], \sigma(\boldsymbol{Y}^{(t)}) = \mathbb{E}\left[\boldsymbol{y} \dot{\otimes} \boldsymbol{y} \,\middle|\, \boldsymbol{y} \in \boldsymbol{Y}^{(t)}\right]$$

to be the $n \times n$ matrices that represent the *level-2 center of masses* of the two sets. Here $\boldsymbol{x} \dot{\otimes} \boldsymbol{y}$ denotes the tensor product $\boldsymbol{x} \otimes \boldsymbol{y}$ with the diagonal zeroed out.[7] To bound the level-2 Fourier growth, it suffices to bound the expected quadratic variation of the martingale $\left(\boldsymbol{z}^{(t)}\right)_t$ defined by taking the inner product of the level-2 center of masses $\boldsymbol{z}^{(t)} := \left\langle \sigma(\boldsymbol{X}^{(t)}), \sigma(\boldsymbol{Y}^{(t)}) \right\rangle$ where $\langle \cdot, \cdot \rangle$ is the inner product of two matrices viewed as vectors.

To this end, we again move to Gaussian space where the inputs $x, y \in \mathbb{R}^n$ and transform the protocol to a clean protocol. First, we need an analog of the *Gram-Schmidt orthogonalization* step — this is achieved in a natural way by Alice sending inner product $\left\langle x \dot{\otimes} x, \sigma(\boldsymbol{Y}^{(t)}) \right\rangle$ with Bob's level-2 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 $\boldsymbol{x}$ is 4-wise clean with parameter λ if the operator norm of the $n^2 \times n^2$ covariance matrix

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

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

$$\mathbb{E}\left[\left\langle M, \boldsymbol{x} \dot{\otimes} \boldsymbol{x} - \mathbb{E}\left[\boldsymbol{x} \dot{\otimes} \boldsymbol{x}\right] \right\rangle^2\right] \leq \lambda \left\|M\right\|_{\mathrm{F}}^2, \tag{4.6}$$

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

We say a set is 4-*wise clean* with parameter λ if (4.6) is preserved for all M with zero diagonal[8]. Combined with this notion, one can define the cleanup in an analogous way to the level-1 cleanup: While there exists some $M \in \mathbb{R}^{n \times n}$ violating (4.6), Alice sends the quadratic form $\left\langle x \dot{\otimes} x, M \right\rangle$ to Bob until her set is 4-wise clean with parameter λ.

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

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

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-1 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 $\lambda \leq d \cdot \mathsf{polylog}(n)$. Analogous to (4.2) and (4.5), this gives a bound of $d^3 \cdot \mathsf{polylog}(n)$ on the expected quadratic variation and implies $L_2(h) \leq d^{3/2} \cdot \mathsf{polylog}(n)$.

Since the parties send values of quadratic forms now, the analysis here is significantly more involved compared to the level-1 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 [Tal20]. This inequality gives a tail bound for sum of squares of quadratic forms: In particular if $M_1, \ldots, M_m$ are matrices with zero diagonal which form an orthonormal set when viewed as n^2 dimensional vectors, then the random variable $\boldsymbol{q} = \sum_{i=1}^m \left\langle \boldsymbol{x} \dot{\otimes} \boldsymbol{x}, M_i \right\rangle^2$ satisfies $\mathbf{Pr}_{\boldsymbol{x} \sim \gamma_n}[\boldsymbol{q} \geq t] \leq e^{-\Omega(\sqrt{t})}$ for any $t \gtrsim m^2$ (see Theorem 4.4.1 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 $\left\langle \boldsymbol{x} \dot{\otimes} \boldsymbol{x}, M \right\rangle$ conditioned on the history is hard to determine, unlike the level-1 case. To handle this, we replace the real communication with Boolean communication of finite precision $\pm 1/\mathsf{poly}(n)$. This means that whenever Alice wants to perform cleanup $\langle \boldsymbol{x} \otimes \boldsymbol{x}, M \rangle$ for some M known to both parties, she sends only $O(\log(n))$ bits. On the one hand, this modification is similar enough to the cleanup protocol with real messages so that most of the argument carries through. On the other hand, now the protocol is completely discrete, which allows us to condition on any particular transcript.

For intuition, assume we fix a transcript of $L = d + O(m \log(n))$ bits which has gone through m cleanups. Typically, this transcript should capture $\approx 2^{-L}$ of the probability mass. More crucially, the matrices $M_1, \ldots, M_m$ for the cleanups are also fixed along the transcript, and one can apply the aforementioned Hanson-Wright inequality on $\boldsymbol{q} = \sum_{i=1}^m \left\langle \boldsymbol{x} \dot{\otimes} \boldsymbol{x}, M_i \right\rangle^2$. Combining the two facts together, we can apply the non-adaptive tail bound above and then condition on obtaining such typical transcript. This shows $\mathbb{E}[\boldsymbol{q}] \leq d^2 \cdot \mathsf{polylog}(n)$. However, each quadratic form comes from a violation of (4.6) and contributes at least λ to $\boldsymbol{q}$ in expectation. This implies that $\mathbb{E}[\boldsymbol{q}] \geq \lambda \cdot m$ and by taking $\lambda = d \cdot \mathsf{polylog}(n)$, we derive that the number of cleanup steps $m \lesssim d$. This shows that the level-2 Fourier growth is $O((m+d) \cdot \sqrt{\lambda}) = d^{3/2} \cdot \mathsf{polylog}(n)$ completing the proof.

Note that if we could take $\lambda = \mathsf{polylog}(n)$ while having the same number of cleanup steps $m = d \cdot \mathsf{polylog}(n)$, then we would obtain an optimal level-2 bound of $d \cdot \mathsf{polylog}(n)$. However, it is not clear how to use current approach to show this. In Section 4.9, 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-2 bound can be further improved, we do not make the effort of generalizing it to higher levels here.

4.4 Concentration for Sum of Squares of Quadratics

Recall that 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 $\left(x \dot{\otimes} x\right)_{ij} = x_i x_j$ if $i \neq j$ and zero if $i = j$.

We will 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.

Theorem 4.4.1. *Let $m \in \mathbb{N}$ be arbitrary. Let $M_1, \ldots, M_m$ 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*

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

Proof. 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

$$K_x = \left\| M(x \dot{\otimes} x) \right\|^2 = \left\| M(x \otimes x) \right\|^2, \tag{4.7}$$

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

Define $f(y) = \|M(y \otimes y)\|$, $g(y) = \sup_{z \in \mathbb{S}^{n-1}} \|M(z \otimes y)\|$, $h(y) = \sup_{z \in \mathbb{S}^{n-1}} \|M(y \otimes z)\|$. Let $F = \mathbb{E}_{\boldsymbol{y} \sim \gamma_n}[f(\boldsymbol{y})]$, $G = \mathbb{E}_{\boldsymbol{y} \sim \gamma_n}[g(\boldsymbol{y})]$, and $H = \mathbb{E}_{\boldsymbol{y} \sim \gamma_n}[h(\boldsymbol{y})]$ 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 (2.1), we have

$$\gamma_n(A + t\mathcal{B}^n) \geq 1 - e^{-t^2/2} \quad \text{holds for all } t \geq 0. \tag{4.8}$$

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

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

where $V = \sup_{z \in \mathbb{S}^{n-1}} \|M(z \otimes z)\|$. This, together with (4.7) and (4.8), implies

$$\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \gamma_n}\left[K_{\boldsymbol{x}} \geq \left(6F + 6t(G + H) + t^2 V\right)^2\right] \leq \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \gamma_n}\left[\boldsymbol{x} \notin A + t\mathcal{B}^n\right] = 1 - \gamma_n(A + t\mathcal{B}^n) \leq e^{-t^2/2}. \tag{4.9}$$

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

Claim 4.4.2. *$F \leq \sqrt{2m}$, $G, H \leq \sqrt{m}$, and $V \leq 1$.*

Plugging Claim 4.4.2 into (4.9), we have

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

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

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

Finally we present the missing proof of Claim 4.4.2.

Proof of Claim 4.4.2. First we observe that rows of M are unit vectors, therefore

$$\|M\|_{\mathrm{F}} = \sqrt{m}. \tag{4.10}$$

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

$$\|M\|_{\mathrm{op}} \leq 1. \tag{4.11}$$

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

$$b_{i,i} = 0^m \quad \text{for all } i \in [n]. \tag{4.12}$$

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

Bounding F. Observe that

$$F^2 = \left(\mathop{\mathbb{E}}_{\boldsymbol{y}\sim\gamma_n}\left[\|M(\boldsymbol{y}\otimes\boldsymbol{y})\|\right]\right)^2 \le \mathop{\mathbb{E}}_{\boldsymbol{y}\sim\gamma_n}\left[\|M(\boldsymbol{y}\otimes\boldsymbol{y})\|^2\right] = \mathop{\mathbb{E}}_{\boldsymbol{y}\sim\gamma_n}\left[\left\|\sum_{i,j\in[n]} b_{i,j}\boldsymbol{y}_i\boldsymbol{y}_j\right\|^2\right] \quad \text{(by convexity)}$$

$$= \mathop{\mathbb{E}}_{\boldsymbol{y}\sim\gamma_n}\left[\sum_{i,j,i',j'\in[n]} \langle b_{i,j}, b_{i',j'}\rangle\, \boldsymbol{y}_i\boldsymbol{y}_j\boldsymbol{y}_{i'}\boldsymbol{y}_{j'}\right] = \sum_{i,j\in[n]}\left(\|b_{i,j}\|^2 + \langle b_{i,j}, b_{j,i}\rangle\right) \quad \text{(by (4.12))}$$

$$\le \sum_{i,j\in[n]}\left(\|b_{i,j}\|^2 + \frac{1}{2}\left(\|b_{i,j}\|^2 + \|b_{j,i}\|^2\right)\right) = 2\sum_{i,j\in[n]}\|b_{i,j}\|^2$$

$$= 2\|M\|_{\mathrm{F}}^2 = 2m. \quad \text{(by (4.10))}$$

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

$$g(y) = \sup_{z\in\mathbb{S}^{n-1}}\left\|\sum_{i,j\in[n]} b_{i,j}z_iy_j\right\| = \sup_{z\in\mathbb{S}^{n-1}}\left\|\sum_{i\in[n]} b_iz_i\right\| = \|B\|_{\mathrm{op}} \le \|B\|_{\mathrm{F}} = \sqrt{\sum_{i\in[n]}\left\|\sum_{j\in[n]} b_{i,j}y_j\right\|^2}. \tag{4.13}$$

Now we bound G:

$$G^2 = \left(\mathop{\mathbb{E}}_{\boldsymbol{y}\sim\gamma_n}[g(\boldsymbol{y})]\right)^2 \le \mathop{\mathbb{E}}_{\boldsymbol{y}\sim\gamma_n}\left[g(\boldsymbol{y})^2\right] \quad \text{(by convexity)}$$

$$\le \mathop{\mathbb{E}}_{\boldsymbol{y}\sim\gamma_n}\left[\sum_{i\in[n]}\left\|\sum_{j\in[n]} b_{i,j}\boldsymbol{y}_j\right\|^2\right] = \mathop{\mathbb{E}}_{\boldsymbol{y}\sim\gamma_n}\left[\sum_{i\in[n]}\sum_{j,j'\in[n]}\langle b_{i,j}, b_{i,j'}\rangle\, \boldsymbol{y}_j\boldsymbol{y}_{j'}\right] \quad \text{(by (4.13))}$$

$$= \sum_{i,j\in[n]}\|b_{i,j}\|^2 = \|M\|_{\mathrm{F}}^2 = m. \quad \text{(by (4.10))}$$

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 (4.11), we have

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

4.5 Fourier Growth via Martingales in Gaussian Space

In this section, we reduce the question of bounding the level-1 and level-2 Fourier growth to bounding the expected quadratic variation of certain martingales. To analyze these martingales and to prove the optimal bound for the level-1 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.

Communication Protocols in Gaussian Space

Let $\mathcal{C} : \{\pm 1\}^n \times \{\pm 1\}^n \to \{\pm 1\}$ be a communication protocol with total communication d and h be its XOR-fiber defined in Definition 4.1.1.

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 $\widetilde{\mathcal{C}}$ runs the original protocol $\mathcal{C}$ with Boolean inputs $\mathsf{sgn}(x)$ and $\mathsf{sgn}(y)$ where $\mathsf{sgn}(v) = (\mathsf{sgn}(v_1), \ldots, \mathsf{sgn}(v_n))$ denotes the sign function applied pointwise to each coordinate for a vector $v \in \mathbb{R}^n$. The behavior of the communication protocol $\widetilde{\mathcal{C}}$ can be defined arbitrarily if any coordinate of $\mathsf{sgn}(x)$ or $\mathsf{sgn}(y)$ 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 $\mathcal{U}_n(S) = \gamma_n(\{x \in \mathbb{R}^n \mid \mathsf{sgn}(x) \in S\})$. 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 4.5.1. *For all $S \subseteq [n]$, we have $\mathbb{E}_{\boldsymbol{z} \sim \mathcal{U}_n}[h(\boldsymbol{z})\boldsymbol{z}_S] = (\pi/2)^{|S|}\, \mathbb{E}_{\boldsymbol{x},\boldsymbol{y} \sim \gamma_n}\left[\widetilde{\mathcal{C}}(\boldsymbol{x}, \boldsymbol{y})\boldsymbol{x}_S \boldsymbol{y}_S\right]$.*

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

$$\begin{aligned}
\mathop{\mathbb{E}}_{\boldsymbol{z} \sim \mathcal{U}_n}[h(\boldsymbol{z})\boldsymbol{z}_S] &= \mathop{\mathbb{E}}_{\boldsymbol{x},\boldsymbol{y} \sim \gamma_n}\left[\mathcal{C}(\mathsf{sgn}(\boldsymbol{x}), \mathsf{sgn}(\boldsymbol{y})) \cdot \prod_{i \in S} \mathsf{sgn}(\boldsymbol{x}_i) \cdot \mathsf{sgn}(\boldsymbol{y}_i)\right] \\
&= (\pi/2)^{|S|} \mathop{\mathbb{E}}_{\boldsymbol{x},\boldsymbol{y} \sim \gamma_n}\left[\mathcal{C}(\mathsf{sgn}(\boldsymbol{x}), \mathsf{sgn}(\boldsymbol{y})) \cdot \prod_{i \in S} \boldsymbol{x}_i \cdot \boldsymbol{y}_i\right] \\
&= (\pi/2)^{|S|} \mathop{\mathbb{E}}_{\boldsymbol{x},\boldsymbol{y} \sim \gamma_n}\left[\widetilde{\mathcal{C}}(\boldsymbol{x}, \boldsymbol{y})\boldsymbol{x}_S \boldsymbol{y}_S\right],
\end{aligned}$$

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:

$$\mathop{\mathbb{E}}_{\boldsymbol{x}_i \sim \gamma} [\boldsymbol{x}_i \,|\, \mathsf{sgn}(\boldsymbol{x}_i) = \eta] = \eta \cdot \int_0^\infty \sqrt{\frac{2}{\pi}} \cdot r \cdot e^{-r^2/2} \,\mathrm{d}\, r = \sqrt{\frac{2}{\pi}} \cdot \eta. \qquad \square$$

Remark 4.5.2. *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 $\boldsymbol{x}, \boldsymbol{y}$ are independently drawn vectors in $\mathbb{R}^n$ where each coordinate is i.i.d. sampled from ξ, then $\mathbb{E}_{\boldsymbol{z} \sim \mathcal{U}_n} [h(\boldsymbol{z}) \boldsymbol{z}_S] = c_\xi^{|S|} \mathbb{E}_{\boldsymbol{x}, \boldsymbol{y} \sim \xi^{\otimes n}} \left[\widetilde{\mathcal{C}}(\boldsymbol{x}, \boldsymbol{y}) \boldsymbol{x}_S \boldsymbol{y}_S \right]$ where $c_\xi = (\mathbb{E}_{\boldsymbol{x}_i \sim \xi}[|\boldsymbol{x}_i|])^{-2}$. In the case of level-2 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.*

Generalized Communication Protocols

In the protocol $\widetilde{\mathcal{C}}$ 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 $\mathsf{sgn}(x)$ and $\mathsf{sgn}(y)$. 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 $\mathsf{sgn}(z)$ 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 $\overline{\mathcal{C}}$ is equivalent to the protocol $\widetilde{\mathcal{C}}$ if $\overline{\mathcal{C}}(x, y) = \widetilde{\mathcal{C}}(x, y)$ 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 $\boldsymbol{x}$ and $\boldsymbol{y}$ 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 $\boldsymbol{x}, \boldsymbol{y}$ is some product measure on $\xi_x^{(t)} \times \xi_y^{(t)}$ supported on a rectangle $\boldsymbol{X}^{(t)} \times \boldsymbol{Y}^{(t)}$ where $\boldsymbol{X}^{(t)}, \boldsymbol{Y}^{(t)} \subseteq \mathbb{R}^n$. We shall refer to the random variable $\boldsymbol{X}^{(t)} \times \boldsymbol{Y}^{(t)}$ as the current rectangle determined by $\mathcal{F}^{(t)}$. Since we will be working with product measures on inputs $\boldsymbol{x}, \boldsymbol{y}$, the reader can think of conditioning on the filtration $\mathcal{F}^{(t)}$ as essentially conditioning on the inputs being in the rectangle $\boldsymbol{X}^{(t)} \times \boldsymbol{Y}^{(t)}$ or equivalently a partial transcript till time t.

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 4.5.1, the level-k Fourier growth of the XOR-fiber h of the original communication protocol is given by

$$\begin{aligned} L_k(h) = \sum_{\substack{S \subseteq [n] \\ |S|=k}} \left| \mathop{\mathbb{E}}_{\boldsymbol{z} \sim \mathcal{U}_n} [h(\boldsymbol{z}) \boldsymbol{z}_S] \right| &= (\pi/2)^k \sum_{\substack{S \subseteq [n] \\ |S|=k}} \left| \mathop{\mathbb{E}}_{\boldsymbol{x}, \boldsymbol{y} \sim \gamma_n} [\overline{\mathcal{C}}(\boldsymbol{x}, \boldsymbol{y}) \boldsymbol{x}_S \boldsymbol{y}_S] \right| \\ &= (\pi/2)^k \max_{(\eta_S)_{|S|=k}} \sum_{\substack{S \subseteq [n] \\ |S|=k}} \eta_S \mathop{\mathbb{E}}_{\boldsymbol{x}, \boldsymbol{y} \sim \gamma_n} \left[\overline{\mathcal{C}}(\boldsymbol{x}, \boldsymbol{y}) \boldsymbol{x}_S \boldsymbol{y}_S \right], \end{aligned} \tag{4.14}$$

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

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

$$\sum_{S \subseteq [n], |S|=k} \eta_S \mathop{\mathbb{E}}_{\boldsymbol{x}, \boldsymbol{y} \sim \gamma_n} \left[\overline{\mathcal{C}}(\boldsymbol{x}, \boldsymbol{y}) \boldsymbol{x}_S \boldsymbol{y}_S \right] \tag{4.15}$$

$$= \mathop{\mathbb{E}}_{\boldsymbol{\ell}} \left[\mathop{\mathbb{E}}_{\boldsymbol{x}, \boldsymbol{y} \sim \gamma} \left[\sum_{S \subseteq [n], |S|=k} \eta_S \cdot \overline{\mathcal{C}}(\boldsymbol{x}, \boldsymbol{y}) \boldsymbol{x}_S \boldsymbol{y}_S \,\middle|\, (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{X}_{\boldsymbol{\ell}} \times \boldsymbol{Y}_{\boldsymbol{\ell}} \right] \right]$$

$$= \mathop{\mathbb{E}}_{\boldsymbol{\ell}} \left[\overline{\mathcal{C}}(\boldsymbol{\ell}) \mathop{\mathbb{E}}_{\boldsymbol{x}, \boldsymbol{y} \sim \gamma} \left[\sum_{S \subseteq [n], |S|=k} \eta_S \cdot \boldsymbol{x}_S \boldsymbol{y}_S \,\middle|\, (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{X}_{\boldsymbol{\ell}} \times \boldsymbol{Y}_{\boldsymbol{\ell}} \right] \right]$$

$$\leq \mathop{\mathbb{E}}_{\boldsymbol{\ell}} \left[\left| \sum_{S \subseteq [n], |S|=k} \eta_S \, \mathbb{E} \left[\boldsymbol{x}_S \,|\, \boldsymbol{x} \in \boldsymbol{X}_{\boldsymbol{\ell}} \right] \cdot \mathbb{E} \left[\boldsymbol{y}_S \,|\, \boldsymbol{y} \in \boldsymbol{Y}_{\boldsymbol{\ell}} \right] \right| \right], \tag{4.16}$$

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

Thus, specializing (4.15) to the level-1 ($k = 1$) and level-2 cases ($k = 2$), from (4.14) we get that

$$L_1(h) \le \frac{\pi}{2} \cdot \max_{\eta} \mathop{\mathbb{E}}_{\boldsymbol{\ell}} \left[\left| \sum_{i=1}^{n} \eta_i \cdot \mathbb{E}\left[\boldsymbol{x}_i \,|\, \boldsymbol{x} \in \boldsymbol{X}_{\boldsymbol{\ell}}\right] \cdot \mathbb{E}\left[\boldsymbol{y}_i \,|\, \boldsymbol{y} \in \boldsymbol{Y}_{\boldsymbol{\ell}}\right] \right| \right],$$
$$L_2(h) \le \frac{\pi^2}{4} \cdot \max_{\eta} \mathop{\mathbb{E}}_{\boldsymbol{\ell}} \left[\left| \sum_{i,j=1}^{n} \eta_{ij} \cdot \ \mathbb{E}\left[\boldsymbol{x}_{ij} \,|\, \boldsymbol{x} \in \boldsymbol{X}_{\boldsymbol{\ell}}\right] \cdot \mathbb{E}\left[\boldsymbol{y}_{ij} \,|\, \boldsymbol{y} \in \boldsymbol{Y}_{\boldsymbol{\ell}}\right] \right| \right],$$

where for L_1 we optimize over $\eta \in \{\pm 1\}^n$ and for L_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-1 and level-2 centers of mass of a set $X \subseteq \mathbb{R}^n$:

$$\mu(X) = \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \gamma_n} \left[\boldsymbol{x} \,|\, \boldsymbol{x} \in X\right] \quad \text{and} \quad \sigma(X) = \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \gamma_n} \left[\boldsymbol{x} \mathbin{\dot{\otimes}} \boldsymbol{x} \,\middle|\, \boldsymbol{x} \in X\right]. \tag{4.17}$$

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

$$\begin{aligned} L_1(h) &\le 4 \cdot \max_{\eta} \mathop{\mathbb{E}}_{\boldsymbol{\ell}} \left[\left| \left\langle \mu(\boldsymbol{X}_{\boldsymbol{\ell}}), \eta \odot \mu(\boldsymbol{Y}_{\boldsymbol{\ell}}) \right\rangle \right| \right], \\ L_2(h) &\le 4 \cdot \max_{\eta} \mathop{\mathbb{E}}_{\boldsymbol{\ell}} \left[\left| \left\langle \sigma(\boldsymbol{X}_{\boldsymbol{\ell}}), \eta \odot \sigma(\boldsymbol{Y}_{\boldsymbol{\ell}}) \right\rangle \right| \right], \end{aligned} \tag{4.18}$$

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 $\left(\boldsymbol{z}_k^{(t)}\right)_t$ 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 $\boldsymbol{X}^{(t)} \times \boldsymbol{Y}^{(t)}$. Then, we have the following proposition.

Proposition 4.5.3. *$\mu(\boldsymbol{X}^{(t)})$ and $\mu(\boldsymbol{Y}^{(t)})$ are vector-valued martingales taking values in $\mathbb{R}^n$ and $\sigma(\boldsymbol{X}^{(t)})$ and $\sigma(\boldsymbol{Y}^{(t)})$ are matrix-valued martingales taking values in $\mathbb{R}^{n \times n}$.*

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

$$\boldsymbol{z}_1^{(t)} = \left\langle \mu(\boldsymbol{X}^{(t)}), \eta \odot \mu(\boldsymbol{Y}^{(t)}) \right\rangle \text{ and } \boldsymbol{z}_2^{(t)} = \left\langle \sigma(\boldsymbol{X}^{(t)}), \eta \odot \sigma(\boldsymbol{Y}^{(t)}) \right\rangle, \tag{4.19}$$

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 $\boldsymbol{\ell}$ given in (4.18). Let $\boldsymbol{d}$ 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 $\Delta \boldsymbol{z}_k^{(t)} = \boldsymbol{z}_k^{(t)} - \boldsymbol{z}_k^{(t-1)}$ from (2.2), 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 4.5.4. *For $k \in \{1,2\}$, we have*

$$\frac{1}{4} \cdot L_k(h) \le \max_{\eta} \sqrt{\mathbb{E}\left[\left(\boldsymbol{z}_k^{(\boldsymbol{d})}\right)^2\right]} = \max_{\eta} \sqrt{\mathbb{E}\left[\sum_{t=1}^{\boldsymbol{d}} \left(\Delta \boldsymbol{z}_k^{(t)}\right)^2\right]}.$$

The martingale implicitly depends on η as used in (4.18) and hence the maximum. Moreover, the martingale also depends on the underlying generalized communication protocol $\overline{\mathcal{C}}$. In the next two sections, we will show that after transforming the original communication protocol into "clean" protocols, the expected quadratic variations of $(\boldsymbol{z}_1^{(t)})_t$ and $(\boldsymbol{z}_2^{(t)})_t$ are $O(d)$ and $O(d^3) \cdot \mathsf{polylog}(n)$ respectively. This will then imply our main theorems.

Remark 4.5.5. *Note that Proposition 4.5.3 still holds even if the input distribution is not the Gaussian distribution, but some other product probability measure on the inputs $\boldsymbol{x}, \boldsymbol{y}$. This also implies that $\boldsymbol{z}_k^{(t)}$ for $k \in \{1,2\}$ is a martingale. In particular, for the level-2 case, we will need to use a truncated Gaussian distribution. In light of Remark 4.5.2, Proposition 4.5.4 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-2 case to a finite precision, so the generalized protocols for the level-2 case only have Boolean communication. However, to obtain the optimal level-1 bound allowing generalized protocols that communicate real values seems to be crucial.*

4.6 Level-One Fourier Growth

In this section, we will give a proof of Theorem 4.2.1 that $L_1(h) = O(\sqrt{d})$. We start with a d-round communication protocol $\widetilde{\mathcal{C}}$ over the Gaussian space as defined in Section 4.5. Given the discussion in the previous section and Proposition 4.5.4, our task ultimately reduces to bounding the expected quadratic variation of the martingale that results from the protocol $\overline{\mathcal{C}}$. For example, one can simply take $\overline{\mathcal{C}} = \widetilde{\mathcal{C}}$, but, as discussed in Section 4.3, 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 $\overline{\mathcal{C}}$ that is equivalent to the original protocol $\widetilde{\mathcal{C}}$ 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.

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

$$\mathop{\mathbb{E}}_{\boldsymbol{x} \sim \gamma} \left[\langle \boldsymbol{x} - \mu(X), a\rangle^2 \,\middle|\, \boldsymbol{x} \in X\right] \le \lambda, \tag{4.20}$$

where we recall that $\mu(X) = \mathbb{E}_{\boldsymbol{x}\sim\gamma}[\boldsymbol{x} \mid \boldsymbol{x} \in X]$ is the level-1 center of mass of X.

The above condition implies that for a random vector $\boldsymbol{x}$ 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 $\boldsymbol{x}$ 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 $\widetilde{\mathcal{C}}$ in the Gaussian space, we will transform it into a pairwise clean protocol $\overline{\mathcal{C}}$ 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 $\widetilde{\mathcal{C}}$. Before Alice sends her bit as in the original protocol $\widetilde{\mathcal{C}}$, she first performs an orthogonalization step by revealing the inner-product between her input and Bob's current level-1 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 $\overline{\mathcal{C}}$ from $\widetilde{\mathcal{C}}$. 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 $\overline{\mathcal{C}}$ at depth t. We use the random variable $\boldsymbol{X}^{(t)} \subseteq \mathbb{R}^n$ (resp., $\boldsymbol{Y}^{(t)} \subseteq \mathbb{R}^n$) to denote the set of inputs of Alice (resp., Bob) reaching the node. If Alice reveals a linear form in this step, we use $\boldsymbol{a}^{(t)} \in \mathbb{R}^n$ to denote the vector of the linear form; otherwise, we set $\boldsymbol{a}^{(t)}$ to be the all-zeroes vector. We define $\boldsymbol{b}^{(t)}$ similarly for Bob. Throughout the protocol, we will abbreviate $\boldsymbol{u}^{(t)} = \mu(\boldsymbol{X}^{(t)})$ and $\boldsymbol{v}^{(t)} = \mu(\boldsymbol{Y}^{(t)})$ 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 $t \leftarrow 0$, $\boldsymbol{X}^{(0)}, \boldsymbol{Y}^{(0)} \leftarrow \mathbb{R}^n$, and $\boldsymbol{a}^{(0)}, \boldsymbol{b}^{(0)} \leftarrow 0^n$.
3. For each phase $i = 1, 2, \ldots, d$: suppose we are starting the cleanup for a node at depth i in the original protocol $\widetilde{\mathcal{C}}$ and suppose we are at a node of depth t in the new protocol $\overline{\mathcal{C}}$. If it is Alice's turn to speak in $\widetilde{\mathcal{C}}$:

   a) **Orthogonalization by Revealing 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 $\eta \odot \boldsymbol{v}^{(t)}$. 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 $\eta \odot \boldsymbol{v}^{(t)}$ that is orthogonal to the previous directions along which Alice announced the inner product. More formally, let $\boldsymbol{a}^{(t+1)}$ be the component of $\eta \odot \boldsymbol{v}^{(t)}$ that is orthonormal to all previous directions $\boldsymbol{a}^{(1)}, \ldots, \boldsymbol{a}^{(t)}$, i.e.,

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

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

b) **Original Communication.** Alice sends the bit $\overline{\boldsymbol{c}}^{(t+1)}$ that she was supposed to send in $\widetilde{\mathcal{C}}$ based on previous messages and the input x. Set $\boldsymbol{a}^{(t+1)}, \boldsymbol{b}^{(t+1)} \leftarrow 0^n$. 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 $\left\langle a, \boldsymbol{a}^{(\tau)} \right\rangle = 0$ for all $\tau \in [t]$) such that $\boldsymbol{X}^{(t)}$ is *not pairwise clean* in direction a, Alice computes $\overline{\boldsymbol{c}}^{(t+1)} \leftarrow \langle x, a \rangle$ and sends this to Bob. Set $\boldsymbol{a}^{(t+1)} \leftarrow a$ and $\boldsymbol{b}^{(t+1)} \leftarrow 0^n$. Increment t by 1. Repeat step (c) as long as $\boldsymbol{X}^{(t)}$ 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 $x, \boldsymbol{a}, \boldsymbol{X}, \boldsymbol{v}$ switched with $y, \boldsymbol{b}, \boldsymbol{Y}, \boldsymbol{U}$.

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

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 $\boldsymbol{X}^{(t)} \times \boldsymbol{Y}^{(t)}$ is pairwise clean for both parties by construction. Also, note that the non-zero vectors in the sequence $(\boldsymbol{a}^{(t)})_t$ (resp., $(\boldsymbol{b}^{(t)})_t$) 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 4.6.1. *Let ℓ be an arbitrary leaf of the protocol $\overline{\mathcal{C}}$ 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 $\boldsymbol{a}^{(t)}$'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 $\overline{\mathcal{C}}$ is $O(d)$ when $\boldsymbol{x}, \boldsymbol{y} \sim \gamma_n$.

Bounding the Expected Quadratic Variation

Consider a random walk on the protocol tree generated by the new protocol $\overline{\mathcal{C}}$ when the parties are given independent inputs $\boldsymbol{x}, \boldsymbol{y} \sim \gamma_n$. Consider the corresponding level-1 martingale process defined in (4.19). Formally, at time t the process is defined by

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

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

The martingale process stops once it hits a leaf of the protocol $\overline{\mathcal{C}}$. Let $\boldsymbol{d}$ denote the (stopping) time when this happens. Note that $\mathbb{E}[\boldsymbol{d}]$ is exactly the expected depth of the protocol $\overline{\mathcal{C}}$. Then, in light of Proposition 4.5.4, to prove Theorem 4.2.1, it suffices to prove the following.

Lemma 4.6.2. $\mathbb{E}\left[\sum_{t=1}^{\boldsymbol{d}} \left(\Delta \boldsymbol{z}_1^{(t)}\right)^2\right] = O(d)$.

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 $\eta \odot \boldsymbol{v}^{(t)}$ so they do not change the value of the martingale $\boldsymbol{z}_1^{(t)}$ as discussed in Section 4.3. Moreover, recalling (4.3), since Alice's node was pairwise clean just before Alice sent the message in step 3(a), the expected change $\mathbb{E}\left[\left(\Delta \boldsymbol{z}_1^{(t+1)}\right)^2\right]$ can be bounded in terms of the squared norm of the change that occurred in $\boldsymbol{u}^{(t)}$ 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 $\boldsymbol{z}_1^{(t)}$ is a Doob martingale and also $\boldsymbol{u}^{(t)}, \boldsymbol{v}^{(t)}$ form vector-valued martingales (recall Proposition 4.5.3). Note that $\mathcal{F}^{(t)}$ fixes all the rectangles encountered during times $0, \ldots, t$ and thus for $\tau \le t$, the random variables $\boldsymbol{u}^{(\tau)}, \boldsymbol{v}^{(\tau)}, \boldsymbol{z}_1^{(\tau)}$ 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 4.6.3 (Step Size)**.** *Let $0 = \boldsymbol{\tau}_1 < \boldsymbol{\tau}_2 < \cdots \le \boldsymbol{d}$ be a sequence of stopping times with $\boldsymbol{\tau}_m$ being the index of the round where Alice speaks in step 3(a) for the m^{th} time or $\boldsymbol{d}$ if there is no such round. Then, for any integer $m \ge 2$,*

$$\mathbb{E}\left[\left(\Delta \boldsymbol{z}_1^{(\boldsymbol{\tau}_m+1)}\right)^2 \;\middle|\; \mathcal{F}^{(\boldsymbol{\tau}_m)}\right] \le \lambda \cdot \left\|\boldsymbol{v}^{(\boldsymbol{\tau}_m)} - \boldsymbol{v}^{(\boldsymbol{\tau}_{m-1})}\right\|^2,$$

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

$$\mathbb{E}\left[\left(\Delta \boldsymbol{z}_1^{(t+1)}\right)^2 \;\middle|\; \mathcal{F}^{(t)}, \boldsymbol{\tau}_{m-1} < t < \boldsymbol{\tau}_m, \textit{Alice speaks at time } t\right] = 0.$$

A similar statement also holds if Bob speaks where $\boldsymbol{v}$ is replaced by $\boldsymbol{U}$ and the sequence $(\boldsymbol{\tau}_m)$ is replaced by $(\boldsymbol{\tau}'_m)$ where $\boldsymbol{\tau}'_m$ is the index of the round where Bob speaks in step 3(a) for the m^{th} time or $\boldsymbol{d}$ 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 $\boldsymbol{u}^{(\boldsymbol{\tau}_1)}$ and $\boldsymbol{v}^{(\boldsymbol{\tau}'_1)}$ are their initial centers of mass which are both zero.

We shall prove the above lemma later and continue with the bound on the quadratic variation here. Using Lemma 4.6.3, we have

$$\mathbb{E}\left[\sum_{t=1}^{\boldsymbol{d}} \left(\Delta \boldsymbol{z}_1^{(t)}\right)^2\right] \le \lambda \cdot \mathbb{E}\left[\sum_{m\ge 2} \left\|\boldsymbol{v}^{(\boldsymbol{\tau}_m)} - \boldsymbol{v}^{(\boldsymbol{\tau}_{m-1})}\right\|^2 + \left\|\boldsymbol{U}^{(\boldsymbol{\tau}'_m)} - \boldsymbol{U}^{(\boldsymbol{\tau}'_{m-1})}\right\|^2\right].$$

On the other hand, by the orthogonality of vector-valued martingale differences from (2.3), we have

$$\mathbb{E}\left[\sum_{m\ge 2} \left\|\boldsymbol{v}^{(\boldsymbol{\tau}_m)} - \boldsymbol{v}^{(\boldsymbol{\tau}_{m-1})}\right\|^2\right] = \mathbb{E}\left[\left\|\boldsymbol{v}^{(\boldsymbol{d})}\right\|^2\right].$$

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

$$\mathbb{E}\left[\sum_{t=1}^{\boldsymbol{d}} \left(\Delta \boldsymbol{z}_1^{(t)}\right)^2\right] \le \lambda \cdot \left(\mathbb{E}\left[\left\|\boldsymbol{U}^{(\boldsymbol{d})}\right\|_{\mathrm{F}}^2\right] + \mathbb{E}\left[\left\|\boldsymbol{v}^{(\boldsymbol{d})}\right\|_{\mathrm{F}}^2\right]\right). \tag{4.21}$$

We will soon prove the following to upper bound the quantity on the right hand side above. Loosely speaking, by an application of level-1 inequalities (see Theorem 2.0.6), the lemma below ultimately boils down to a bound on the expected number of cleanup steps.

Lemma 4.6.4 (Final Center of Mass)**.** $\mathbb{E}\left[\left\|\boldsymbol{u}^{(\boldsymbol{d})}\right\|^2 + \left\|\boldsymbol{v}^{(\boldsymbol{d})}\right\|^2\right] = O(d)$.

Since $\lambda = 100$, plugging in the bounds from the above into (4.21) readily implies Lemma 4.6.2. Together with Proposition 4.5.4, this completes the proof of Theorem 4.2.1.

Bounds on Step Sizes (Proof of Lemma 4.6.3)

Let us abbreviate $\boldsymbol{\tau} = \boldsymbol{\tau}_m$. Observe that

$$\mathbb{E}\left[\left(\Delta \boldsymbol{z}_1^{(\boldsymbol{\tau}+1)}\right)^2 \;\middle|\; \mathcal{F}^{(\boldsymbol{\tau})}\right] = \mathbb{E}\left[\left\langle \boldsymbol{U}^{(\boldsymbol{\tau}+1)} - \boldsymbol{U}^{(\boldsymbol{\tau})}, \eta \odot \boldsymbol{v}^{(\boldsymbol{\tau})}\right\rangle^2 \;\middle|\; \mathcal{F}^{(\boldsymbol{\tau})}\right]$$

$$= \mathbb{E}\left[\left\langle \boldsymbol{U}^{(\boldsymbol{\tau}+1)}, \eta \odot \boldsymbol{v}^{(\boldsymbol{\tau})}\right\rangle^2 - \left\langle \boldsymbol{U}^{(\boldsymbol{\tau})}, \eta \odot \boldsymbol{v}^{(\boldsymbol{\tau})}\right\rangle^2 \middle| \mathcal{F}^{(\boldsymbol{\tau})}\right], \tag{4.22}$$

where the second line is due to $(\boldsymbol{u}^{(t)})_t$ being a vector-valued martingale and thus

$$\mathbb{E}\left[\boldsymbol{U}^{(\boldsymbol{\tau}+1)} \middle| \mathcal{F}^{(\boldsymbol{\tau})}\right] = \boldsymbol{U}^{(\boldsymbol{\tau})}.$$

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

We now claim that components of $\boldsymbol{u}^{(\boldsymbol{\tau}+1)}$ and $\boldsymbol{u}^{(\boldsymbol{\tau})}$ are the same along any of the previous directions $\boldsymbol{a}^{(0)}, \ldots, \boldsymbol{a}^{(\boldsymbol{\tau})}$. So in (4.22), they cancel out and the only relevant quantity is the component in the direction $\boldsymbol{a}^{(\boldsymbol{\tau}+1)}$. This follows since, in all the previous steps $t \le \boldsymbol{\tau}$, Alice has already fixed $\left\langle x, \boldsymbol{a}^{(t)}\right\rangle$. This implies that for any $\boldsymbol{X}^{(\boldsymbol{\tau})}$ and $\boldsymbol{X}^{(\boldsymbol{\tau}+1)}$ that are determined by $\mathcal{F}^{(\boldsymbol{\tau}+1)}$, the inner product with all the previous $\boldsymbol{a}^{(0)}, \ldots, \boldsymbol{a}^{(\boldsymbol{\tau})}$ is fixed over the choice of x from both rectangles. Formally, we have that for any $x \in \boldsymbol{X}^{(\boldsymbol{\tau})}$ and $x' \in \boldsymbol{X}^{(\boldsymbol{\tau}+1)}$, it holds that $\left\langle x, \boldsymbol{a}^{(t)}\right\rangle = \left\langle x', \boldsymbol{a}^{(t)}\right\rangle$ for any $t \le \boldsymbol{\tau}$. In particular, since $\boldsymbol{U}^{(\boldsymbol{\tau})} = \mu(\boldsymbol{X}^{(\boldsymbol{\tau})})$ and $\boldsymbol{U}^{(\boldsymbol{\tau}+1)} = \mu(\boldsymbol{X}^{(\boldsymbol{\tau}+1)})$ are the corresponding centers of mass, we have that

$$\left\langle \boldsymbol{U}^{(\boldsymbol{\tau}+1)}, \boldsymbol{a}^{(t)}\right\rangle = \left\langle \boldsymbol{U}^{(\boldsymbol{\tau})}, \boldsymbol{a}^{(t)}\right\rangle \text{ for all } t \le \boldsymbol{\tau}. \tag{4.23}$$

This, together with (4.22) and recalling that $\boldsymbol{\beta}^{(\boldsymbol{\tau}+1)}$ is determined by $\mathcal{F}^{(\boldsymbol{\tau})}$, implies that

$$\mathbb{E}\left[\left(\Delta \boldsymbol{z}_1^{(\boldsymbol{\tau}+1)}\right)^2 \middle| \mathcal{F}^{(\boldsymbol{\tau})}\right] = \left(\boldsymbol{\beta}^{(\boldsymbol{\tau}+1)}\right)^2 \cdot \mathbb{E}\left[\left\langle \boldsymbol{U}^{(\boldsymbol{\tau}+1)}, \boldsymbol{a}^{(\boldsymbol{\tau}+1)}\right\rangle^2 - \left\langle \boldsymbol{U}^{(\boldsymbol{\tau})}, \boldsymbol{a}^{(\boldsymbol{\tau}+1)}\right\rangle^2 \middle| \mathcal{F}^{(\boldsymbol{\tau})}\right]. \tag{4.24}$$

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

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

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

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

$$\left(\boldsymbol{\beta}^{(\boldsymbol{\tau}+1)}\right)^2 \le \left\|\boldsymbol{v}^{(\boldsymbol{\tau})} - \boldsymbol{v}^{(\boldsymbol{\tau}_{m-1})}\right\|^2. \tag{4.25}$$

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

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

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

$$\begin{aligned}
\mathbb{E}\left[\left\langle \boldsymbol{U}^{(\boldsymbol{\tau}+1)} - \boldsymbol{U}^{(\boldsymbol{\tau})}, \boldsymbol{a}^{(\boldsymbol{\tau}+1)}\right\rangle^2 \,\middle|\, \mathcal{F}^{(\boldsymbol{\tau})}\right] &= \mathbb{E}\left[\left\langle \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma}\left[\boldsymbol{x} \,\middle|\, \boldsymbol{x} \in \boldsymbol{X}^{(\boldsymbol{\tau}+1)}\right] - \boldsymbol{u}^{(\tau)}, \boldsymbol{a}^{(\boldsymbol{\tau}+1)}\right\rangle^2 \,\middle|\, \mathcal{F}^{(\boldsymbol{\tau})}\right] \\
&= \mathbb{E}\left[\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma}\left[\left\langle \boldsymbol{x} - \boldsymbol{u}^{(\boldsymbol{\tau})}, \boldsymbol{a}^{(\boldsymbol{\tau}+1)}\right\rangle^2 \,\middle|\, \boldsymbol{x} \in \boldsymbol{X}^{(\boldsymbol{\tau}+1)}\right] \,\middle|\, \mathcal{F}^{(\boldsymbol{\tau})}\right] \\
&= \mathbb{E}\left[\left\langle \boldsymbol{x} - \boldsymbol{u}^{(\boldsymbol{\tau})}, \boldsymbol{a}^{(\boldsymbol{\tau}+1)}\right\rangle^2 \,\middle|\, \mathcal{F}^{(\boldsymbol{\tau})}\right], \qquad (4.26)
\end{aligned}$$

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

Recall that $\boldsymbol{u}^{(\boldsymbol{\tau})} = \mu(\boldsymbol{X}^{(\boldsymbol{\tau})})$ is the center of mass. Moreover, the unit vector $\boldsymbol{a}^{(\tau+1)}$ is determined by $\mathcal{F}^{(\tau)}$ and also the conditional distribution of $\boldsymbol{x}$ conditioned on $\mathcal{F}^{(\tau)}$ is that of $\boldsymbol{x} \sim \gamma$ conditioned on $\boldsymbol{x} \in \boldsymbol{X}^{(\tau)}$. Thus, using the fact that $\boldsymbol{X}^{(\boldsymbol{\tau})}$ is pairwise clean since Alice is in step 3(a), the right hand side in (4.26) is at most λ.

Final Bound. Substituting the above in (4.24), we have

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

where the second inequality follows from (4.25). This completes the proof of the first statement.

For the moreover part, let us condition on the event $\boldsymbol{\tau}_{m-1} < t < \boldsymbol{\tau}_m$ 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 $\boldsymbol{\tau}_{m-1}$ till t, i.e., $\boldsymbol{v}^{(t+1)} = \boldsymbol{v}^{(\boldsymbol{\tau}_{m-1})}$. Thus we have $\Delta \boldsymbol{z}_1^{(t+1)} = \left\langle \boldsymbol{U}^{(t+1)} - \boldsymbol{U}^{(t)}, \eta \odot \boldsymbol{v}^{(\boldsymbol{\tau}_{m-1})}\right\rangle$. Analogous to (4.23), the component of Alice's center of mass along the previous directions are fixed. Thus $\left\langle \boldsymbol{U}^{(t+1)}, \boldsymbol{a}^{(r)}\right\rangle = \left\langle \boldsymbol{U}^{(t)}, \boldsymbol{a}^{(r)}\right\rangle$ for all $r \le t$. Furthermore, by construction, $\eta \odot \boldsymbol{v}^{(\boldsymbol{\tau}_{m-1})}$ lies in the linear subspace spanned by $\boldsymbol{a}^{(0)}, \ldots, \boldsymbol{a}^{(\boldsymbol{\tau}_{m-1}+1)}$. Therefore, since $\boldsymbol{\tau}_{m-1} + 1 \le t$, it follows that $\Delta \boldsymbol{z}_1^{(t+1)} = 0$.

Expected Norm of Final Center of Mass (Proof of Lemma 4.6.4)

Let $\boldsymbol{H}_A = \boldsymbol{H}_A^{(\boldsymbol{d})}$ be the (random) linear subspace spanned by the vectors $\boldsymbol{a}^{(0)}, \ldots, \boldsymbol{a}^{(\boldsymbol{d})}$ and similarly, let $\boldsymbol{H}_B = \boldsymbol{H}_B^{(\boldsymbol{d})}$ be the linear subspace spanned by the vectors $\boldsymbol{b}^{(0)}, \ldots, \boldsymbol{b}^{(\boldsymbol{d})}$. 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

$$\left\|\boldsymbol{u}^{(\boldsymbol{d})}\right\|^2 = \left\|\mathbf{\Pi}_{H_A}\boldsymbol{u}^{(\boldsymbol{d})}\right\|^2 + \left\|\mathbf{\Pi}_{H_A^\perp}\boldsymbol{u}^{(\boldsymbol{d})}\right\|^2 \text{ and } \left\|\boldsymbol{v}^{(\boldsymbol{d})}\right\|^2 = \left\|\mathbf{\Pi}_{H_B}\boldsymbol{v}^{(\boldsymbol{d})}\right\|^2 + \left\|\mathbf{\Pi}_{H_B^\perp}\boldsymbol{v}^{(\boldsymbol{d})}\right\|^2.$$

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

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

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

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

In an analogous fashion,

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

We next show that both $\mathbb{E}[\boldsymbol{p}_A + \boldsymbol{p}_B]$ and $\mathbb{E}[\boldsymbol{q}_A + \boldsymbol{q}_B]$ 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 $\mathbb{E}\left[\left\langle \boldsymbol{x}, \boldsymbol{a}^{(t)}\right\rangle^2\right]$ is large. The latter follows from an application of level-1 inequalities.

We will bound the norm of the projection on the subspaces $\boldsymbol{H}_A$ and $\boldsymbol{H}_B$, which corresponds to the quantity $\mathbb{E}[\boldsymbol{p}_A + \boldsymbol{p}_B]$, and bound the norm of the projection on the orthogonal subspaces $\boldsymbol{H}_A^\perp$ and $\boldsymbol{H}_B^\perp$, which corresponds to the quantity $\mathbb{E}[\boldsymbol{q}_A + \boldsymbol{q}_B]$.

Projection on the Subspaces $\boldsymbol{H}_A$ and $\boldsymbol{H}_B$

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

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

Towards this end, define the random variable $\boldsymbol{k}_t = \boldsymbol{k}_t(\boldsymbol{x}, \boldsymbol{y}) = \left\langle \boldsymbol{x}, \boldsymbol{a}^{(t)} \right\rangle^2 + \left\langle \boldsymbol{y}, \boldsymbol{b}^{(t)} \right\rangle^2$ for each $t \in \mathbb{N}$. Note that the vectors $\boldsymbol{a}^{(t)}$'s are being chosen adaptively depending on the previous inner products $\left\langle \boldsymbol{x}, \boldsymbol{a}^{(\tau)} \right\rangle$ for $\tau < t$, as well as the Boolean communication bits from step 3(b), thus they are functions of $\boldsymbol{x}$ and $\boldsymbol{y}$ as well here. Observe that

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

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

Claim 4.6.5. *For any $r \in \mathbb{N}$, we have*

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

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 4.6.6. *For any $r \in \mathbb{N}$, we have* $\mathbf{Pr}\left[\boldsymbol{d} > \binom{r}{2} \cdot 4d\right] \leq 2 \cdot 2^{-r}$.

Proof of Claim 4.6.6. We bound $\mathbf{Pr}\left[\boldsymbol{d} > \binom{r}{2} \cdot 4d\right]$ by induction on r. The claim trivially holds for $r = 1$.

Now we proceed to analyze the event $\boldsymbol{d} \geq \binom{r+1}{2} \cdot 4d$. Observe that Claim 4.6.1 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 determined[9] by $\mathcal{F}^{(t-1)} \cap \{\boldsymbol{d} > \binom{r}{2} \cdot 4d\}$ where the process is in step 3(c) and Alice speaks, then

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

where $\lambda = 100$ is the cleanup parameter and $\mu(X) = \mathbb{E}_{\boldsymbol{x}\sim\gamma}[\boldsymbol{x} \mid \boldsymbol{x} \in X]$ is the center of mass. This is because $\boldsymbol{a}^{(t)}$ 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 $\left\langle \boldsymbol{y}, \boldsymbol{b}^{(t)} \right\rangle^2$ where $\boldsymbol{y}$ is sampled from γ conditioned on Y.

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

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

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

Recall that Claim 4.6.5 implies that the right hand side is at most $20dr + 4\ln\left(\frac{1}{\mathbf{Pr}[\boldsymbol{d} > \binom{r}{2} \cdot 4d]}\right)$. We consider two cases:

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

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

$$\mathbf{Pr}\left[\boldsymbol{d} > \tbinom{r+1}{2} \cdot 4d \,\middle|\, \boldsymbol{d} > \tbinom{r}{2} \cdot 4d\right] \leq 1/2,$$

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

These claims imply that

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

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 $\mathbb{E}[\boldsymbol{p}_A + \boldsymbol{p}_B]$ assuming Claim 4.6.5 which we prove next.

Proof of Claim 4.6.5. To prove the claim, we need to analyze the expectation of $\sum_{t \in I_r} \boldsymbol{k}_t$ under $\boldsymbol{x}, \boldsymbol{y}$ sampled from γ conditioned on the event $\boldsymbol{d} \geq \binom{r}{2} \cdot 4d$.

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 $\overline{\mathcal{C}}$ is solely determined by the previous Boolean communication, since it is specified by the original protocol $\widetilde{\mathcal{C}}$ (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 $\widetilde{\mathcal{C}}$. 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 $\widetilde{\mathcal{C}}$. If we sample $\boldsymbol{x}, \boldsymbol{y} \sim \gamma$

conditioned on $\boldsymbol{d} > \binom{r}{2} \cdot 4d$ and output the unique $(\boldsymbol{X}_c, \boldsymbol{Y}_c)$ such that $(\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{X}_c \times \boldsymbol{Y}_c$, we obtain a distribution on rectangles. We use $\gamma(X_c \times Y_c \,|\, \boldsymbol{d} > \binom{r}{2} \cdot 4d)$ to denote the probability of obtaining $X_c \times Y_c$ by this sampling process so that $\sum_c \gamma(X_c \times Y_c \,|\, \boldsymbol{d} > \binom{r}{2} \cdot 4d) = 1$.

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 $\boldsymbol{x}, \boldsymbol{y}$ sampled from γ_n conditioned on the event that $\{(\boldsymbol{x}, \boldsymbol{y}) \in X_c \times Y_c\} \wedge \{\boldsymbol{d} > \binom{r}{2} \cdot 4d\}$. We shall show the following claim for any rectangle $X_c \times Y_c$ that could be sampled in the first step.

Claim 4.6.7. $\mathbb{E}_{\boldsymbol{x},\boldsymbol{y}\sim\gamma}\left[\sum_{t\in I_r} \boldsymbol{k}_t \,\middle|\, \boldsymbol{d} > 4d\binom{r}{2}, (\boldsymbol{x}, \boldsymbol{y}) \in X_c \times Y_c\right]$ *is at most*

$$12dr + 4\ln\left(\frac{1}{\mathbf{Pr}[\boldsymbol{d} > 4d\binom{r}{2}, (\boldsymbol{x},\boldsymbol{y}) \in X_c \times Y_c]}\right).$$

Assuming the above, and taking an expectation over $X_c \times Y_c$ drawn with probability $\gamma(X_c \times Y_c \,|\, \boldsymbol{d} > \binom{r}{2} \cdot 4d)$, we immediately obtain Claim 4.6.5:

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\boldsymbol{x},\boldsymbol{y}\sim\gamma}\left[\sum_{t\in I_r} \boldsymbol{k}_t \,\middle|\, \boldsymbol{d} > \binom{r}{2} \cdot 4d\right] \\
&\leq 12dr + \\
&\quad 4 \cdot \sum_{c\in\{0,1\}^*, |c|\leq d} \gamma(X_c \times Y_c | \boldsymbol{d} > \binom{r}{2} \cdot 4d) \cdot \left(\ln\left(\frac{1}{\gamma(X_c \times Y_c | \boldsymbol{d} > \binom{r}{2} \cdot 4d)}\right) + \ln\left(\frac{1}{\mathbf{Pr}[\boldsymbol{d} > \binom{r}{2} \cdot 4d]}\right)\right) \\
&\leq 12dr + 4 \cdot \ln(3^d) + 4 \cdot \ln\left(\frac{1}{\mathbf{Pr}[\boldsymbol{d} > \binom{r}{2} \cdot 4d]}\right) && \text{(by concavity of } \ln(\cdot)\text{)} \\
&\leq 20dr + 4 \cdot \ln\left(\frac{1}{\mathbf{Pr}[\boldsymbol{d} > \binom{r}{2} \cdot 4d]}\right). && \square
\end{aligned}$$

To complete the proof, we now prove Claim 4.6.7.

Proof of Claim 4.6.7. Fix any c such that $\gamma(X_c \times Y_c \,|\, \boldsymbol{d} > \binom{r}{2} \cdot 4d) > 0$. We will bound the expectation of the quantity $\sum_{t\in I_r} \boldsymbol{k}_t = \sum_{t\in I_r} \left\langle \boldsymbol{x}, \boldsymbol{a}^{(t)} \right\rangle^2 + \left\langle \boldsymbol{y}, \boldsymbol{b}^{(t)} \right\rangle^2$ where $\boldsymbol{x}, \boldsymbol{y}$ are sampled from γ_n conditioned on the event that $\{(\boldsymbol{x}, \boldsymbol{y}) \in X_c \times Y_c\} \wedge \{\boldsymbol{d} > \binom{r}{2} \cdot 4d\}$. Note that $\boldsymbol{a}^{(t)}, \boldsymbol{b}^{(t)}, \boldsymbol{d}$ are functions of the previous messages of the protocol and hence also the inputs $\boldsymbol{x}, \boldsymbol{y}$. 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 $\sum_{t\in I_r} \boldsymbol{k}_t$ 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 $\overline{\mathcal{C}}$ on any input x, y. Consider the random walk on this new protocol tree where the inputs $\boldsymbol{x}, \boldsymbol{y} \sim \gamma$ (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 $\boldsymbol{a}^{(t)}$ and $\boldsymbol{b}^{(t)}$ 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 $\boldsymbol{a}^{(t)}$ and $\boldsymbol{b}^{(t)}$ 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 $\boldsymbol{a}^{(\cdot)}$ is chosen to be a unit vector orthogonal to the previously chosen $\boldsymbol{a}^{(\cdot)}$'s and the same holds for the vectors $\boldsymbol{b}^{(\cdot)}$.

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

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 $\boldsymbol{x}$ and $\boldsymbol{y}$ are standard Gaussian distributions on an affine subspace of $\mathbb{R}^n$ (defined by the previous messages). Then, at time t, since $\boldsymbol{a}^{(t)}$ is orthogonal to the directions used in all previous real messages, it follows that the distribution of $\left\langle \boldsymbol{x}, \boldsymbol{a}^{(t)} \right\rangle$ conditioned on any event in $\mathcal{G}^{(t-1)}$ is an independent standard Gaussian for every t if $\boldsymbol{a}^{(t)}$ is non-zero. The same holds for $\left\langle \boldsymbol{y}, \boldsymbol{b}^{(t)} \right\rangle$ 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 $\left\langle \boldsymbol{x}, \boldsymbol{a}^{(t)} \right\rangle^2$ and $\left\langle \boldsymbol{y}, \boldsymbol{b}^{(t)} \right\rangle^2$ 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 2.0.5 implies that

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

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

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

$$\begin{aligned} &\mathop{\mathbf{Pr}}_{(\boldsymbol{x},\boldsymbol{y})\sim\gamma}\left[\sum_{t\in I_r} \boldsymbol{k}_t(\boldsymbol{x},\boldsymbol{y}) \geq 8dr + s \,\middle|\, \boldsymbol{d} > \tbinom{r}{2} \cdot 4d, (\boldsymbol{x},\boldsymbol{y}) \in X_c \times Y_c\right] \\ &= \frac{\mathbf{Pr}_{\boldsymbol{x},\boldsymbol{y}\sim\gamma}\left[\sum_{t\in I_r} \boldsymbol{k}_t(\boldsymbol{x},\boldsymbol{y}) \geq 8dr + s, (\boldsymbol{x},\boldsymbol{y}) \in X_c \times Y_c, \boldsymbol{d} > \binom{r}{2} \cdot 4d\right]}{\mathbf{Pr}_{\boldsymbol{x},\boldsymbol{y}\sim\gamma}\left[(\boldsymbol{x},\boldsymbol{y}) \in X_c \times Y_c, \boldsymbol{d} > \binom{r}{2} \cdot 4d\right]}. \end{aligned} \tag{4.27}$$

We then bound the numerator by

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

Note that the inequality gives us an exponential tail on (4.27):

$$(4.27)\leq e^{-s/4}\cdot\left(\Pr_{\boldsymbol{x},\boldsymbol{y}\sim\gamma}\left[(\boldsymbol{x},\boldsymbol{y})\in X_c\times Y_c,\boldsymbol{d}>\binom{r}{2}\cdot 4d\right]\right)^{-1}.$$

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

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

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

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

This completes the proof of Claim 4.6.7. □

Projection on the Orthogonal Subspaces $\boldsymbol{H}_A^\perp$ and $\boldsymbol{H}_B^\perp$

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

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

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

The proof will follow in two steps. We will first show that one can bound the norm of the projection $\boldsymbol{\Pi}_{\boldsymbol{H}_A^\perp}\boldsymbol{u}^{(d)}$, which turns out to be the Gaussian center of mass of a set that lives in the subspace $\boldsymbol{H}_A^\perp$, 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 $\gamma_{\boldsymbol{H}_A^\perp}$ on the subspace $\boldsymbol{H}_A^\perp$. The case for $\boldsymbol{\Pi}_{\boldsymbol{H}_B^\perp}\boldsymbol{u}^{(d)}$ 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 $\boldsymbol{x}, \boldsymbol{y} \sim \gamma$ 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 $\boldsymbol{\ell} = (\boldsymbol{c}, \boldsymbol{r})$ denote the random variable representing the full transcript of the generalized protocol where $\boldsymbol{c}$ is the Boolean communication and $\boldsymbol{r}$ is the real communication. For any given transcript $\boldsymbol{\ell}$, let $\boldsymbol{X}_{\boldsymbol{\ell}} \times \boldsymbol{Y}_{\boldsymbol{\ell}}$ denote the corresponding rectangle consists of inputs reaching the leaf, and let $\boldsymbol{X}_{\boldsymbol{c}} \times \boldsymbol{Y}_{\boldsymbol{c}}$ (for $\boldsymbol{X}_{\boldsymbol{c}}, \boldsymbol{Y}_{\boldsymbol{c}} \subseteq \mathbb{R}^n$) denote the rectangle consisting of all pairs of inputs to Alice and Bob that result in the Boolean transcript $\boldsymbol{c}$. Note that the real communication $\boldsymbol{r}$ together with $\boldsymbol{c}$ fixes the subspaces $\boldsymbol{H}_A$ and $\boldsymbol{H}_B$ and particular affine shifts $\boldsymbol{s}_A$ and $\boldsymbol{s}_B$ of those subspaces depending on the value of the inner products determined by the full transcript. In particular, the rectangle $\boldsymbol{X}_{\boldsymbol{\ell}} \times \boldsymbol{Y}_{\boldsymbol{\ell}}$ consistent with the full transcript $\boldsymbol{\ell} = (\boldsymbol{c}, \boldsymbol{r})$ is given by $\boldsymbol{X}_{\boldsymbol{\ell}} = \boldsymbol{X}_{\boldsymbol{c}} \cap (\boldsymbol{H}_A + \boldsymbol{s}_A)$ and $\boldsymbol{Y}_{\boldsymbol{\ell}} = \boldsymbol{Y}_{\boldsymbol{c}} \cap (\boldsymbol{H}_B + \boldsymbol{s}_B)$, i.e., taking (random) affine slices of the original sets.

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

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{q}_A + \boldsymbol{q}_B\right] &= \mathbb{E}\left[\left\|\boldsymbol{\Pi}_{\boldsymbol{H}_A^\perp}\boldsymbol{u}^{(\boldsymbol{d})}\right\|^2 + \left\|\boldsymbol{\Pi}_{\boldsymbol{H}_A^\perp}\boldsymbol{u}^{(\boldsymbol{d})}\right\|^2\right] \\ &= \mathop{\mathbb{E}}_{\boldsymbol{\ell}}\left[\|\mu_{\boldsymbol{H}_A^\perp}(\boldsymbol{\Pi}_{\boldsymbol{H}_A^\perp}\boldsymbol{X}_{\boldsymbol{\ell}})\|^2 + \|\mu_{\boldsymbol{H}_B^\perp}(\boldsymbol{\Pi}_{\boldsymbol{H}_B^\perp}\boldsymbol{Y}_{\boldsymbol{\ell}})\|^2\right].\end{aligned}$$

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

Claim 4.6.8. $\|\mu_{\boldsymbol{H}_A^\perp}(\boldsymbol{\Pi}_{\boldsymbol{H}_A^\perp}\boldsymbol{X}_{\boldsymbol{\ell}})\|^2 \le 2e^2 \ln\left(\dfrac{e}{\gamma_{\mathrm{rel}}\left(\boldsymbol{X}_{\boldsymbol{\ell}}\right)}\right)$ *and*

$$\|\mu_{\boldsymbol{H}_B^\perp}(\boldsymbol{\Pi}_{\boldsymbol{H}_B^\perp}\boldsymbol{Y}_{\boldsymbol{\ell}})\|^2 \le 2e^2 \ln\left(\frac{e}{\gamma_{\mathrm{rel}}\left(\boldsymbol{Y}_{\boldsymbol{\ell}}\right)}\right).$$

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

Using the previous claim,

$$\mathbb{E}\left[\boldsymbol{q}_A + \boldsymbol{q}_B\right] = \mathbb{E}\left[\left\|\boldsymbol{\Pi}_{\boldsymbol{H}_A^\perp}\boldsymbol{u}^{(\boldsymbol{d})}\right\|^2 + \left\|\boldsymbol{\Pi}_{\boldsymbol{H}_A^\perp}\boldsymbol{u}^{(\boldsymbol{d})}\right\|^2\right] \leq 2e^2 \cdot \mathop{\mathbb{E}}_{\boldsymbol{\ell}}\left[\ln\left(\frac{e}{\gamma_{\mathrm{rel}}\left(\boldsymbol{X}_{\boldsymbol{\ell}} \times \boldsymbol{Y}_{\boldsymbol{\ell}}\right)}\right)\right].$$

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 4.6.9. $\mathbb{E}_{\boldsymbol{\ell}}\left[\ln\left(\frac{e}{\gamma_{\mathrm{rel}}\left(\boldsymbol{X}_{\boldsymbol{\ell}} \times \boldsymbol{Y}_{\boldsymbol{\ell}}\right)}\right)\right] = O(d).$

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

Proof of Claim 4.6.8. We can bound the norm of the above projection by an application of the Gaussian level-1 inequality (Theorem 2.0.6), which, by rotational symmetry, implies that if A is a subset of a linear subspace V with non-zero measure, then

$$\|\mu_V(A)\|^2 \leq 2e^2 \ln\left(\frac{e}{\gamma_V(A)}\right), \tag{4.28}$$

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

If we run the generalized protocol on $\boldsymbol{x}, \boldsymbol{y} \sim \gamma$ and condition on getting the full transcript $\boldsymbol{\ell}$, the conditional probability measure on $\boldsymbol{\Pi}_{\boldsymbol{H}_A^\perp}\boldsymbol{x}$ is that of the Gaussian measure $\gamma_{\boldsymbol{H}_A^\perp}$ conditioned on $\boldsymbol{x} \in \boldsymbol{X}_{\boldsymbol{\ell}} - \boldsymbol{s}_A$ and $\boldsymbol{\Pi}_{\boldsymbol{H}_A^\perp}\boldsymbol{y}$ is that of the Gaussian measure $\gamma_{\boldsymbol{H}_B^\perp}$ conditioned on $\boldsymbol{y} \in \boldsymbol{Y}_{\boldsymbol{\ell}} - \boldsymbol{s}_B$ 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 $\boldsymbol{H}_A$ and $\boldsymbol{H}_B$ and the fact the projection of a standard Gaussian on orthogonal subspaces are independent.

Thus, applying (4.28), we have

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

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

Proof of Claim 4.6.9. For this claim, it will be convenient to consider a different generalized protocol $\mathcal{C}'$ that generates the same distribution on the leaves $\boldsymbol{\ell}$. In particular, since the Boolean messages in the generalized protocol $\overline{\mathcal{C}}$ only depend on the previous Boolean messages, one can first send all the Boolean messages $\boldsymbol{c}$, and then send all the real messages $\boldsymbol{r}$

choosing them according to the protocol $\overline{\mathcal{C}}$ depending on the previous real messages and the (partial) Boolean transcript. Note that the protocol $\mathcal{C}'$ generates the same distribution on the leaves $\boldsymbol{\ell}$ when the inputs $\boldsymbol{x}, \boldsymbol{y} \sim \gamma_n$. In particular, the real communication only partitions [10] 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 $\boldsymbol{c} = c$ and by induction show that

$$\mathop{\mathbb{E}}_{\boldsymbol{r}}\left[\ln\left(\frac{e}{\gamma_{\mathrm{rel}}(\boldsymbol{X}_{(c,\boldsymbol{r})} \times \boldsymbol{Y}_{(c,\boldsymbol{r})})}\right) \middle| \boldsymbol{c} = c\right] \le \ln\left(\frac{e}{\gamma_{\mathrm{rel}}(X_c \times Y_c)}\right), \tag{4.29}$$

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 $\gamma_{\mathrm{rel}}(X_c \times Y_c)$ 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,

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

where the last line follows from concavity.

Induction. To complete the proof, we now show (4.29) 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 $r_{<t} := r_1, \ldots, r_{t-1}$. 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:

$$\mathop{\mathbb{E}}_{\boldsymbol{r}_{\le t}}\left[\ln\left(\frac{e}{\gamma_{\mathrm{rel}}(\boldsymbol{X}_{(c,\boldsymbol{r}_{\le t})})}\right) \middle| \boldsymbol{c} = c, \boldsymbol{r}_{<t} = r_{<t}\right] \le \ln\left(\frac{e}{\gamma_{\mathrm{rel}}(X_{(c,r_{<t})})}\right), \tag{4.30}$$

and an analogous statement holds when Bob speaks. Taking an expectation over $\boldsymbol{r}_{<t}$, the above directly applies (4.29) by a straightforward backward induction:

$$\mathop{\mathbb{E}}_{\boldsymbol{r}_{\le t}}\left[\ln\left(\frac{e}{\gamma_{\mathrm{rel}}(\boldsymbol{X}_{(c,\boldsymbol{r}_{\le t})} \times \boldsymbol{Y}_{(c,\boldsymbol{r}_{\le t})})}\right) \middle| \boldsymbol{c} = c\right] \le \mathop{\mathbb{E}}_{\boldsymbol{r}_{<t}}\left[\ln\left(\frac{e}{\gamma_{\mathrm{rel}}(\boldsymbol{X}_{(c,\boldsymbol{r}_{<t})} \times \boldsymbol{Y}_{(c,\boldsymbol{r}_{<t})})}\right) \middle| \boldsymbol{c} = c\right]$$

[10] We remark that this protocol $\mathcal{C}'$ suffices for proving this claim since we are looking only at the leaves. However, unlike Lemma 4.6.3, directly bounding the expected quadratic variation of the martingale corresponding to the protocol $\mathcal{C}'$ is difficult.

$$\leq \cdots \leq \ln\left(\frac{e}{\gamma_{\mathrm{rel}}(X_c \times Y_c)}\right).$$

To see (4.30), 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 $a^{(1)}, \ldots, a^{(t-1)}$ 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 $a^{(1)}, \ldots, a^{(t-1)}$ 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 $\boldsymbol{x} \in \mathbb{R}^n$ can be described as follows: the projections corresponding to the non-zero vectors in the sequence $a^{(1)}, \ldots, a^{(t-1)}$, i.e., the inner products $\langle \boldsymbol{x}, a^{(\tau)}\rangle$ 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 $\boldsymbol{x} \in X - s$ (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 $\boldsymbol{x} \in \mathbb{R}^{n-k}$ where $\boldsymbol{x}$ is sampled from γ_{n-k} conditioned on $\boldsymbol{x} \in X$. In this case, γ_{rel} in the left hand side of (4.30) is exactly $\gamma_{\mathrm{rel}}(X \cap \{x_1 = r\})$ if Alice sends r as the message, while for the right hand side of (4.30), we have $\gamma_{\mathrm{rel}}(X) = \gamma_{n-k}(X)$. Denoting by $\mathrm{d}\,\mu_{x_1}$ the probability density function of $\boldsymbol{x}_1$, our statement boils down to showing

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

We show the above by explicitly writing the probability density function $\mathrm{d}\,\mu_{x_1}$. Denote by $\mathrm{d}\,\gamma_{n-k}(x_1, \ldots, x_{n-k})$ the standard Gaussian density function[11] in $\mathbb{R}^{n-k}$. The density function of the random vector $\boldsymbol{x}$ sampled from γ_{n-k} conditioned on $x \in X$, is given $\gamma_{n-k}(X)^{-1} \cdot \mathrm{d}\,\gamma_{n-k}(x_1, \ldots, x_{n-k})$ for $x \in X$ and zero outside. Thus, we have

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

Then, by concavity, the left hand side of (4.30) is exactly given by

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

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

4.7 Level-Two Fourier Growth

In this section, we prove Theorem 4.2.2 that $L_2(h) = O\left(d^{3/2}\log^3(n)\right)$. Similar to the proof of level-1 bound Theorem 4.2.1, we start with a d-round communication protocol $\widetilde{\mathcal{C}}$ over the Gaussian space as defined in Section 4.5. Note that $\widetilde{\mathcal{C}}$ 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 Section 4.5, to bound the second-level Fourier growth, one can attempt to bound the expected quadratic variation of the martingale that results from the protocol $\overline{\mathcal{C}}$ directly, but similar to the case of level-1 it is hard to leverage cancellations here to prove the bound we aim for. So, starting from $\widetilde{\mathcal{C}}$, we will define a communication protocol $\overline{\mathcal{C}}$ that computes the same function as $\widetilde{\mathcal{C}}$, but satisfies some additional "clean" property where it is easier to control the quadratic variation. This new protocol will differ from $\widetilde{\mathcal{C}}$ in two ways. Firstly, the protocol $\overline{\mathcal{C}}$ will consist of additional "cleanup steps" where Alice and Bob reveal certain *quadratic forms* of their input. Secondly, the protocol $\overline{\mathcal{C}}$ 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 $\overline{\mathcal{C}}$ from the corresponding level-1 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 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 4.5.2 and Proposition 4.5.4, it still suffices to analyze the corresponding martingale under the truncated Gaussian distribution.

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

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

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

where we recall that $\sigma(X) = \mathbb{E}_{\boldsymbol{x}\sim\gamma}\left[\boldsymbol{x}\mathbin{\dot\otimes}\boldsymbol{x} \,\middle|\, \boldsymbol{x}\in X\right]$ is the level-2 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

$\left\langle x \dot{\otimes} x, a \right\rangle$ 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 $\mathrm{trunc}_L(z)$ function defined at $z \in \mathbb{R}$ as

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

Construct $\overline{\mathcal{C}}$ from $\widetilde{\mathcal{C}}$. As described before, $\overline{\mathcal{C}}$ 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 $\boldsymbol{X}^{(t)}$ (resp., $\boldsymbol{Y}^{(t)}$) 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 $\boldsymbol{a}^{(t)}$ to denote the matrix of the quadratic form revealed at this node, otherwise set $\boldsymbol{a}^{(t)}$ to be the all-zeroes matrix. We define $\boldsymbol{b}^{(t)}$ similarly for Bob. Throughout the protocol, we will always set $\boldsymbol{u}^{(t)}$ and $\boldsymbol{v}^{(t)}$ to denote $\sigma(\boldsymbol{X}^{(t)})$ and $\sigma(\boldsymbol{Y}^{(t)})$ 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 $\overline{\mathcal{C}}$ 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 $t \leftarrow 0$, $\boldsymbol{X}^{(0)}, \boldsymbol{Y}^{(0)} \leftarrow [-T, T]^n$, and $\boldsymbol{a}^{(0)}, \boldsymbol{b}^{(0)} \leftarrow 0^{n \times n}$.
3. For each phase $i = 1, 2, \ldots, d$: suppose we are starting the cleanup for a node at depth i in the original protocol $\widetilde{\mathcal{C}}$ and suppose we are at a node of depth t in the new protocol $\overline{\mathcal{C}}$. If it is Alice's turn to speak in $\widetilde{\mathcal{C}}$:

    a) **Orthogonalization by Revealing Correlation with Bob's Center of Mass.** Alice begins by revealing the inner product of her input x with Bob's current (signed) level-2 center of mass $\eta \odot \boldsymbol{v}^{(t)}$. 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 $\eta \odot \boldsymbol{v}^{(t)}$ that is orthogonal to the previous directions along which Alice announced the inner product. More formally, let $\boldsymbol{a}^{(t+1)}$ be the component of $\eta \odot \boldsymbol{v}^{(t)}$ that is orthonormal to the span of the previous directions $\boldsymbol{a}^{(\tau)}$ for $\tau \le t$, i.e.,

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

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

b) **Original Communication.** Alice sends the bit $\overline{\boldsymbol{c}}^{(t+1)}$ that she was supposed to send in $\widetilde{\mathcal{C}}$ based on previous messages and x. Set $\boldsymbol{a}^{(t+1)}, \boldsymbol{b}^{(t+1)} \leftarrow 0^{n\times n}$. 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., $\left\langle a, \boldsymbol{a}^{(\tau)}\right\rangle = 0$ for all $\tau \le t$, and $\boldsymbol{X}^{(t)}$ is *not* 4-*wise clean* in direction a, Alice computes $\overline{\boldsymbol{c}}^{(t+1)} \leftarrow \mathrm{trunc}_L\left(\left\langle x \dot{\otimes} x, a\right\rangle\right)$ and sends $\overline{\boldsymbol{c}}^{(t+1)}$ to Bob. Set $\boldsymbol{a}^{(t+1)} \leftarrow a$ and $\boldsymbol{b}^{(t+1)} \leftarrow 0^{n\times n}$. Increment t by 1. Repeat step (c) while $\boldsymbol{X}^{(t)}$ 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 $x, \boldsymbol{a}, \boldsymbol{X}, \boldsymbol{U}$ switched with $y, \boldsymbol{b}, \boldsymbol{Y}, \boldsymbol{v}$.

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

Remark 4.7.1. *Note that by construction, the non-zero matrices among $\boldsymbol{a}^{(1)}, \boldsymbol{a}^{(2)}, \ldots$ form an orthonormal set when viewed as n^2-dimensional vectors (similarly for $\boldsymbol{b}^{(1)}, \boldsymbol{b}^{(2)}, \ldots$) and moreover, their diagonals are zero. Lastly, $\boldsymbol{a}^{(t)}$ and $\boldsymbol{b}^{(t)}$ 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 $\boldsymbol{X}^{(t)} \times \boldsymbol{Y}^{(t)}$ is 4-wise clean for both parties by construction.

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

Fact 4.7.2. *$\boldsymbol{u}^{(0)} = \boldsymbol{v}^{(0)} = 0^{n\times n}$ and each $\boldsymbol{u}^{(t)}, \boldsymbol{v}^{(t)}$ has zero diagonal.*

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

Fact 4.7.3. *Let $\gamma^* = \gamma(\boldsymbol{X}^{(0)}) \cdot \gamma(\boldsymbol{Y}^{(0)})$. Then $\gamma^* \ge 1 - O\left(n \cdot e^{-T^2/2}\right)$.*

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 4.7.4. *In step 3(a) and 3(c), any $x, x' \in \boldsymbol{X}^{(t+1)}$ satisfies*

$$\left|\left\langle x \dot{\otimes} x, \boldsymbol{a}^{(t+1)}\right\rangle - \left\langle x' \dot{\otimes} x', \boldsymbol{a}^{(t+1)}\right\rangle\right| < 2^{-L}.$$

Similarly any $y, y' \in \boldsymbol{Y}^{(t+1)}$ *satisfies* $\left|\left\langle y \dot{\otimes} y, \boldsymbol{b}^{(t+1)}\right\rangle - \left\langle y' \dot{\otimes} y', \boldsymbol{b}^{(t+1)}\right\rangle\right| < 2^{-L}$.

The next claim bounds the maximum attainable norms for Alice and Bob's level-2 center of masses at any point in the protocol. This uses the fact that the inputs come from the truncated Gaussian distribution.

Claim 4.7.5. $\left\|\boldsymbol{u}^{(t)}\right\|_{\mathrm{F}} = \left\|\eta \odot \boldsymbol{u}^{(t)}\right\|_{\mathrm{F}} < nT$ *and* $\left\|\boldsymbol{v}^{(t)}\right\|_{\mathrm{F}} = \left\|\eta \odot \boldsymbol{v}^{(t)}\right\|_{\mathrm{F}} < nT$ *for all possible* t *and* $\boldsymbol{u}^{(t)}, \boldsymbol{v}^{(t)}$ *throughout the communication.*

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

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

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

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

Claim 4.7.6. *For any* $x \in \boldsymbol{X}^{(0)}$ *and* $y \in \boldsymbol{Y}^{(0)}$*, any message in* $\overline{\mathcal{C}}(x, y)$ *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 $\mathrm{trunc}_L(\left\langle x \dot{\otimes} x, a\right\rangle)$ for some $a \in \mathbb{S}^{n\times n-1}$ and send the result. Since

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

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 $\overline{\mathcal{C}}$.

Claim 4.7.7. *Let* ℓ *be an arbitrary leaf of the protocol* $\overline{\mathcal{C}}$ *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* $\boldsymbol{a}^{(t)}$ *and at most* $n^2 - n$ *many non-zero* $\boldsymbol{b}^{(t)}$ *for* $t \in \{1, \ldots, D(\ell)\}$.

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 $\boldsymbol{a}^{(t)}$'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 $d \cdot \mathsf{polylog}(n)$.

Bounding the Expected Quadratic Variation

Consider the martingale process defined in (4.19) from a random walk on the protocol tree generated by $\overline{\mathcal{C}}$ where the inputs $\boldsymbol{x}, \boldsymbol{y}$ are sampled from γ_n conditioned on being in the bounded cube $[-T, T]^n$. Recall that Proposition 4.5.3 still holds (see Remark 4.5.5).

Formally, at time t the process is defined by

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

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

In light of Remark 4.5.2 and Proposition 4.5.4, to prove Theorem 4.2.2, it suffices to prove the following.

Lemma 4.7.8. $\mathbb{E}\left[\sum_{t=1}^{\boldsymbol{d}} \left(\Delta \boldsymbol{z}_2^{(t)}\right)^2\right] = O\left(d^3 \log^6(n)\right).$

Lemma 4.7.8 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-1 bound: Alice's messages sent in step 3(b) and 3(c) are always near-orthogonal to Bob's current level-2 center of mass, thus they do not change the value of the martingale $\boldsymbol{z}_2^{(t)}$ much. Moreover, by level-2 analog of (4.3), since Alice's current node was clean just before Alice sent the message in step 3(a), the expected change $\mathbb{E}\left[\left(\Delta \boldsymbol{z}_2^{(t+1)}\right)^2\right]$ can be bounded in terms of the squared norm of the change that occurred in $\boldsymbol{u}^{(t)}$ (or $\boldsymbol{v}^{(t)}$) 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 $\boldsymbol{z}_2^{(t)}$ is a Doob martingale and also $\boldsymbol{u}^{(t)}, \boldsymbol{v}^{(t)}$ form vector-valued martingales (recall Proposition 4.5.3). Note that $\mathcal{F}^{(t)}$ fixes all the rectangles encountered during times $0, \ldots, t$ and thus for $\tau \le t$, the random variables $\boldsymbol{u}^{(\tau)}, \boldsymbol{v}^{(\tau)}, \boldsymbol{z}_2^{(\tau)}$ 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: $\boldsymbol{u}^{(0)} = \boldsymbol{v}^{(0)} = 0$.

Lemma 4.7.9 (Step Size)**.** *Let $0 = \boldsymbol{\tau}_1 < \boldsymbol{\tau}_2 < \cdots \le \boldsymbol{d}$ be a sequence of stopping times with $\boldsymbol{\tau}_m$ being the index of the round where Alice speaks in step 3(a) for the m^{th} time or $\boldsymbol{d}$ if there is no such round. Then, for any integer $m \ge 2$,*

$$\mathbb{E}\left[\left(\Delta \boldsymbol{z}_2^{(\boldsymbol{\tau}_m+1)}\right)^2 \,\middle|\, \mathcal{F}^{(\boldsymbol{\tau}_m)}\right] \le \lambda \cdot \left\|\boldsymbol{v}^{(\boldsymbol{\tau}_m)} - \boldsymbol{v}^{(\boldsymbol{\tau}_{m-1})}\right\|^2 + 16n^7T^3 \cdot 2^{-L}.$$

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

$$\mathbb{E}\left[\left(\Delta \boldsymbol{z}_2^{(t+1)}\right)^2 \,\middle|\, \mathcal{F}^{(t)}, \boldsymbol{\tau}_{m-1} < t < \boldsymbol{\tau}_m, \textit{Alice speaks at time } t\right] \le 4n^6T^2 \cdot 2^{-2L}$$

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

We indeed see that, if $L = \Omega(\log(n))$ and $T = O(\sqrt{\log(n)})$, then $\mathsf{poly}(T, n) \cdot 2^{-L} = o(1)$, 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 $\boldsymbol{u}^{(\boldsymbol{\tau}_1)}$ and $\boldsymbol{v}^{(\boldsymbol{\tau}'_1)}$ are their initial centers of mass which are both zero.

We shall prove the above lemma later and continue with the bound on the quadratic variation here. Using the bounds on the step sizes from Lemma 4.7.9,

$$\begin{aligned}
&\mathbb{E}\left[\sum_{t=1}^{\boldsymbol{d}} \left(\Delta \boldsymbol{z}_2^{(t)}\right)^2\right] \\
&\le \lambda \cdot \mathbb{E}\left[\sum_{m\ge 2} \left\|\boldsymbol{v}^{(\boldsymbol{\tau}_m)} - \boldsymbol{v}^{(\boldsymbol{\tau}_{m-1})}\right\|^2 + \left\|\boldsymbol{U}^{(\boldsymbol{\tau}'_m)} - \boldsymbol{U}^{(\boldsymbol{\tau}'_{m-1})}\right\|^2\right] + 16n^7T^3 \cdot 2^{-L} \cdot \mathbb{E}[\boldsymbol{d}] \\
&\le \lambda \cdot \mathbb{E}\left[\sum_{m\ge 2} \left\|\boldsymbol{v}^{(\boldsymbol{\tau}_m)} - \boldsymbol{v}^{(\boldsymbol{\tau}_{m-1})}\right\|^2 + \left\|\boldsymbol{U}^{(\boldsymbol{\tau}'_m)} - \boldsymbol{U}^{(\boldsymbol{\tau}'_{m-1})}\right\|^2\right] + 16n^7T^3 \cdot 2^{-L} \cdot 2n^2. \\
&\qquad\qquad \text{(by Claim 4.7.7)}
\end{aligned}$$

On the other hand, using the orthogonality of vector-valued martingale differences from (2.3),

$$\mathbb{E}\left[\sum_{m\ge 2} \left\|\boldsymbol{v}^{(\boldsymbol{\tau}_m)} - \boldsymbol{v}^{(\boldsymbol{\tau}_{m-1})}\right\|^2\right] = \mathbb{E}\left[\left\|\boldsymbol{v}^{(\boldsymbol{d})}\right\|^2\right].$$

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

$$\mathbb{E}\left[\sum_{t=1}^{\boldsymbol{d}} \left(\Delta \boldsymbol{z}_2^{(t)}\right)^2\right] \le \lambda \cdot \left(\mathbb{E}\left[\left\|\boldsymbol{U}^{(\boldsymbol{d})}\right\|_{\mathrm{F}}^2\right] + \mathbb{E}\left[\left\|\boldsymbol{v}^{(\boldsymbol{d})}\right\|_{\mathrm{F}}^2\right]\right) + 64n^9T^3 \cdot 2^{-L}. \tag{4.31}$$

Then we will apply level-2 inequalities (see Theorem 2.0.6) to convert the bounding $\mathbb{E}\left[\left\|\boldsymbol{U}^{(\boldsymbol{d})}\right\|_{\mathrm{F}}^2 + \left\|\boldsymbol{v}^{(\boldsymbol{d})}\right\|_{\mathrm{F}}^2\right]$ into bounding the second moment $\mathbb{E}[\boldsymbol{d}^2]$. This reduction is formalized as Lemma 4.7.10 below and its proof is similar to [GRT22, Claim 1].

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

Lemma 4.7.10. $\mathbb{E}\left[\left\|\boldsymbol{u}^{(\boldsymbol{d})}\right\|_{\mathrm{F}}^2 + \left\|\boldsymbol{v}^{(\boldsymbol{d})}\right\|_{\mathrm{F}}^2\right] \leq O\left(\frac{1}{\gamma^*} + L^2\,\mathbb{E}[\boldsymbol{d}^2]\right)$.

Finally, we will bound the second moment $\mathbb{E}[\boldsymbol{d}^2]$ for a suitable choice of parameters.

Lemma 4.7.11. *It holds that* $\mathbb{E}[\boldsymbol{d}^2] = O(d^2)$ *and* $\gamma^* \geq \frac{3}{4}$ *for*

$$L = \Theta(\log(n)), \quad T = \Theta(\sqrt{\log(n)}), \quad \text{and} \quad \lambda = \Theta(d\log^4(n)).$$

Given Lemmas 4.7.10 and 4.7.11,the proof of Lemma 4.7.8 naturally follows.

Proof of Lemma 4.7.8. With the parameters chosen in Lemma 4.7.11, we have

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{\boldsymbol{d}}\left(\Delta\boldsymbol{z}_2^{(t)}\right)^2\right] &\leq O(d\log^4(n))\cdot\left(\mathbb{E}\left[\left\|\boldsymbol{U}^{(\boldsymbol{d})}\right\|_{\mathrm{F}}^2\right] + \mathbb{E}\left[\left\|\boldsymbol{v}^{(\boldsymbol{d})}\right\|_{\mathrm{F}}^2\right]\right) + 1 && \text{(by (4.31))}\\
&\leq O(d\log^4(n))\cdot\left(1+\log^2(n)\cdot\mathbb{E}[\boldsymbol{d}^2]\right) + 1 && \text{(by Lemma 4.7.10)}\\
&\leq O(d\log^4(n))\cdot\left(1+\log^2(n)\cdot d^2\right) + 1 && \text{(by Lemma 4.7.11)}\\
&= O(d^3\log^6(n)). && \square
\end{aligned}$$

Remark 4.7.12. *Note that our proof for level-2 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-2 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.*

Bounds on Step Sizes (Proof of Lemma 4.7.9)

Let us abbreviate $\boldsymbol{\tau} = \boldsymbol{\tau}_m$ and note that at time $\boldsymbol{\tau}$ a new phase starts for Alice. By construction, this means that the current rectangle $\boldsymbol{X}^{(\boldsymbol{\tau})} \times \boldsymbol{Y}^{(\boldsymbol{\tau})}$ determined by $\mathcal{F}^{(\boldsymbol{\tau})}$ is 4-wise clean with parameter λ, and since Alice is in step 3(a) at the start of a new phase,

$\boldsymbol{a}^{(\boldsymbol{\tau}+1)}$ is chosen to be the (normalized) component of $\eta \odot \boldsymbol{v}^{(\boldsymbol{\tau})}$ that is orthogonal to previous directions $\boldsymbol{a}^{(1)}, \ldots, \boldsymbol{a}^{(\boldsymbol{\tau})}$.

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

$$\begin{aligned}\mathbb{E}\left[\left(\Delta \boldsymbol{z}_2^{(\boldsymbol{\tau}+1)}\right)^2 \middle| \mathcal{F}^{(\boldsymbol{\tau})}\right] &= \mathbb{E}\left[\left\langle \boldsymbol{U}^{(\boldsymbol{\tau}+1)} - \boldsymbol{U}^{(\boldsymbol{\tau})}, \eta \odot \boldsymbol{v}^{(\boldsymbol{\tau})} \right\rangle^2 \middle| \mathcal{F}^{(\boldsymbol{\tau})}\right] \\ &= \mathbb{E}\left[\left(\sum_{r=1}^{\boldsymbol{\tau}+1} \boldsymbol{\beta}^{(r)} \cdot \left\langle \boldsymbol{u}^{(\boldsymbol{\tau}+1)} - \boldsymbol{u}^{(\boldsymbol{\tau})}, \boldsymbol{a}^{(r)} \right\rangle\right)^2 \middle| \mathcal{F}^{(\boldsymbol{\tau})}\right]. \end{aligned} \tag{4.32}$$

Similar to the level-1 proof, the components of $\boldsymbol{u}^{(\boldsymbol{\tau}+1)}$ and $\boldsymbol{u}^{(\boldsymbol{\tau})}$ are roughly the same along any of the previous directions $\boldsymbol{a}^{(1)}, \ldots, \boldsymbol{a}^{(\boldsymbol{\tau})}$ and so they almost cancel out and the major quantity is in the direction $\boldsymbol{a}^{(\boldsymbol{\tau}+1)}$. This follows since, in all the previous steps $r \le \boldsymbol{\tau}$, Alice has already fixed $\left\langle x \dot{\otimes} x, \boldsymbol{a}^{(r)} \right\rangle$ with precision 2^{-L}. This implies that for any $\boldsymbol{X}^{(\boldsymbol{\tau})}$ and $\boldsymbol{X}^{(\boldsymbol{\tau}+1)}$ that are determined by $\mathcal{F}^{(\boldsymbol{\tau}+1)}$, the inner product with all the previous $\boldsymbol{a}^{(1)}, \ldots, \boldsymbol{a}^{(\boldsymbol{\tau})}$ is fixed with precision 2^{-L} over the choice of x. Formally, by Fact 4.7.4, we have that for any $x \in \boldsymbol{X}^{(\boldsymbol{\tau})}$ and $x' \in \boldsymbol{X}^{(\boldsymbol{\tau}+1)}$, it holds that $\left|\left\langle x \dot{\otimes} x, \boldsymbol{a}^{(r)} \right\rangle - \left\langle x' \dot{\otimes} x', \boldsymbol{a}^{(r)} \right\rangle\right| \le 2^{-L}$ for all $r \le \boldsymbol{\tau}$. In particular, since $\boldsymbol{u}^{(\boldsymbol{\tau})} = \sigma(\boldsymbol{X}^{(\boldsymbol{\tau})})$ and $\boldsymbol{u}^{(\boldsymbol{\tau}+1)} = \sigma(\boldsymbol{X}^{(\boldsymbol{\tau}+1)})$ are the corresponding centers of mass, we have that

$$\left|\left\langle \boldsymbol{u}^{(\boldsymbol{\tau}+1)} - \boldsymbol{u}^{(\boldsymbol{\tau})}, \boldsymbol{a}^{(r)} \right\rangle\right| \le 2^{-L} \quad \text{for all } r \le \boldsymbol{\tau}. \tag{4.33}$$

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

$$\left|\left\langle \boldsymbol{u}^{(\boldsymbol{\tau}+1)} - \boldsymbol{u}^{(\boldsymbol{\tau})}, \boldsymbol{a}^{(\boldsymbol{\tau}+1)} \right\rangle\right| \le \left\|\boldsymbol{u}^{(\boldsymbol{\tau}+1)} - \boldsymbol{u}^{(\boldsymbol{\tau})}\right\| \le 2nT. \tag{4.34}$$

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

$$\left|\boldsymbol{\beta}^{(r)}\right| = \left|\left\langle \eta \odot \boldsymbol{v}^{(\boldsymbol{\tau})}, \boldsymbol{a}^{(r)} \right\rangle\right| \le \left\|\eta \odot \boldsymbol{v}^{(\boldsymbol{\tau})}\right\| \le \left\|\boldsymbol{v}^{(\boldsymbol{\tau})}\right\| \le nT \quad \text{for all } r \le \boldsymbol{\tau}+1. \tag{4.35}$$

Expanding the square in (4.32) and plugging these estimates to each one of the $(\boldsymbol{\tau}+1)^2$ terms gives

$$\mathbb{E}\left[\left(\Delta \boldsymbol{z}_2^{(\boldsymbol{\tau}+1)}\right)^2 \middle| \mathcal{F}^{(\boldsymbol{\tau})}\right] \tag{4.36}$$

$$\le \mathbb{E}\left[\left(\boldsymbol{\beta}^{(\boldsymbol{\tau}+1)}\right)^2 \left\langle \boldsymbol{u}^{(\boldsymbol{\tau}+1)} - \boldsymbol{u}^{(\boldsymbol{\tau})}, \boldsymbol{a}^{(\boldsymbol{\tau}+1)} \right\rangle^2 + ((\boldsymbol{\tau}+1)^2 - 1) \cdot \tfrac{2(nT)^3}{2^L} \middle| \mathcal{F}^{(\boldsymbol{\tau})}\right]$$

$$\le \left(\boldsymbol{\beta}^{(\boldsymbol{\tau}+1)}\right)^2 \mathbb{E}\left[\left\langle \boldsymbol{u}^{(\boldsymbol{\tau}+1)} - \boldsymbol{u}^{(\boldsymbol{\tau})}, \boldsymbol{a}^{(\boldsymbol{\tau}+1)} \right\rangle^2 \middle| \mathcal{F}^{(\boldsymbol{\tau})}\right] + 12n^7T^3 \cdot 2^{-L}, \tag{4.37}$$

where the second line follows from Claim 4.7.7.

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

Term Outside the Expectation. Recall that $\boldsymbol{a}^{(\boldsymbol{\tau}+1)}$ is chosen to be the (normalized) component of $\eta \odot \boldsymbol{v}^{(\boldsymbol{\tau})}$ that is orthogonal to the span of $\boldsymbol{a}^{(1)}, \ldots, \boldsymbol{a}^{(\boldsymbol{\tau})}$. Since $\eta \odot \boldsymbol{v}^{(\boldsymbol{\tau}_m - 1)}$ is in the span of $\boldsymbol{a}^{(1)}, \ldots, \boldsymbol{a}^{(\boldsymbol{\tau}_{m-1}+1)}$ and $\boldsymbol{\tau}_{m-1} + 1 \le \boldsymbol{\tau} = \boldsymbol{\tau}_m$, it is orthogonal to $\boldsymbol{a}^{(\boldsymbol{\tau}+1)}$. Hence

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

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

$$\left(\boldsymbol{\beta}^{(\boldsymbol{\tau}+1)}\right)^2 \le \left\| \boldsymbol{v}^{(\boldsymbol{\tau})} - \boldsymbol{v}^{(\boldsymbol{\tau}_{m-1})} \right\|^2. \tag{4.38}$$

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

$$\begin{aligned}
\left\langle \boldsymbol{u}^{(\boldsymbol{\tau}+1)}, \boldsymbol{a}^{(\boldsymbol{\tau}+1)} \right\rangle^2 &= \left(\mathop{\mathbb{E}}_{\boldsymbol{x} \sim \gamma} \left[\left\langle \boldsymbol{x} \dot{\otimes} \boldsymbol{x}, \boldsymbol{a}^{(\boldsymbol{\tau}+1)} \right\rangle \,\middle|\, \boldsymbol{x} \in \boldsymbol{X}^{(\boldsymbol{\tau}+1)} \right] \right)^2 \\
&= \left(\left\langle x \dot{\otimes} x, \boldsymbol{a}^{(\boldsymbol{\tau}+1)} \right\rangle + \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \gamma} \left[\varepsilon_{\boldsymbol{x}} \,\middle|\, \boldsymbol{x} \in \boldsymbol{X}^{(\boldsymbol{\tau}+1)} \right] \right)^2 \\
&\qquad\qquad (|\varepsilon_{\boldsymbol{x}}| \le 2^{-L} \text{ is the truncation error by Fact 4.7.4}) \\
&\le \left\langle x \dot{\otimes} x, \boldsymbol{a}^{(\boldsymbol{\tau}+1)} \right\rangle^2 + 2^{-2L} + 2^{1-L} \cdot \left| \left\langle x \dot{\otimes} x, \boldsymbol{a}^{(\boldsymbol{\tau}+1)} \right\rangle \right| \\
&\le \left\langle x \dot{\otimes} x, \boldsymbol{a}^{(\boldsymbol{\tau}+1)} \right\rangle^2 + nT \cdot 2^{2-L},
\end{aligned} \tag{4.39}$$

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

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

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

Then by (4.39), we upper bound the right hand side by

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

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

$$\mathbb{E}\left[\left\langle \boldsymbol{u}^{(\boldsymbol{\tau}+1)} - \boldsymbol{u}^{(\boldsymbol{\tau})}, \boldsymbol{a}^{(\boldsymbol{\tau}+1)} \right\rangle^2 \,\middle|\, \mathcal{F}^{(\boldsymbol{\tau})} \right] \le nT \cdot 2^{2-L} + \lambda \tag{4.40}$$

Putting everything together, we have

$$\begin{aligned}
\mathbb{E}\left[\left(\Delta \boldsymbol{z}_2^{(\boldsymbol{\tau}+1)}\right)^2 \middle| \mathcal{F}^{(\boldsymbol{\tau})}\right] &\le \left(\boldsymbol{\beta}^{(\boldsymbol{\tau}+1)}\right)^2 \mathbb{E}\left[\left\langle \boldsymbol{u}^{(\boldsymbol{\tau}+1)} - \boldsymbol{u}^{(\boldsymbol{\tau})}, \boldsymbol{a}^{(\boldsymbol{\tau}+1)}\right\rangle^2 \middle| \mathcal{F}^{(\boldsymbol{\tau})}\right] + 12n^7T^3 \cdot 2^{-L} \\
&\qquad \text{(by (4.37))} \\
&\le \left(\boldsymbol{\beta}^{(\boldsymbol{\tau}+1)}\right)^2 \cdot \left(nT \cdot 2^{2-L} + \lambda\right) + 12n^7T^3 \cdot 2^{-L} \qquad \text{(by (4.40))} \\
&\le \lambda \cdot \left(\boldsymbol{\beta}^{(\boldsymbol{\tau}+1)}\right)^2 + n^3T^3 \cdot 2^{2-L} + 12n^7T^3 \cdot 2^{-L} \qquad \text{(by (4.35))} \\
&\le \lambda \cdot \left\|\boldsymbol{v}^{(\boldsymbol{\tau})} - \boldsymbol{v}^{(\boldsymbol{\tau}_{m-1})}\right\|^2 + n^3T^3 \cdot 2^{2-L} + 12n^7T^3 \cdot 2^{-L} \qquad \text{(by (4.38))} \\
&\le \lambda \cdot \left\|\boldsymbol{v}^{(\boldsymbol{\tau})} - \boldsymbol{v}^{(\boldsymbol{\tau}_{m-1})}\right\|^2 + 16n^7T^3 \cdot 2^{-L}.
\end{aligned}$$

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

For the moreover part, let us condition on the event $\boldsymbol{\tau}_{m-1} < t < \boldsymbol{\tau}_m$ 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 $\boldsymbol{\tau}_{m-1}$ till t, i.e., $\boldsymbol{v}^{(t+1)} = \boldsymbol{v}^{(\boldsymbol{\tau}_{m-1})}$. Thus we have

$$\Delta \boldsymbol{z}_2^{(t+1)} = \left\langle \boldsymbol{u}^{(t+1)} - \boldsymbol{u}^{(t)}, \eta \odot \boldsymbol{v}^{(\boldsymbol{\tau}_{m-1})}\right\rangle. \tag{4.41}$$

Analogous to (4.33), the component of Alice's center of mass along the previous directions are fixed with precision 2^{-L}. Thus by Fact 4.7.4,

$$\left|\left\langle \boldsymbol{u}^{(t+1)} - \boldsymbol{u}^{(t)}, \boldsymbol{a}^{(r)}\right\rangle\right| \le 2^{-L} \quad \text{for all } r \le t. \tag{4.42}$$

Furthermore, by construction, $\eta \odot \boldsymbol{v}^{(\boldsymbol{\tau}_{m-1})}$ lies in the space spanned by $\boldsymbol{a}^{(1)}, \ldots, \boldsymbol{a}^{(\boldsymbol{\tau}_{m-1}+1)}$. Note that $\boldsymbol{\tau}_{m-1} + 1 \le t$. Similar to the previous analysis, for each $r = 1, \ldots, t$, let $\boldsymbol{\beta}^{(r)} := \left\langle \eta \odot \boldsymbol{v}^{(t)}, \boldsymbol{a}^{(r)}\right\rangle$ be the length of $\eta \odot \boldsymbol{v}^{(t)}$ along direction $\boldsymbol{a}^{(r)}$. Then (4.35) also holds here. Therefore

$$\begin{aligned}
\left|\Delta \boldsymbol{z}_2^{(t+1)}\right| &= \left|\sum_{r=1}^{t} \boldsymbol{\beta}^{(r)} \cdot \left\langle \boldsymbol{u}^{(t+1)} - \boldsymbol{u}^{(t)}, \boldsymbol{a}^{(r)}\right\rangle\right| \qquad \text{(by (4.41))} \\
&\le \sum_{r=1}^{t} \left|\boldsymbol{\beta}^{(r)}\right| \cdot \left|\left\langle \boldsymbol{u}^{(t+1)} - \boldsymbol{u}^{(t)}, \boldsymbol{a}^{(r)}\right\rangle\right| \le \sum_{r=1}^{t} nT \cdot 2^{-L} \qquad \text{(by (4.35) and (4.42))} \\
&\le 2n^3T \cdot 2^{-L}. \qquad \text{(by Claim 4.7.7)}
\end{aligned}$$

Conversion to Bounds on Depth (Proof of Lemma 4.7.10)

Recall $\gamma^* = \gamma(\boldsymbol{X}^{(0)}) \times \gamma(\boldsymbol{Y}^{(0)})$ and $\gamma(\ell) = \gamma(\boldsymbol{X}^{(D(\ell))}) \cdot \gamma(\boldsymbol{Y}^{(D(\ell))})$ for each leaf ℓ. The goal of this subsection is to prove Lemma 4.7.10.

We first note the following basic fact.

Fact 4.7.13. $\sum_\ell \gamma(\ell) = \gamma^*$ *and*

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

Now we apply Theorem 2.0.6 with $k = 2$ to relate the LHS of Lemma 4.7.10 with an entropy-type bound.

Lemma 4.7.14. $\mathbb{E}\left[\left\|\boldsymbol{u}^{(\boldsymbol{d})}\right\|_{\mathrm{F}}^2 + \left\|\boldsymbol{v}^{(\boldsymbol{d})}\right\|_{\mathrm{F}}^2\right] \le \frac{4e^2}{\gamma^*}\sum_\ell \gamma(\ell)\cdot\ln^2\left(\frac{e}{\gamma(\ell)}\right)$.

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 $1_\ell\colon \mathbb{R}^{2n}\to\{0,1\}$ by

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

Then we have

$$\begin{aligned}
&\left\|u^{(D)}\right\|_{\mathrm{F}}^2 + \left\|v^{(D)}\right\|_{\mathrm{F}}^2\\
&= \left\|\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma}\left[\boldsymbol{x}\mathbin{\dot\otimes}\boldsymbol{x}\,\middle|\,\boldsymbol{x}\in X^{(D)}\right]\right\|_{\mathrm{F}}^2 + \left\|\mathop{\mathbb{E}}_{\boldsymbol{y}\sim\gamma}\left[\boldsymbol{y}\mathbin{\dot\otimes}\boldsymbol{y}\,\middle|\,\boldsymbol{y}\in Y^{(D)}\right]\right\|_{\mathrm{F}}^2\\
&= \sum_{\substack{i,j=1\\ i\neq j}}^n\left(\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma}\left[\boldsymbol{x}_i\boldsymbol{x}_j\,\middle|\,\boldsymbol{x}\in X^{(D)}\right]\right)^2 + \sum_{\substack{i,j=1\\ i\neq j}}^n\left(\mathop{\mathbb{E}}_{\boldsymbol{y}\sim\gamma}\left[\boldsymbol{y}_i\boldsymbol{y}_j\,\middle|\,\boldsymbol{y}\in Y^{(D)}\right]\right)^2\\
&= \sum_{\substack{i,j=1\\ i\neq j}}^n\left(\mathop{\mathbb{E}}_{\boldsymbol{x},\boldsymbol{y}\sim\gamma}\left[\boldsymbol{x}_i\boldsymbol{x}_j\,\middle|\,(\boldsymbol{x},\boldsymbol{y})\in X^{(D)}\times Y^{(D)}\right]\right)^2 + \sum_{\substack{i,j=1\\ i\neq j}}^n\left(\mathop{\mathbb{E}}_{\boldsymbol{x},\boldsymbol{y}\sim\gamma}\left[\boldsymbol{y}_i\boldsymbol{y}_j\,\middle|\,(\boldsymbol{x},\boldsymbol{y})\in X^{(D)}\times Y^{(D)}\right]\right)^2\\
&= \frac{2}{\gamma(\ell)^2}\left(\sum_{S\in\binom{[n]}{2}}\left(\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma,\boldsymbol{y}\sim\gamma}\left[1_\ell(\boldsymbol{x},\boldsymbol{y})\boldsymbol{x}_S\right]\right)^2 + \sum_{S\in\binom{[n]}{2}}\left(\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma,\boldsymbol{y}\sim\gamma}\left[1_\ell(\boldsymbol{x},\boldsymbol{y})\boldsymbol{y}_S\right]\right)^2\right)\\
&\le \frac{2}{\gamma(\ell)^2}\sum_{S\in\binom{[2n]}{2}}\left(\mathop{\mathbb{E}}_{\boldsymbol{w}\sim\gamma_n\times\gamma_n}\left[1_\ell(\boldsymbol{w})\boldsymbol{w}_S\right]\right)^2\\
&\le \frac{2}{\gamma(\ell)^2}\cdot 2e^2\gamma(\ell)^2\cdot\ln^2\left(\frac{e}{\gamma(\ell)}\right) \qquad \text{(by Theorem 2.0.6)}\\
&= 4e^2\cdot\ln^2\left(\frac{e}{\gamma(\ell)}\right).
\end{aligned}$$

Therefore taking expectation over a random ℓ, by Fact 4.7.13, we have

$$\mathbb{E}\left[\left\|\boldsymbol{u}^{(\boldsymbol{d})}\right\|_{\mathrm{F}}^2 + \left\|\boldsymbol{v}^{(\boldsymbol{d})}\right\|_{\mathrm{F}}^2\right] \le 4e^2\cdot\mathop{\mathbb{E}}_{\boldsymbol{\ell}}\left[\ln^2\left(\frac{e}{\gamma(\boldsymbol{\ell})}\right)\right] = \frac{4e^2}{\gamma^*}\sum_\ell \gamma(\ell)\cdot\ln^2\left(\frac{e}{\gamma(\ell)}\right). \qquad \square$$

Now in the next lemma, we bound the right hand side of Lemma 4.7.14 in terms of the second moment of the depth, which immediately proves Lemma 4.7.10.

Lemma 4.7.15. *Assume that $Tn \le 2^L$. Then, $\sum_\ell \gamma(\ell) \cdot \ln^2(e/\gamma(\ell)) \le O(1 + \gamma^* \cdot L^2\, \mathbb{E}[\boldsymbol{d}^2])$.*

Proof. By Claim 4.7.6, and the assumption $Tn \le 2^L$ each message is of length at most $L + \log(Tn) \le 2L$. We divide ℓ into two cases based on $\gamma(\ell)$:

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

and

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

Adding up the two estimates above gives the desired bound. □

Second Moment Bounds for the Depth (Proof of Lemma 4.7.11)

The final ingredient is an estimate for the second moment $\mathbb{E}[\boldsymbol{d}^2]$. This subsection is devoted to this goal and proving Lemma 4.7.11.

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

Lemma 4.7.16. *There exists a universal constant $\alpha > 0$ such that the following holds. Let $0 \le d_1 < d_2$ be two arbitrary integers with $d_2 - d_1 \ge 2d + 1$. Let $\ell^* = (\overline{\boldsymbol{c}}^{(1)}, \dots, \overline{\boldsymbol{c}}^{(d_1)})$ be arbitrary messages of the first d_1 communication steps. Assume $2^L \ge 8n^4T^2$. Then*

$$\mathbf{Pr}\left[\boldsymbol{d} \ge d_2 \,|\, \ell^*\right] \le \frac{\alpha \cdot d_2^2 L^2}{\lambda \cdot (d_2 - d_1 - 2d)} + \frac{1}{4} \cdot \frac{2^{-3L \cdot d_1}}{\gamma(\ell^*)}.$$

Proof. Let $\boldsymbol{x}, \boldsymbol{y}$ be sampled from γ conditioned on $\boldsymbol{x} \in \boldsymbol{X}^{(0)}, \boldsymbol{y} \in \boldsymbol{Y}^{(0)}$. Let $\boldsymbol{\ell}$ be its corresponding leaf in $\overline{\mathcal{C}}$ and $\boldsymbol{d}$ be the depth of $\boldsymbol{\ell}$. By Claim 4.7.7, $\boldsymbol{\ell}$ always has finite depth. We extend $\boldsymbol{a}^{(t)} = \boldsymbol{b}^{(t)} = 0^{n \times n}$ and $\boldsymbol{X}^{(t)} = \boldsymbol{X}^{(\boldsymbol{d})}, \boldsymbol{Y}^{(t)} = \boldsymbol{Y}^{(\boldsymbol{d})}$ for all $t > \boldsymbol{d}$. Then define

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

where $\boldsymbol{a}^{(\cdot)}$'s and $\boldsymbol{b}^{(\cdot)}$'s depend only on $\boldsymbol{\ell}$.[12] Equivalently, we can write K as

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

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

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

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

Note that we can further expand $\boldsymbol{\ell}'$ to a leaf $\boldsymbol{\ell}$ as a full communication path, and obtain the following equivalent sampling process:

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

As a result, we have

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

[12] Note that $\boldsymbol{\ell}$ specifies all the communication messages, which allows us to simulate the protocol and obtain each $\boldsymbol{a}^{(\cdot)}$ and $\boldsymbol{b}^{(\cdot)}$.

Observe that there are at most $2d$ many step 3(a) and 3(b) in $\boldsymbol{\ell}$. This means, if $\boldsymbol{d} \geq d_2$, 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

$$K \geq \lambda \cdot (d_2 - d_1 - 2d) \cdot \mathbf{Pr}\left[\boldsymbol{d} \geq d_2 \,|\, \ell^*\right]. \tag{4.43}$$

On the other hand by Claim 4.7.7, there are at most n^2 non-zero $\boldsymbol{a}^{(\cdot)}$'s and at most n^2 non-zero $\boldsymbol{b}^{(\cdot)}$'s in each communication path. Thus

$$\boldsymbol{k}(\boldsymbol{x},\boldsymbol{y}) \leq n^2 \cdot \left(\max_{x \in \boldsymbol{X}^{(0)}} \left\| x \dot{\otimes} x \right\|_{\mathrm{F}}^2 + \max_{y \in \boldsymbol{Y}^{(0)}} \left\| y \dot{\otimes} y \right\|_{\mathrm{F}}^2 \right) < 2n^4 T^2. \tag{4.44}$$

We now obtain another upper bound using Theorem 4.4.1. Let $\overline{\boldsymbol{\ell}} = (\overline{\boldsymbol{c}}^{(1)}, \ldots, \overline{\boldsymbol{c}}^{(d_2)})$ extend ℓ^* for the next $d_2 - d_1$ messages.[13] Then $K = \mathbb{E}_{\overline{\boldsymbol{\ell}}}\left[\boldsymbol{k}(\overline{\boldsymbol{\ell}}) \,\middle|\, \ell^*\right]$ where $\boldsymbol{k}(\overline{\ell}) := \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}\sim\gamma}\left[\boldsymbol{k}(\boldsymbol{x},\boldsymbol{y}) \,\middle|\, \overline{\ell}\right]$. Note that $\overline{\ell}$ fixes $a^{(\cdot)}$'s and $b^{(\cdot)}$'s in $\boldsymbol{k}(\boldsymbol{x},\boldsymbol{y})$. Therefore we use $\boldsymbol{k}_{\overline{\ell}}(\boldsymbol{x},\boldsymbol{y})$ to denote $\boldsymbol{k}(\boldsymbol{x},\boldsymbol{y})$ with the directions $a^{(\cdot)}$'s and $b^{(\cdot)}$'s fixed by $\overline{\ell}$. We now continue the bound on $\boldsymbol{k}(\overline{\ell})$:

$$\begin{aligned}
\boldsymbol{k}(\overline{\ell}) &\leq \sum_{t=0}^{\infty} \mathop{\mathbf{Pr}}_{\boldsymbol{x},\boldsymbol{y}\sim\gamma}\left[\boldsymbol{k}_{\overline{\ell}}(\boldsymbol{x},\boldsymbol{y}) \geq t \,\middle|\, \overline{\ell}\right] = \sum_{t=0}^{\infty} \frac{\mathbf{Pr}_{\boldsymbol{x},\boldsymbol{y}\sim\gamma}\left[\boldsymbol{k}_{\overline{\ell}}(\boldsymbol{x},\boldsymbol{y}) \geq t, \overline{\ell}\right]}{\mathbf{Pr}_{\boldsymbol{x},\boldsymbol{y}\sim\gamma}\left[\overline{\ell}\right]} \\
&= \sum_{t=0}^{\infty} \min\left\{1, \frac{\mathbf{Pr}_{\boldsymbol{x},\boldsymbol{y}\sim\gamma}\left[\boldsymbol{k}_{\overline{\ell}}(\boldsymbol{x},\boldsymbol{y}) \geq t, \overline{\ell}\right]}{\gamma(\overline{\ell})}\right\} \qquad \text{(by the definition of } \gamma(\cdot)) \\
&\leq \sum_{t=0}^{\infty} \min\left\{1, \frac{\mathbf{Pr}_{\boldsymbol{x},\boldsymbol{y}\sim\gamma}\left[\boldsymbol{k}_{\overline{\ell}}(\boldsymbol{x},\boldsymbol{y}) \geq t\right]}{\gamma(\overline{\ell})}\right\}.
\end{aligned} \tag{4.45}$$

We now analyze $\mathbf{Pr}_{\boldsymbol{x},\boldsymbol{y}\sim\gamma}\left[\boldsymbol{k}_{\overline{\ell}}(\boldsymbol{x},\boldsymbol{y}) \geq t\right]$ using Theorem 4.4.1. Since $a^{(t)}, b^{(t)}$ cannot be non-zero simultaneously, we rearrange the matrices and assume $a^{(d_1+1)}, \ldots, a^{(d')}, b^{(d'+1)}, \ldots, b^{(d'')}$ are the only non-zero matrices where $d'' \leq d_2$. Then

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

Note that a's (resp., b's) satisfy the condition in Theorem 4.4.1. Let $1/\kappa$ be the constant[14] in Ω in Theorem 4.4.1. Hence

$$\begin{aligned}
\mathbf{Pr}\left[\boldsymbol{k}_{\overline{\ell}}(\boldsymbol{x},\boldsymbol{y}) \geq t\right] &\leq \mathbf{Pr}\left[\sum_{t=d_1+1}^{d'} \left\langle \boldsymbol{x} \dot{\otimes} \boldsymbol{x}, a^{(t)} \right\rangle^2 \geq t/2\right] + \mathbf{Pr}\left[\sum_{t=d'+1}^{d''} \left\langle \boldsymbol{y} \dot{\otimes} \boldsymbol{y}, b^{(t)} \right\rangle^2 \geq t/2\right] \\
&\leq 2\exp\left\{-\frac{1}{\kappa} \cdot \frac{t/2}{d' - d_1 + \sqrt{t/2}}\right\} + 2\exp\left\{-\frac{1}{\kappa} \cdot \frac{t/2}{d'' - d' + \sqrt{t/2}}\right\} \\
&\qquad \text{(by Theorem 4.4.1 and assuming } t \geq 196 \cdot \max\{d' - d_1, d'' - d'\})
\end{aligned}$$

[13] If $\overline{\boldsymbol{\ell}}$ becomes a leaf before d_2, then we can simply pad dummy messages to it.

[14] In particular $\kappa = 56448$ suffices from our proof in Section 4.4.

$$\leq 4\exp\left\{-\frac{1}{\kappa}\cdot\frac{t/2}{d_2-d_1+\sqrt{t/2}}\right\}. \qquad \text{(since } d_1\leq d'\leq d''\leq d_2\text{)}$$

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

$$\mathbf{Pr}\left[\boldsymbol{k}_{\bar{\ell}}(\boldsymbol{x},\boldsymbol{y})\geq t\right]\leq 4\exp\left\{-\frac{1}{\kappa}\cdot\frac{t/2}{d_2-d_1+\sqrt{t/2}}\right\}. \tag{4.46}$$

For $\gamma(\bar{\ell})\geq 2^{-3L\cdot d_2}$, we plug (4.46) into (4.45) and obtain

$$\begin{aligned}
\boldsymbol{k}(\bar{\ell}) &\leq \sum_{t=0}^{196\cdot(d_2-d_1)^2} 1+\sum_{t>196\cdot(d_2-d_1)^2}\min\left\{1,2^{3L\cdot d_2+1}\cdot\exp\left\{-\frac{1}{\kappa}\cdot\frac{t/2}{d_2-d_1+\sqrt{t/2}}\right\}\right\} \qquad \text{(by (4.46))}\\
&\leq 196\cdot(d_2-d_1)^2+1+\sum_{t\geq 196\cdot(d_2-d_1)^2}\min\left\{1,2^{3L\cdot d_2+1}\cdot e^{-\frac{1}{\kappa}\cdot\frac{t/2}{2\sqrt{t/2}}}\right\}\\
&\leq 197\cdot d_2^2+\sum_{t\geq 1}\min\left\{1,2^{3L\cdot d_2+1}\cdot e^{-\frac{\sqrt{t/2}}{2\kappa}}\right\}\\
&\leq \alpha\cdot d_2^2L^2, \qquad (4.47)
\end{aligned}$$

where α is another universal constant. Now we have

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

where the first summation can be bounded by

$$\begin{aligned}
\sum_{\bar{\ell}:\gamma(\bar{\ell})<2^{-3L\cdot d_2}}\frac{\gamma(\bar{\ell})}{\gamma(\ell^*)}\cdot\boldsymbol{k}(\bar{\ell}) &\leq \frac{2^{-3L\cdot d_1}}{\gamma(\ell^*)}\cdot\sum_{\bar{\ell}}2^{-3L\cdot(d_2-d_1)}\cdot n^4T^2 \qquad \text{(by (4.44))}\\
&\leq \frac{2^{-3L\cdot d_1}}{\gamma(\ell^*)}\cdot 2^{2L\cdot(d_2-d_1)}\cdot 2^{-3L\cdot(d_2-d_1)}\cdot n^4T^2\\
&\quad \text{(since } \ell^* \text{ is fixed and each message is at most } 2L \text{ bits)}\\
&= \frac{2^{-3L\cdot d_1}}{\gamma(\ell^*)}\cdot\frac{2n^4T^2}{2^L} \qquad \text{(since } d_2-d_1\geq 1\text{)}
\end{aligned}$$

and the second summation is bounded by

$$\sum_{\bar{\ell}:\gamma(\bar{\ell})\geq 2^{-3L\cdot d_2}}\frac{\gamma(\bar{\ell})}{\gamma(\ell^*)}\cdot\boldsymbol{k}(\bar{\ell})\leq\sum_{\bar{\ell}}\frac{\gamma(\bar{\ell})}{\gamma(\ell^*)}\cdot\alpha\cdot d_2^2L^2=\alpha\cdot d_2^2L^2. \qquad \text{(by (4.47))}$$

Then combining (4.43), we have

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

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

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

Corollary 4.7.17. *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 $k = 0, 1, \ldots, 4\log(n)$, we have*

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

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

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

and

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

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 4.7.16 with ℓ^* and $d_1 = 4(k-1)d, d_2 = 4kd$, we have

$$\begin{aligned} Q &\leq \sum_{\ell^* : \gamma(\ell^*)/\gamma^* \geq 2^{-3L \cdot 4(k-1)d}} \frac{\gamma(\ell^*)}{\gamma^*} \cdot \left(\frac{16\alpha \cdot k^2 d^2 L^2}{2dR} + \frac{1}{4} \cdot \frac{2^{-3L \cdot 4(k-1)d}}{\gamma(\ell^*)} \right) \\ &\leq \sum_{\ell^*} \frac{\gamma(\ell^*)}{\gamma^*} \cdot \left(\frac{8\alpha \cdot k^2 dL^2}{\lambda} + \frac{1}{4\gamma^*} \right) \\ &= \mathbf{Pr}\left[\boldsymbol{d} \geq 4(k-1)d\right] \cdot \left(\frac{8\alpha \cdot k^2 dL^2}{\lambda} + \frac{1}{4\gamma^*} \right) \end{aligned}$$

$$\leq \mathbf{Pr}\left[\boldsymbol{d} \geq 4(k-1)d\right] \cdot \frac{1}{2} \qquad \text{(since } \gamma^* \geq 3/4 \text{ and } \lambda \geq \Theta(dL^2\log^2(n)), k \leq 4\log(n)\text{)}$$

$$\leq \left(2^{-(k-1)} + \frac{k-1}{n^5}\right) \cdot \frac{1}{2} \leq 2^{-k} + \frac{k-1}{n^5}. \qquad \text{(by induction hypothesis)}$$

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

Given Corollary 4.7.17 and suitable choice of the parameters, we now prove the second moment bound.

Proof of Lemma 4.7.11. With $L = \Theta(\log(n))$, $T = \Theta(\sqrt{\log(n)})$, and $\lambda = \Theta(d\log^4(n))$, by Fact 4.7.3, we have $\gamma^* \geq 3/4$. Therefore the second moment of $\boldsymbol{d}$ is

$$\mathbb{E}[\boldsymbol{d}^2] \leq \sum_{k=0}^{4\log(n)} (4(k+1)d)^2 \cdot \mathbf{Pr}\left[\boldsymbol{d} \geq 4kd\right] + \mathbf{Pr}\left[\boldsymbol{d} \geq 16d\log(n)\right] \cdot (2n^2)^2 \quad \text{(by Claim 4.7.7)}$$

$$\leq \sum_{k=0}^{4\log(n)} (4(k+1)d)^2 \cdot \left(2^{-k} + \frac{k}{n^5}\right) + \left(n^{-4} + \frac{4\log(n)}{n^5}\right) \cdot (2n^2)^2$$

(by Corollary 4.7.17)

$$= O(d^2).$$ □

4.8 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 a, b be two positive integers. Let $g\colon \{\pm1\}^a \times \{\pm1\}^b \to \{\pm1\}$ be a gadget. We define the g-fiber of communication protocols similar to the XOR-fiber:

Definition 4.8.1. *For any randomized two-party protocol $\mathcal{C}\colon (\{\pm1\}^a)^n \times (\{\pm1\}^b)^n \to [-1,1]$, its g-fiber, denoted by $\mathcal{C}_{\downarrow g}\colon \{\pm1\}^n \to [-1,1]$, is defined by*

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

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

To compare the Fourier growth bounds between gadgets, we use $L_k(g,d,a,b,n)$ to denote the upper bound of the level-k Fourier growth for the g-fiber of an arbitrary randomized communication protocol $\mathcal{C}\colon (\{\pm1\}^a)^n \times (\{\pm1\}^b)^n \to [-1,1]$ with at most d bits of communication, where $g\colon \{\pm1\}^a \times \{\pm1\}^b \to \{\pm1\}$ is the gadget. Since randomized protocols are convex combinations of deterministic protocols of the same cost, using this notation, our main results Theorems 4.2.1 and 4.2.2 can be rephrased as

$$L_1(\mathrm{XOR}, d, 1, 1, n) \leq O\left(\sqrt{d}\right) \quad \text{and} \quad L_2(\mathrm{XOR}, d, 1, 1, n) \leq O\left(d^{3/2}\log^3(n)\right).$$

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

$$g(x,y) = \sum_{S \subseteq [a], T \subseteq [b]} \widehat{g}(S,T) \cdot x_S y_T, \quad \text{where} \quad \widehat{g}(S,T) = \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mathcal{U}_a, \boldsymbol{y} \sim \mathcal{U}_b} [g(\boldsymbol{x}, \boldsymbol{y}) \cdot \boldsymbol{x}_S \boldsymbol{y}_T] .$$

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

Assumption 4.8.2. *$\widehat{g}(S,T) = 0$ if $S = \emptyset$ or $T = \emptyset$.*

Remark 4.8.3. *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 $g' \colon \{\pm 1\}^{a+1} \times \{\pm 1\}^{b+1} \to \{\pm 1\}$, where we XOR the last bit of Alice and the last bit of Bob to the old gadget g applied to Alice's first a bits and Bob's first b bits, i.e.,

$$g'(x,y) = x_{a+1} y_{b+1} \cdot g(x_{\le a}, y_{\le b}).$$

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

Now for a protocol $\mathcal{C} \colon (\{\pm 1\}^a)^n \times (\{\pm 1\}^b)^n \to [-1,1]$, it is also a two-party function and thus admitting similar Fourier representation. We view an input from $(\{\pm 1\}^a)^n$ as indexed by a tuple in $[n] \times [a]$. Therefore any subset of $(\{\pm 1\}^a)^n$ is uniquely identified as $\bigcup_{i \in [n]} \{i\} \times S_i$, where each $S_i \subseteq [a]$. 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

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

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

Under this notation and assuming Assumption 4.8.2, we can effectively compute the Fourier coefficients of any g-fiber.

Fact 4.8.4. *Assume gadget $g \colon \{\pm 1\}^a \times \{\pm 1\}^b \to \{\pm 1\}$ satisfies Assumption 4.8.2. 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

$$\begin{aligned}\widehat{\mathcal{C}_{\downarrow g}}(I) &= \mathop{\mathbb{E}}_{\boldsymbol{z}\sim\mathcal{U}_n}\left[\mathcal{C}_{\downarrow g}(\boldsymbol{z})\cdot\prod_{i\in I}\boldsymbol{z}_i\right]\\ &= \mathop{\mathbb{E}}_{\boldsymbol{z}\sim\mathcal{U}_n}\left[\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mathcal{U}_{an},\boldsymbol{y}\sim\mathcal{U}_{bn}}[\mathcal{C}(\boldsymbol{x},\boldsymbol{y})\,|\,g(\boldsymbol{x}_i,\boldsymbol{y}_i)=\boldsymbol{z}_i,\ \forall i]\cdot\prod_{i\in I}\boldsymbol{z}_i\right]\\ &= \mathop{\mathbb{E}}_{\boldsymbol{z}\sim\mathcal{U}_n}\left[\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mathcal{U}_{an},\boldsymbol{y}\sim\mathcal{U}_{bn}}\left[\mathcal{C}(\boldsymbol{x},\boldsymbol{y})\cdot\prod_{i\in I}g(\boldsymbol{x}_i,\boldsymbol{y}_i)\,\middle|\,g(\boldsymbol{x}_i,\boldsymbol{y}_i)=\boldsymbol{z}_i,\ \forall i\right]\right].\end{aligned}$$

Since $\widehat{g}(\emptyset,\emptyset)=0$ by Assumption 4.8.2, every pair (x,y) is sampled with the same probability under the conditional distribution. Thus we get

$$\widehat{\mathcal{C}_{\downarrow g}}(I) = \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mathcal{U}_{an},\boldsymbol{y}\sim\mathcal{U}_{bn}}\left[\mathcal{C}(\boldsymbol{x},\boldsymbol{y})\cdot\prod_{i\in I}g(\boldsymbol{x}_i,\boldsymbol{y}_i)\right].$$

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

$$\begin{aligned}&\widehat{\mathcal{C}_{\downarrow g}}(I)\\ &= \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mathcal{U}_{an},\boldsymbol{y}\sim\mathcal{U}_{bn}}\left[\left(\sum_{S^{[n]},T^{[n]}}\widehat{\mathcal{C}}(S^{[n]},T^{[n]})\prod_{i\in[n]}\boldsymbol{x}_{i,S_i}\prod_{j\in[n]}\boldsymbol{y}_{j,T_j}\right)\cdot\prod_{i\in I}\left(\sum_{S_i,T_i}\widehat{g}(S_i,T_i)\boldsymbol{x}_{i,S_i}\boldsymbol{y}_{i,T_i}\right)\right]\\ &= \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mathcal{U}_{an},\boldsymbol{y}\sim\mathcal{U}_{bn}}\left[\left(\sum_{S^{[n]},T^{[n]}}\widehat{\mathcal{C}}(S^{[n]},T^{[n]})\prod_{i\in[n]}\boldsymbol{x}_{i,S_i}\prod_{j\in[n]}\boldsymbol{y}_{j,T_j}\right)\left(\sum_{S^I,T^I}\prod_{i\in I}\widehat{g}(S_i,T_i)\boldsymbol{x}_{i,S_i}\boldsymbol{y}_{i,T_i}\right)\right]\\ &= \sum_{S^I,T^I}\widehat{\mathcal{C}}(S^I,T^I)\cdot\prod_{i\in I}\widehat{g}(S_i,T_i)\\ &= \sum_{\substack{S^I,T^I\\ S_i\neq\emptyset,T_i\neq\emptyset,\forall i\in I}}\widehat{\mathcal{C}}(S^I,T^I)\cdot\prod_{i\in I}\widehat{g}(S_i,T_i), \qquad \text{(by Assumption 4.8.2)}\end{aligned}$$

as desired. □

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

Theorem 4.8.5. *Assume gadget $g\colon\{\pm1\}^a\times\{\pm1\}^b\to\{\pm1\}$ satisfies Assumption 4.8.2. Then*

$$\begin{aligned}L_k(\mathrm{XOR},d,1,1,n) &\le \left(\max_{S,T}|\widehat{g}(S,T)|\right)^{-k}\cdot L_k(g,d,a,b,n)\\ &\le 2^{(a+b)\cdot k/2}\cdot L_k(g,d,a,b,n).\end{aligned}$$

Proof. Let $\mathcal{C}\colon \{\pm1\}^n \times \{\pm1\}^n \to [-1,1]$ be an arbitrary protocol of cost at most d. Then for a fixed set $I \subseteq [n]$, by Fact 4.8.4 applied to the XOR gadget, we have

$$\widehat{\mathcal{C}_{\downarrow \mathrm{XOR}}}(I) = \widehat{\mathcal{C}}(1^I, 1^I). \tag{4.48}$$

Let $S \subseteq [a]$ and $T \subseteq [b]$ maximize $|\widehat{g}(S,T)|$. Since g satisfies Assumption 4.8.2, we know S and T are not empty sets.

Now define a different protocol $\mathcal{C}'\colon (\{\pm1\}^a)^n \times (\{\pm1\}^b)^n \to [-1,1]$ 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 (4.48) and Fact 4.8.4 applied to $\mathcal{C}'$ with gadget g, we have

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

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

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

Since $\mathcal{C}$ is arbitrary, this proves the first half of Theorem 4.8.5. To prove the second half, we use an averaging argument and Parseval's identity on g:

$$|\widehat{g}(S,T)| \geq \sqrt{2^{-a-b} \sum_{S',T'} \widehat{g}(S',T')^2} = \sqrt{2^{-a-b}}. \qquad \square$$

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

Theorem 4.8.6. *Assume gadget $g\colon \{\pm1\}^a \times \{\pm1\}^b \to \{\pm1\}$ satisfies Assumption 4.8.2. Then*

$$\begin{aligned} L_k(g,d,a,b,n) &\leq \left(\sum_{S,T} |\widehat{g}(S,T)|\right)^k \cdot L_k(\mathrm{XOR}, d, 1, 1, n) \\ &\leq 2^{(a+b)\cdot k/2} \cdot L_k(\mathrm{XOR}, d, 1, 1, n). \end{aligned}$$

Proof. Let $\mathcal{C}\colon (\{\pm1\}^a)^n \times (\{\pm1\}^b)^n \to [-1,1]$ be an arbitrary protocol of cost at most d. Then for a fixed set $I \subseteq [n]$, by Fact 4.8.4 applied to gadget g and using Assumption 4.8.2, 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

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

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

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

Then we can rewrite $L_k(\mathcal{C}_{\downarrow g})$ as

$$\begin{aligned} L_k(\mathcal{C}_{\downarrow g}) &\le \sum_{I\subseteq[n]:|I|=k} \mathop{\mathbb{E}}_{(\boldsymbol{S}^I,\boldsymbol{T}^I)\sim\rho^I} \left[\left|\widehat{\mathcal{C}}(\boldsymbol{S}^I,\boldsymbol{T}^I)\right| \cdot M^k\right] \\ &= M^k \cdot \mathop{\mathbb{E}}_{(\boldsymbol{S}^{[n]},\boldsymbol{T}^{[n]})\sim\rho^{[n]}} \left[\sum_{I\subseteq[n]:|I|=k} \left|\widehat{\mathcal{C}}(\boldsymbol{S}^I,\boldsymbol{T}^I)\right|\right]. \end{aligned} \tag{4.49}$$

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 4.8.2. Then define a different protocol $\mathcal{C}'\colon \{\pm1\}^n \times \{\pm1\}^n \to [-1,1]$ as follows: After receiving input x, Alice samples $x' \in (\{\pm1\}^a)^n$ uniformly conditioned on $x'_{i,S_i} = x_i$ for all $i\in[n]$; and Bob samples similarly $y' \in (\{\pm1\}^b)^n$ conditioned on $y'_{i,T_i} = y_i$ for all $i\in[n]$. Then they execute the protocol $\mathcal{C}$ on x' and y'. That is, $\mathcal{C}'(x,y) = \mathbb{E}_{\boldsymbol{x}',\boldsymbol{y}'}[\mathcal{C}(\boldsymbol{x}',\boldsymbol{y}')]$. Therefore, for any $I\subseteq[n]$, we have

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

By Fact 4.8.4 applied to $\mathcal{C}'$ and the XOR gadget, we have

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

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

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

Putting back to (4.49), we have

$$L_k(\mathcal{C}_{\downarrow g}) \le M^k \cdot L_k(\mathrm{XOR},d,1,1,n),$$

which proves the first half of Theorem 4.8.6 since $\mathcal{C}$ is arbitrary. To prove the second half, we use Cauchy-Schwarz inequality and Parseval's identity on g:

$$M = \sum_{S,T} |\widehat{g}(S,T)| \le \sqrt{2^{a+b}\sum_{S,T} \widehat{g}(S,T)^2} = \sqrt{2^{a+b}}. \qquad \square$$

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

Corollary 4.8.7. *Assume Assumption 4.8.2 holds for gadgets* $g\colon \{\pm1\}^a \times \{\pm1\}^b \to \{\pm1\}$ *and* $g'\colon \{\pm1\}^{a'} \times \{\pm1\}^{b'} \to \{\pm1\}$. *Then*

$$L_k(g, d, a, b, n) \le 2^{(a+b+a'+b')\cdot k/2} \cdot L_k(g', d, a', b', n).$$

4.9 Directions Towards Further Improvements

In this section we propose potential directions for further improving our second level bounds. In one approach, 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 4.8. In another direction, we examine the bottleneck in our analysis and identify major obstacles within.

Better Lifting Theorems Imply Better Fourier Growth

Let $f : \{\pm1\}^n \to \{\pm1\}$ be a Boolean function. Let $g : \{\pm1\}^a \times \{\pm1\}^b \to \{\pm1\}$ 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 $\mathcal{C} : (\{\pm1\}^a)^n \times (\{\pm1\}^b)^n \to [-1, 1]$ be a randomized two-party protocol. Recall Definition 4.8.1, the g-fiber of $\mathcal{C}$, denoted $\mathcal{C}_{\downarrow g}(z) : \{\pm1\}^n \to [-1, 1]$, is defined by

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

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 $1/\mathsf{poly}(n)$ (see e.g., [GPW20]).

Theorem 4.9.1. *Assume gadget* $g\colon \{\pm1\}^a \times \{\pm1\}^b \to \{\pm1\}$ *satisfies Assumption 4.8.2. Assume for any randomized protocol* $\mathcal{C}\colon (\{\pm1\}^a)^n \times (\{\pm1\}^b)^n \to [-1, 1]$ *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.,*

$$|\mathcal{T}(z) - \mathcal{C}_{\downarrow g}(z)| \le n^{-k} \quad \forall z \in \{\pm1\}^n.$$

Then, for any randomized protocol $\mathcal{C}'\colon \{\pm1\}^n \times \{\pm1\}^n \to [-1, 1]$ *with at most* d *bits of communication, its XOR-fiber* $\mathcal{C}'_{\downarrow\mathrm{XOR}}$ *has level-*k *Fourier growth*

$$\begin{aligned} L_k(\mathcal{C}'_{\downarrow\mathrm{XOR}}) &\le \left(\max_{S,T} |\widehat{g}(S, T)|\right)^{-k} \cdot \sqrt{D^k \cdot O\left(\log(n)\right)^{k-1}} \\ &\le 2^{(a+b)\cdot k/2} \cdot \sqrt{D^k \cdot O\left(\log(n)\right)^{k-1}}. \end{aligned}$$

As a simple corollary, we see that if the assumption of Theorem 4.9.1 holds with $k = 2$, $D = d \cdot \mathsf{polylog}(n)$, and a polylogarithmic-sized gadget g (i.e., $2^a, 2^b \leq \mathsf{polylog}(n)$), then the second level Fourier growth of the XOR-fiber of any randomized protocol of cost d is at most $d \cdot \mathsf{polylog}(n)$ as desired. In addition, by a majority vote of $O(k \log(n))$ independent copies of the randomized decision tree, the error condition above can be relaxed to any small constant.

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

- The inner product on $a = b = O(\log(n))$ bits [CFK+19]. However, for such g the largest Fourier coefficient squared is $1/\mathsf{poly}(n)$, which yields a trivial bound in Theorem 4.9.1.
- The index function with $a = \mathsf{poly}(n)$, $b = \log(a)$ [GPW20].[15] In this case the largest Fourier coefficient squared is $1/a^2$, which again yields a trivial bound in Theorem 4.9.1. Nonetheless, even a polynomial improvement on a, say $a = n^{0.01}$, would give new non-trivial bounds in Theorem 4.9.1 and in turn improves our lower bound on the XOR-lift of Forrelation.

Proof of Theorem 4.9.1. Let $\mathcal{C} \colon (\{\pm 1\}^a)^n \times (\{\pm 1\}^b)^n \to [-1, 1]$ 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 $1/n^k$. Therefore by the level-k Fourier growth bounds on randomized decision trees [Tal20, SSW23], we have

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

Since $\mathcal{C}$ is arbitrary, the claimed bound for $\mathcal{C}'_{\downarrow \mathrm{XOR}}$ follows from Theorem 4.8.5. □

Sums of Squares of Quadratic Forms for Pairwise Clean Sets

In our analysis for the level-2 bound, we showed that one can transform a general protocol to a 4-wise clean protocol with parameter $\lambda = d \cdot \mathsf{polylog}(n)$ by adding $O(d)$ additional cleanup steps in expectation. If one could show that with essentially the same number of steps, one could take $\lambda = \mathsf{polylog}(n)$, then we would obtain the optimal level-2 bound of $d \cdot \mathsf{polylog}(n)$.

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 4.4.1), which says that if $M_1, \ldots, M_m$ are matrices with zero diagonal and form an orthonormal set when viewed as n^2 dimensional vectors, then the random variable $\boldsymbol{q} = \sum_{i=1}^m \left\langle \boldsymbol{x} \mathbin{\dot{\otimes}} \boldsymbol{x}, M_i \right\rangle^2$ satisfies $\mathbf{Pr}_{\boldsymbol{x} \sim \gamma_n}[\boldsymbol{q} \geq t] \leq e^{-\Omega(\sqrt{t})}$ for any $t \gtrsim m^2$. Using this tail bound for $m = \Theta(d)$ and conditioning on $\boldsymbol{x} \in X$ where X is an arbitrary subset of $\mathbb{R}^n$ with Gaussian measure $\approx 2^{-d}$, we obtained a bound $\mathbb{E}_{\boldsymbol{x} \sim \gamma}[\boldsymbol{q} \mid \boldsymbol{x} \in X] \lesssim d^2$. This shows that there can be at most $O(d)$ such quadratic forms M_i's

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

where the value $\mathbb{E}_{\boldsymbol{x}\sim\gamma}\left[\left\langle \boldsymbol{x}\,\dot{\otimes}\,\boldsymbol{x}, M_i\right\rangle^2 \middle| \boldsymbol{x}\in X\right]$ 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 $\boldsymbol{x}\in X$ is $\Theta(d^2)$.

Example 4.9.2. *For $1 \le i < j \le \sqrt{d}$, 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 $X = \left\{x \in \mathbb{R}^n \,\middle|\, |x_i| \gtrsim d^{1/4} \text{ for all } i \le d^{1/2}\right\}$. Then, the Gaussian measure $\gamma(X) = 2^{-\Theta(d)}$ but*

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

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-2 bounds.

Chapter 5

The Power of Adaptivity in Quantum Query Algorithms

Motivated by limitations on the depth of near-term quantum devices, we study the depth-computation trade-off in the query model, where the depth corresponds to the number of adaptive query rounds and the computation per layer corresponds to the number of parallel queries per round. We achieve the strongest known separation between quantum algorithms with r versus $r-1$ rounds of adaptivity. We do so by using the k-fold Forrelation problem introduced by Aaronson and Ambainis (SICOMP'18). For $k = 2r$, this problem can be solved using an r round quantum algorithm with only one query per round, yet we show that any $r-1$ round quantum algorithm needs an exponential (in the number of qubits) number of parallel queries per round.

Our results are proven following the Fourier analytic machinery developed in recent works on quantum-classical separations. The key new component in our result are bounds on the Fourier weights of quantum query algorithms with bounded number of rounds of adaptivity. These may be of independent interest as they distinguish the polynomials that arise from such algorithms from arbitrary bounded polynomials of the same degree.

Organization. In Section 5.1, we give a brief introduction on adaptivity in quantum query algorithms. In Section 5.2, we explain our main results. An overview of our Fourier growth analysis is provided in Section 5.3. Section 5.4 contains the full proof of our Fourier growth bounds. In Section 5.5, we discuss the tightness of our result in the non-adaptive setting. In Section 5.6, we generalize our result to quantum query algorithms with classical preprocessing power.

5.1 Introduction

A quantum query algorithm has "black-box access" to the input and is only charged for quantum queries to the input, while any intermediate computation is considered free. Most

well-known quantum algorithms, such as Grover's search [Gro96], Deutsch-Josza's algorithm [DJ92], Bernstein-Vazirani's algorithm [BV97], Simon's Algorithm [Sim97], and Shor's period-finding algorithm [Sho99], are captured by this black-box access model. There are slightly different models of black-box access to the input and in this work, we consider the most basic access model where each query returns a *bit* of the input.

Our focus is to identify the exact limits of quantum depth in the query model. One of the primary motivations for studying the power of depth comes from near-term quantum hardware which is restricted to quantum circuits of small depth in order to combat decoherence due to noise. Because of depth limitations, one needs to use wider circuits with more gates in each layer to perform computation, thus making parallel operations quite desirable. This makes optimizing the depth-width trade-off a fundamental task in quantum circuit synthesis for the near-term: reducing circuit depth allows the computation to be completed before the qubits decohere too much, but it also requires more quantum gates per layer.

On the positive side, Cleve and Watrous [CW00] showed how to implement the quantum Fourier transform in a parallel fashion, which leads to the parallelization of Shor's factoring algorithm [Sho99]. Regev [Reg25] employed parallelization followed by polynomial-time classical post-processing, to design a more efficient quantum algorithm for factoring under certain number-theoretic conjectures. On the other hand, Moore and Nilsson [MN01] conjectured that certain staircase-shaped quantum circuits cannot be efficiently parallelized.

In the query model abstraction, the circuit depth corresponds to the number of adaptive rounds, denoted by r, and the circuit width corresponds to the maximal number of parallel queries, denoted by t, per round. An extreme case $r = 1$ is the non-adaptive quantum query algorithm, where all queries are made in parallel. Perhaps surprisingly, van Dam [vD98] showed that any n-bit Boolean function can be computed with bounded error using only $t \leq n/2 + O(\sqrt{n})$ non-adaptive quantum queries, which is essentially tight for total functions [Mon10]. Techniques have been developed to establish lower bounds for various problems in this non-adaptive setting [NY04, KLPY10, Bur19], but less is known when we have more adaptive rounds. Zalka [Zal99] considered the unordered search problem on n-bit database and showed that $t = \Omega(n/r^2)$ is needed. This matches the simple divide-and-search algorithm: Partition the space into $O(n/r^2)$ parts of $O(r^2)$ size each and execute Grover's algorithm [Gro96] on each part in parallel in r steps. Jeffery, Magniez, and de Wolf [JMdW17] proved tight $t = \Theta(n/r^{3/2})$ trade-off for the element distinctness problem and tight $t = \Theta(n/r^{1+1/k})$ trade-off for the k-sum problem.

The above results show that being more adaptive indeed reduces the need of quantum queries. However the improvement is quite marginal: Even if we double the number of rounds, the saving is still only a constant factor. This naturally leads to the following question:

What is the largest possible saving in queries offered by more rounds of adaptivity?

5.2 Our Results

We answer the above question in the strongest sense and along the way prove structural theorems about the Fourier spectrum of polynomials that arise from low-depth quantum algorithms.

Our main result shows that the aforementioned k-fold Forrelation problem separates different levels of quantum computational power, measured in terms of adaptivity. Informally, the saving in the number of parallel queries can be unbounded, even when we just have one more adaptive round.

Theorem 5.2.1. *For any constant $r \geq 2$, the $2r$-fold Forrelation problem on n-bit inputs*

1. *can be solved with advantage 2^{-10r} by r adaptive rounds of queries with one quantum query per round, yet*
2. *any quantum query algorithm with $r-1$ adaptive rounds requires $\widetilde{\Omega}(n^{1/r^2})$ parallel queries to approximate it.*

Item 2 continues to hold even in the presence of a large amount of classical pre-processing. In more detail, we consider algorithms that are allowed to first make classical queries and based on the outputs, choose a quantum algorithm to run that has $k-1$ rounds of t parallel queries each. We show that any such algorithm must either make $\Omega(n^{1/(2r)})$ classical queries or $\widetilde{\Omega}(n^{1/r^2})$ quantum queries. See Section 5.6 for more details.

We note two easy modifications of the above theorem that also follow from our work, which we do not state in the theorem statement above for brevity. First, in the first item above, one can boost the advantage of the quantum algorithm to any constant close to 1 by making $2^{O(r)}$ parallel queries per round without increasing the number of rounds since error amplification can be done by making parallel queries. Second, we can more generally obtain an r versus r' separation for any $r' < r$ where the lower bound in the second item improves as r' decreases and is of the form $\widetilde{\Omega}(n^{c(r,r')})$ where

$$c(r,r') = \begin{cases} 1-\frac{1}{r} & \text{for } r'=1, \\ \frac{r-r'}{rr'+r/2} \geq \frac{1}{r^2} & \text{for } 2 \leq r' \leq r-1. \end{cases} \tag{5.1}$$

For example, reducing the number of rounds by a factor of 2, i.e., when $r = 2r'$, gives $c(r,r') = 1/(r+1)$. Furthermore, notice that the case when $r' = 1$ corresponds to a non-adaptive lower bound: here we obtain that any non-adaptive quantum algorithm that solves $2r$-fold Forrelation must make $\widetilde{\Omega}(n^{1-1/r})$ parallel queries.

Remark 5.2.2. *We recall that k-fold Forrelation is a partial function and being a partial function is necessary for Item 1. [Zal99, JMdW17] showed that for any total Boolean function f, the number of parallel quantum queries needed with r rounds is $t = \Omega(\mathsf{bs}(f)/r^2)$, where $\mathsf{bs}(f)$ is the block sensitivity complexity of f. Note that Simon [Sim83] proved that $\mathsf{bs}(f) = \Omega(\log n)$ if f is a non-degenerate n-bit Boolean function. This implies that $t = \Omega(\log n)$*

when r is a constant. Similarly, Ambainis and de Wolf [AdW14] showed that any non-degenerate n-bit total function requires $\Omega(\log n/\log\log n)$ quantum queries in total, which implies $t = \Omega(\log n/(r\log\log n))$. In summary, for total functions and constant rounds, the best possible separation is only logarithmic-vs-polynomial, instead of the $O(1)$-vs-polynomial separation we obtain.

As mentioned previously, Item 1 of Theorem 5.2.1 was already known since the work of [AA18] and the crux of our result is the lower bound in Item 2. Lower bounds for k-fold Forrelation are quite non-trivial to prove even for classical query algorithms and the known techniques rely on the polynomial method. The polynomial method cannot be directly applied since k-fold Forrelation is a low-degree bounded polynomial and as such one needs to find a way to distinguish it from the polynomials of much higher degree that are computed by the computational model of interest. In particular, we use Fourier growth as the distinguisher.

Fourier Growth of Low-Depth Quantum Algorithms

Following [RT22, Tal20], Bansal and Sinha [BS21] successfully related the advantage of approximating k-fold Forrelation for $k = 2r$ with the (low-level) Fourier growth of the model of computation in question. See Theorem 2.0.12 for detail. As a direct application, Item 2 of Theorem 5.2.1 follows from the following Fourier growth bounds.

Theorem 5.2.3. *Let $\mathcal{A}$ be a quantum query algorithm on n-bit inputs with arbitrarily many auxiliary qubits. Assume $\mathcal{A}$ has r adaptive rounds of $t \le n$ parallel queries. Define $f\colon \{\pm1\}^n \to [0,1]$ by $f(x) = \mathbf{Pr}[\mathcal{A}$ accepts $x]$. Then*

$$L_\ell(f) \le O_{r,\ell}\left(t^\ell \cdot \left(\sqrt{n/t}\right)^{\left\lfloor\left(1-\frac{1}{2r}\right)\ell\right\rfloor}\right).$$

Moreover, this bound holds when some bits of x are fixed in advance.

Remark 5.2.4. *In the non-adaptive case (i.e., $r = 1$), the bound in Theorem 5.2.3 can be improved (see Section 5.3 for detail) to*

$$L_\ell(f) \le O_\ell\left(t^{\ell/4} \cdot n^{\ell/4}\right).$$

This is also tight as shown in Section 5.5.

Proof of Theorem 5.2.1 and (5.1). Since Item 1 follows from Fact 2.0.11, we focus on Item 2. By Theorem 2.0.12, it suffices to show

$$\left(\frac{1}{\sqrt{n}}\right)^{1-\frac{1}{2r}} \cdot (L_\ell(f))^{1/\ell} \le r^{-20} \quad \text{for all } 2r \le \ell \le 2r\cdot(2r-1) \text{ and } t \le O_r\left(n^{c(r,r')}\right) \tag{5.2}$$

where the Fourier growth bound $L_\ell(f)$ is from Theorem 5.2.3 and satisfies

$$L_\ell(f) \le O_{r',\ell}\left(t^\ell \cdot \left(\sqrt{n/t}\right)^{\left\lfloor\left(1-\frac{1}{2r'}\right)\ell\right\rfloor}\right) \le O_r\left(t^{\frac{1}{2}+\frac{1}{4r'}} \cdot n^{\frac{1}{2}-\frac{1}{4r'}}\right)^\ell$$

for low levels of $\ell \le O(r^2)$. Putting $c(r,r') = \frac{r-r'}{rr'+r/2}$ gives the desired bound in (5.1).

For the special case $r' = 1$, we apply the improved bound from Remark 5.2.4:

$$L_\ell(f) \le O_\ell\left(t^{\ell/4} \cdot n^{\ell/4}\right) \le O_r\left(t^{1/4} \cdot n^{1/4}\right)^\ell$$

for low levels of $\ell \le O(r^2)$. Now (5.2) holds with $c(r,r') = 1 - 1/r$ as desired. □

The acceptance probability of any quantum query algorithm that makes d queries can be expressed as a degree-$2d$ bounded polynomial. Most of the techniques in the literature do not distinguish polynomials that come from quantum algorithms from general bounded polynomials and we lack a sufficiently good understanding of such distinctions.

Our Fourier growth bounds are far better than the bounds that can be obtained by directly applying the Fourier growth estimates for low-degree bounded polynomials [IRR+21, EI22]. Thus, this points to one way in which polynomials computed by low-depth quantum algorithms are different than general bounded polynomials of the same degree.

Classically Simulating Low-Depth Quantum Algorithms

We mention an open problem related to the question of where the exact limits of the trade-offs between depth and the number of parallel queries lie. As mentioned before, if there is only one query per round ($t = 1$), then Aaronson and Ambainis [AA18] conjectured that any r-round quantum algorithm can be simulated with $O(n^{1-1/2r})$ classical queries and this conjecture was proved by [BGGS22]. Does such a classical simulation continue to exist for low-depth quantum algorithms that make multiple parallel queries per round? We believe this is the case and make the following conjecture.

Conjecture 5.2.5. *Any quantum query algorithm on n-bit inputs with r adaptive rounds and t parallel queries per round can be classically simulated with $\widetilde{O}_{t,r}\left(n^{1-1/2r}\right)$ queries.*

It is worth mentioning that the Fourier growth bounds of classical query models (aka decision trees) [Tal20, SSW23] scales roughly like $(D \cdot \log n)^{\ell/2}$ where D is the number of classical queries. Our Fourier bound matches the Fourier bound for decision trees of depth $\widetilde{O}_{r,t}\left(n^{1-1/2r}\right)$ giving some support to the above conjecture.

Related Works in Communication Models. Aside from the aforementioned results in the quantum query complexity, the round-query trade-off in the query model can also be deduced from the round-communication trade-off in the model of communication complexity. In this model, Alice and Bob are given n-bit inputs x and y separately and their goal is to evaluate some function $F(x,y)$ by communication.

Given such a communication task F, we immediately get a query task f by letting $z = (x, y)$ and defining $f(z) = F(x, y)$. Then each quantum query to z can be implemented in the communication setting by Alice and Bob exchanging one round of $O(\log n)$ qubits.[1] Therefore if F requires sending $t \log n$ qubits in each round, then the corresponding f requires $\Omega(t)$ parallel queries in each round. Via this reduction, the pointer chasing problem with r jumps needs $\widetilde{\Omega}_r(n)$ parallel queries with $r-1$ adaptive rounds [KNTZ01, JRS02], whereas it can be solved with r adaptive rounds of $O(\log n)$ queries. Since the pointer chasing problem is a total function, by Remark 5.2.2 this logarithmic-vs-polynomial separation cannot be further improved to an $O(1)$-vs-polynomial separation.

We remark that[2] it is possible to define a variant of the pointer chasing problem which only uses one quantum query per round. This is achieved by using the Bernstein-Vazirani trick (see [Wat09]) to encode the address of each jump by the Hadamard code. Note that this is a partial function (due to the Bernstein-Vazirani trick), and it is conceivable that it will require $n^{\Omega(1)}$ queries if the number of adaptive rounds is reduced. In light of this, we highlight that our results generalize to the setting of *quantum query algorithms with classical preprocessing*, where the algorithm is allowed to first perform $n^{\Omega(1)}$ classical queries, then adaptively choose a quantum query algorithm with prescribed number of rounds and parallel queries. See details in Section 5.6. In this setting, variants of the pointer chasing problem would be solved already in the classical preprocessing phase, whereas the $2r$-fold Forrelation problem still exhibits an $O(1)$-vs-polynomial separation.

Related Works in Hybrid Models. There is another line of work on hybrid quantum-classical query algorithms that is related to the questions studied here. In particular, this line of work [CM20, CH23, HG22, ACC+23] considers the trade-off between quantum depth and the number of classical queries in a model that allows both. Although some of these works prove a fine-grained depth separation that seems similar to ours, the models considered in these works do not allow parallel queries (or only allow $\mathsf{polylog}(n)$-parallel queries in [CM20]) and they do not study the trade-offs between depth and parallel quantum queries. Consequently, these results are not comparable to ours.

5.3 Proof Overview

Describing Quantum Algorithms with Parallel Queries. Quantum algorithms which make parallel queries have the following form. First, we have an initial state $|u\rangle$; this state has some registers to index coordinates of the input and some registers for workspace. The algorithm has several rounds, where each round consists of a few parallel oracle queries followed by a unitary operator. The parallel queries are modelled by $O_x^{\otimes t} \otimes I$. Here, O_x is an

[1] Here $\log n$ is required for indexing an n-bit string in superposition, which is not needed classically. By switching the role of Alice and Bob between communication rounds, we can simulate r queries in r rounds of communication and one party in the end will compute the answer.

[2] We thank an anonymous QIP'24 reviewer for pointing this out.

$(n+1) \times (n+1)$ unitary that maps $|i\rangle$ to $x_i |i\rangle$ for all $i \in [n]$ and keeps $|0\rangle$ fixed, and this is equivalent to the usual quantum query oracle. The operator $O_x^{\otimes t}$ implements t parallel oracle queries and I acts as the identity matrix on the workspace. Finally, the algorithm applies some two-outcome measurement and returns the outcome as the output. See Figure 5.1 for depiction.

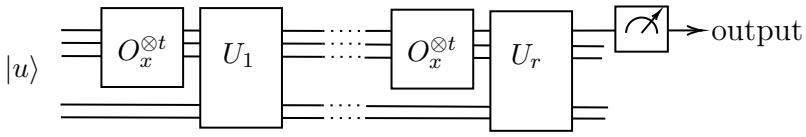


Figure 5.1: Quantum algorithm with r adaptive rounds of t parallel queries each.

For simplicity, let us imagine that there is no workspace memory. Additionally, let us ignore the action of the oracle O_x on the basis state $|0\rangle$ and treat O_x as an $n \times n$ unitary matrix. These simplifications are only for the proof overview, and our proof works in full generality. In this case, the acceptance probability of this algorithm can be expressed as

$$f(x) = u^\dagger O_x^{\otimes t} M_1 O_x^{\otimes t} \cdots M_{k-1} O_x^{\otimes t} v, \tag{5.3}$$

where k is twice the number of rounds, $u = v$ corresponds to the initial state, $M_1 = M_{k-1}^\dagger, M_2 = M_{k-2}^\dagger, \ldots, M_{k/2-1} = M_{k/2+1}^\dagger$ are the $\frac{k}{2} - 1$ unitary operators applied by the quantum algorithm and $M_{k/2}$ is the final measurement operator. For the rest of our proof, we can forget about the exact details of these matrices, we will only need that $M_1, \ldots, M_{k-1}$ have bounded operator norm and u, v are unit vectors.

Fourier Growth of Quantum Algorithms. Let us now understand the Fourier growth of functions as in (5.3) where $M_1, \ldots, M_{k-1}$ have bounded operator norm and u, v are unit vectors. We first set up some notation. We use $I \in [n]^t$ to denote a t-tuple of elements in $[n]$. We can view I as an ordered multiset of $[n]$ of size t (when counted with multiplicity). Accordingly, we use $\oplus I$ to denote the set of elements that appear an odd number of times in I and use $\oplus I \oplus I'$ to denote $(\oplus I) \oplus (\oplus I')$ for $I, I' \in [n]^t$.

When we expand the matrix multiplication in (5.3), many variables cancel out due to the identity $x_i^2 = 1$. Assume for simplicity that u and v are real vectors, i.e., $(u[I])^* = u[I]$. Thus, for all $S \subseteq [n]$, the coefficient of the monomial $\prod_{i \in S} x_i$ in (5.3) is given by

$$\widehat{f}(S) = \sum_{\substack{I_1, \ldots, I_k \in [n]^t \\ \oplus I_1 \oplus \ldots \oplus I_k = S}} u[I_1] M_1[I_1, I_2] M_2[I_2, I_3] \cdots M_{k-1}[I_{k-1}, I_k] v[I_k].$$

Fix complex numbers $\alpha_S = \widehat{f}(S)^* / |\widehat{f}(S)|$ for each $S \subseteq [n]$ of size ℓ. We wish to upper bound $L_\ell(f) = \sum_{|S|=\ell} \alpha_S \cdot \widehat{f}(S)$, which by the above is

$$L_\ell(f) = \sum_{\substack{I_1, \ldots, I_k \in [n]^t \\ |\oplus I_1 \oplus \cdots \oplus I_k| = \ell}} \alpha[\oplus I_1 \oplus \cdots \oplus I_k] \cdot u[I_1] M_1[I_1, I_2] M_2[I_2, I_3] \cdots M_{k-1}[I_{k-1}, I_k] v[I_k]. \tag{5.4}$$

To highlight the difficulties in upper bounding (5.4), we first present a few failed approaches and then describe our high-level proof approach. First, let us focus on the base case $k = 2$. For ease of notation, we will switch from indices I_1, I_2 to indices I, J and from the matrix M_1 to M. Our goal is to upper bound

$$L_\ell(f) = \sum_{\substack{I,J\in[n]^t \\ |\oplus I\oplus J|=\ell}} \alpha[\oplus I \oplus J] \cdot u[I]M[I,J]v[J].$$

One natural approach is to express $L_\ell(f)$ as a product of matrices (with bounded operator norms). One way to do this is to incorporate the phases $\alpha[\oplus I \oplus J]$ and the constraint $|\oplus I \oplus J| = \ell$ into the matrix $M[I, J]$. For instance, define $\widetilde{M}$ such that

$$\widetilde{M}[I,J] := \alpha[\oplus I \oplus J] \cdot \mathsf{1}[|\oplus I \oplus J| = \ell] \cdot M[I,J].$$

It is easy to see that $L_\ell(f) = u^\dagger \widetilde{M} v$ and consequently, $L_\ell(f) \le \|\widetilde{M}\|$. What is the best upper bound that we can prove for $\|\widetilde{M}\|$? At first glance, it might seem that we cannot do better than $\sqrt{n^t}$. Indeed, given an $n^t \times n^t$ unitary matrix M, if we multiply each entry by arbitrary numbers in the unit disk, this could blow up the operator norm by as much as $\sqrt{n^t}$ (the Hadamard matrix gives a tight example of this). However, we can do much better. This is because the terms multiplying each entry of M are highly constrained; the term multiplying the (I, J)-th entry depends only on $\oplus I \oplus J$.

To get an improved bound, consider the matrix D whose rows and columns are indexed by all possible $\oplus I$ and $\oplus J$ respectively, and the $(\oplus I, \oplus J)$-th entry is $\alpha[\oplus I \oplus J] \cdot \mathsf{1}[|\oplus I \oplus J| = \ell]$. It is not too difficult to convince oneself that $\widetilde{M}$ is a sub-matrix of $M \otimes D$. Therefore, $\|\widetilde{M}\| \le \|M\| \cdot \|D\| \le \|D\|$. Now, what is the best upper bound we can show for $\|D\|$? Consider the row corresponding to $\oplus I = \emptyset$. For this row, we need to choose a column $\oplus J$ such that $|\oplus J| = \ell$ and there are $\binom{n}{\ell}$ such columns. This already means that $\|D\| \ge \sqrt{\binom{n}{\ell}}$ (and this turns out to be tight). While a bound of $L_\ell(f) \le \sqrt{\binom{n}{\ell}}$ would already be a great improvement over the previous bound, it is still a trivial bound that holds for all bounded functions! Indeed, all Boolean functions which map into the complex unit disk satisfy $L_\ell(f) \le \sqrt{\binom{n}{\ell}}$.

To get the optimal bound of $n^{\ell/4} \cdot t^{\ell/4}$, the idea is to reduce the operator norm of D. For instance, suppose we defined $\widetilde{D}$ to be D, except that we zero out entries for which $|\oplus I \setminus \oplus J| \neq \ell/2$ (or equivalently $|\oplus J \setminus \oplus I| \neq \ell/2$). In this case, for any fixed $\oplus I$, the number of possibilities for $\oplus J$ is only $\binom{n}{\ell/2} \cdot \binom{t}{\ell/2}$ and we can actually prove that $\|\widetilde{D}\| \le n^{\ell/4} \cdot t^{\ell/4}$ as desired. Of course this doesn't suffice as we also need to sum over terms zeroed out.

In the full proof, the idea is to implicitly consider all possible values of $|\oplus I \setminus \oplus J|$. We fix any ℓ_1, ℓ_2 such that $|\oplus I \setminus \oplus J| = \ell_1$ and $|\oplus J \setminus \oplus I| = \ell_2$. Since $\ell_1 + \ell_2 = \ell$, either (1) $\ell_1 \le \ell/2$ or (2) $\ell_2 \le \ell/2$. We will define two different matrix product decompositions to handle each of these cases separately. It will turn out that the decomposition for case (1) satisfies an

operator norm bound of $n^{\ell_1/2} \cdot t^{\ell_2/2}$ and the decomposition for case (2) satisfies a bound of $n^{\ell_2/2} \cdot t^{\ell_1/2}$. Together, taking the geometric mean of the two bounds would give the desired bound of $n^{\ell/4} \cdot t^{\ell/4}$.

We remark that our proof does not explicitly list out these cases; instead, it defines two different decompositions and simply takes the minimum of the two bounds which essentially captures these two cases. We describe the details of this soon. For $k > 2$, it turns out that there is a subtle but crucial over-counting issue that is too technical to describe at this point. To address this, we need to introduce new matrices in the decompositions as well as carry out a step similar to Möbius inversion to undo the over-counting. We will highlight this issue later.

Technical Overview: $k = 2$

Recall from (5.4) that we wish to upper bound

$$L_\ell(f) = \sum_{\substack{I,J \in [n]^t \\ |\oplus I \oplus J| = \ell}} \alpha[\oplus I \oplus J] \cdot u[I] M[I,J] v[J]. \tag{5.5}$$

The high-level idea is as follows. We will express $L_\ell(f)$ as $\sum_{\substack{s_1,s_2 \in \mathbb{N} \\ s_1+s_2=\ell}} g(s)$ for some function $g(s)$, where $s = (s_1, s_2)$ and we shall group the terms based on the sizes s_1 and s_2 of the sets $\oplus I \setminus \oplus J$ and $\oplus J \setminus \oplus I$ respectively. We shall then upper bound $g(s)$ for any $s_1, s_2 \in \mathbb{N}$ satisfying $s_1 + s_2 = \ell$. To do this, we will express $g(s)$ in two different ways, namely, as $u^\dagger W R' v$ and as $u^\dagger W' R v$, for some matrices W, W', R, R' with bounded operator norms, and we will upper bound these by $\|u\|\|W\|\|R'\|\|v\|$ and $\|u\|\|W'\|\|R\|\|v\|$ respectively. Recall that $\|u\| = \|v\| = 1$. We will show that $\|R\|, \|R'\| \leq 1$ and

$$\|W\| \leq \sqrt{\binom{n}{s_2} \cdot \binom{t}{s_1}} \quad \text{and} \quad \|W'\| \leq \sqrt{\binom{n}{s_1} \cdot \binom{t}{s_2}}.$$

We upper bound the minimum of the two bounds by their geometric mean and use the fact that $s_1 + s_2 = \ell$ to obtain

$$g(s) \leq \sqrt{n^{s_2} t^{s_1} \cdot n^{s_1} t^{s_2}} = n^{\ell/4} t^{\ell/4}$$

as desired. We now describe the function $g(s)$ and the matrices W, W', R, R' in more detail.

We group the terms in (5.5) based on the sizes of $\oplus I \setminus \oplus J$ and $\oplus J \setminus \oplus I$. For any $(s_1, s_2) \in \mathbb{N} \times \mathbb{N}$, define the indicator function $\mathsf{Size}^s(S_1, S_2)$ for any subsets $S_1, S_2 \subseteq [n]$ by

$$\mathsf{Size}^s(S_1, S_2) = \mathsf{1}\left[|S_1 \setminus S_2| = s_1 \text{ and } |S_2 \setminus S_1| = s_2\right].$$

We will consider $\mathsf{Size}^s(\oplus I, \oplus J)$ as depicted in Figure 5.2.

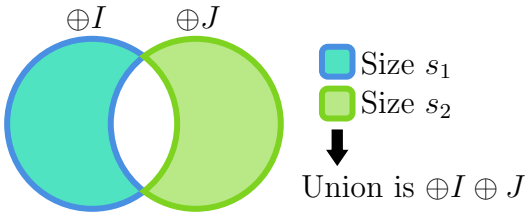


Figure 5.2: The constraint $\mathsf{Size}^s(\oplus I, \oplus J) = 1$.

Let $g(s)$ denote the contribution to (5.5) from terms satisfying $\mathsf{Size}^s(\oplus I, \oplus J) = 1$, that is,

$$g(s) := \sum_{I,J \in [n]^t} \mathsf{Size}^s(\oplus I, \oplus J) \cdot \alpha[\oplus I \oplus J] \cdot u[I] M[I,J] v[J].$$

From (5.5), we have $L_\ell(f) = \sum_{\substack{s_1,s_2 \in \mathbb{N} \\ s_1+s_2=\ell}} g(s)$. Fix any $s_1, s_2 \in \mathbb{N}$ such that $s_1 + s_2 = \ell$. We will now bound $g(s)$. As described before, we will express $g(s)$ in two different ways, namely, as $u^\dagger W R' v$ and as $u^\dagger W' R v$, for some matrices W, W', R, R' with bounded operator norms.

Expressing $g(s)$ as $u^\dagger W R' v$. The rows and columns of W are indexed by I and $(I', \oplus J)$ respectively, and those of R' by $(I', \oplus J)$ and J' respectively. These matrices are defined as follows

$$W[I, (I', \oplus J)] = \mathsf{1}[I = I'] \cdot \mathsf{Size}^s(\oplus I, \oplus J) \cdot \alpha[\oplus I \oplus J],$$
$$R'[(I', \oplus J), J'] = \mathsf{1}[\oplus J' = \oplus J] \cdot M[I', J'].$$

Intuitively, W is a matrix that multiplies by the signs $\alpha[\oplus I \oplus J]$ as well as enforces the Size^s constraint on $\oplus I$ and $\oplus J$, and R' is a matrix that implements the action of M, as well propagates information about $\oplus J$ backwards. This is depicted in Figure 5.3.

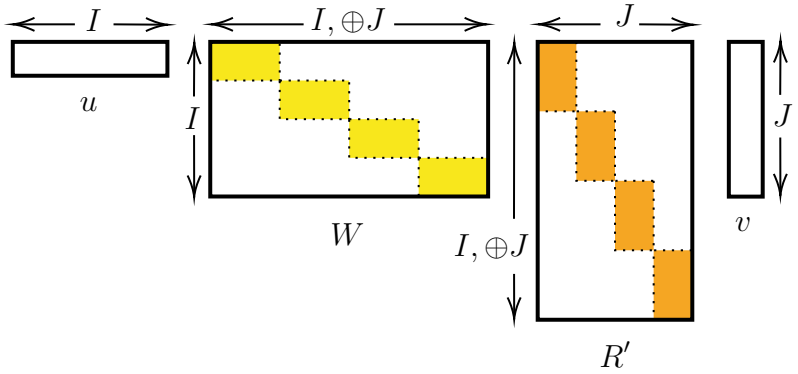


Figure 5.3: Expressing $g(s)$ as $u^\dagger W R' v$.

It is not too difficult to see that indeed $g(s) = u^\dagger W R' v$. We now show the desired upper bounds of $\|R'\| \leq 1$ and $\|W\| \leq \sqrt{\binom{n}{s_2} \cdot \binom{t}{s_1}}$.

- **Bounding $\|R'\|$:** We rearrange the columns of R' according to $\oplus J$. Under this ordering of the columns, observe that R' is a block diagonal matrix, where each block is a submatrix of M. Since $\|M\| \leq 1$, this implies that $\|R'\| \leq 1$.
- **Bounding $\|W\|$:** We rearrange the columns of W according to I and with this ordering, W is block-diagonal. We now use the fact that $\|W\| \leq \sqrt{\|W\|_1 \cdot \|W\|_\infty}$ where $\|W\|_1$ and $\|W\|_\infty$ are the max-column-norm and the max-row-norm respectively. Observe that $\|W\|_1 \leq 1$, since each column has at most one non-zero entry, which in turn is of unit magnitude. We now bound $\|W\|_\infty$. For any row $I \in [n]^t$, observe that there are at most $\binom{n}{s_2} \cdot \binom{t}{s_1}$ many columns $\oplus J$ such that $\mathsf{Size}^s(\oplus I, \oplus J) \neq 0$. Since each non-zero entry of W is of unit magnitude, this implies that $\|W\|_\infty \leq \binom{n}{s_2} \cdot \binom{t}{s_1}$. This gives us the desired bound of

$$\|W\| \leq \sqrt{\binom{n}{s_2} \cdot \binom{t}{s_1}}. \tag{5.6}$$

Expressing $g(s)$ as $u^\dagger R W' v$. The rows and columns of R are indexed by I and $(J', \oplus I')$ respectively and those of W' are indexed by $(J', \oplus I')$ and J respectively, and

$$W'[(J', \oplus I'), J] = \mathsf{1}[J = J'] \cdot \mathsf{Size}^s(\oplus I', \oplus J) \cdot \alpha[\oplus I' \oplus J],$$
$$R[I, (J', \oplus I')] = \mathsf{1}[\oplus I = \oplus I'] \cdot M[I, J'].$$

Here, W' implements $\alpha[\oplus I \oplus J]$ as well as enforces the Size^s constraint on $\oplus I$ and $\oplus J$, and R implements the action of M, as well propagates information about $\oplus I$ forward. This is depicted in Figure 5.4. A calculation similar to the previous case implies the desired bound of

$$\|W'\| \leq \sqrt{\binom{n}{s_1} \cdot \binom{t}{s_2}}. \tag{5.7}$$

This completes the proof overview for $k = 2$.

Technical Overview: $k = 3$

For simplicity of notation, we will switch from indices I_1, I_2, I_3 to indices I, J, K. We need to upper bound

$$L_\ell(f) = \sum_{\substack{I,J,K \in [n]^t \\ |\oplus I \oplus J \oplus K| = \ell}} \alpha[\oplus I \oplus J \oplus K] \cdot u[I] M_1[I, J] M_2[J, K] v[K].$$

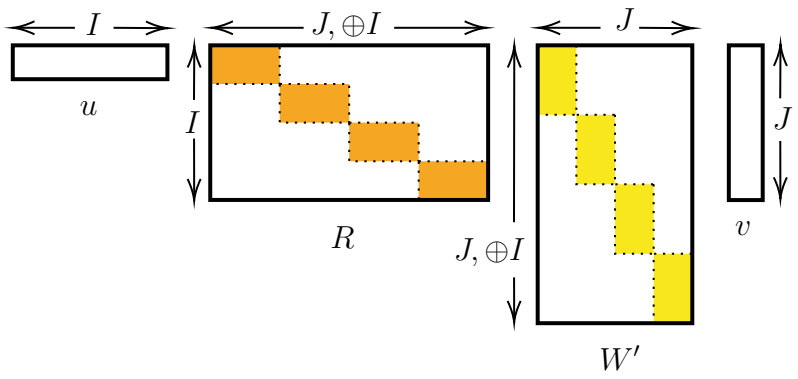


Figure 5.4: Expressing $g(s)$ as $u^\dagger RW'v$.

As before, we will express $L_\ell(f)$ as $\sum_{\substack{s_1,\ldots,s_4\in\mathbb{N}\\ s_1+\cdots+s_4=\ell}} g(s)$ grouping terms based on sizes of certain sets and in order to bound each $g(s)$, we will try to express it in three different ways as $u^\dagger W_1 R_1' R_2' v$, $u^\dagger R_1 W_2 R_2' v$ and $u^\dagger R_1 R_2 W_3 v$. It will turn out that $\|R_1\|, \|R_1'\|, \|R_2\|, \|R_2'\| \le 1$ and that

$$\|W_1\| \le (n/t)^{(s_2+s_3)/2} \cdot t^\ell, \qquad \|W_2\| \le (n/t)^{(s_1+s_3)/2} \cdot t^\ell, \qquad \|W_3\| \le (n/t)^{(s_1+s_2)/2} \cdot t^\ell. \quad (5.8)$$

Since $s_1 + s_2 + s_3 \le \ell$, taking the minimum of the three bounds would give us the desired bound of $(n/t)^{\ell/3} \cdot t^\ell$. There is an issue that comes up that we will later highlight. To describe it now in a nutshell, it turns out we *cannot* express $g(s)$ in the form of a matrix product with operator norms bounded as desired. Nevertheless, with some additional work, we can express a different function $h(s)$ in this form, furthermore, $h(s) = \sum_{s'} P[s,s']g(s')$ for some invertible matrix P such that P^{-1} has bounded norms. Therefore, using bounds on $h(s)$, we can derive the desired bounds on $g(s)$. We describe all this in more detail.

We start with the description of $g(s)$. Similar to the previous case, we will fix the sizes of certain sets in the Venn diagram of $\oplus I, \oplus J, \oplus K$ as depicted in Figure 5.5. More formally,

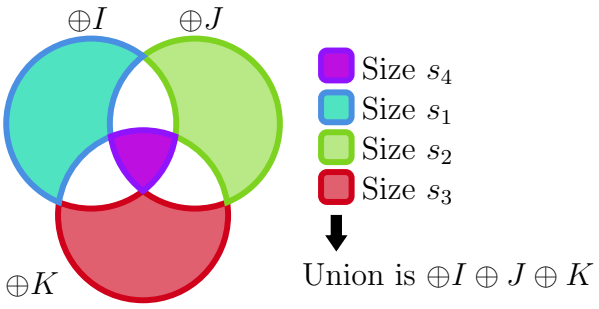


Figure 5.5: The constraint $\mathsf{Size}^s(\oplus I, \oplus J) = 1$.

let $s \in \mathbb{N}^4$. Define $\mathsf{Size}^s(S_1, S_2, S_3)$ to be the indicator function of

$$|S_1 \setminus (S_2 \cup S_3)| = s_1, \quad |S_2 \setminus (S_1 \cup S_3)| = s_2, \quad |S_3 \setminus (S_1 \cup S_2)| = s_3, \quad |S_1 \cap S_2 \cap S_3| = s_4.$$

Let

$$g(s) := \sum_{\substack{I,J,K \in [n]^t \\ |\oplus I \oplus J \oplus K| = \ell}} \mathsf{Size}^s(\oplus I, \oplus J, \oplus K) \cdot \alpha[\oplus I \oplus J \oplus K] \cdot u[I] M_1[I, J] M_2[J, K] v[J].$$

We attempt to express $g(s)$ in three different ways as

$$u^\dagger W_1 R_1' R_2' v, \quad u^\dagger R_1 W_2 R_2' v, \quad \text{and} \quad u^\dagger R_1 R_2 W_3 v.$$

The simplest to describe is the second expression. Here, we have matrices R_1, W_2, R_2' whose indices are as depicted in Figure 5.6.

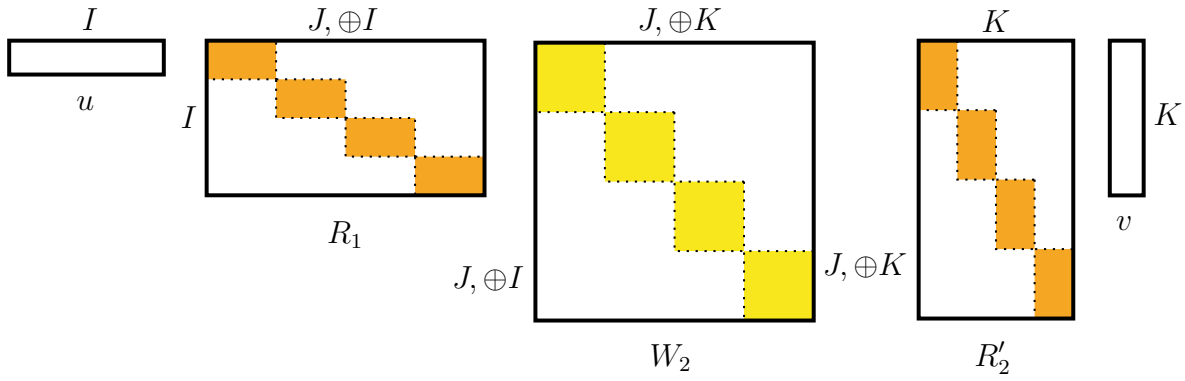


Figure 5.6: Expressing $g(s)$ as $u^\dagger R_1 W_2 R_2' v$.

Based on the intuition from before, there is a very natural way to define these matrices, namely,

$$R_1[I, (\oplus I', J)] = \mathbf{1}[\oplus I = \oplus I'] \cdot M_1[I, J] \text{ and } R_2'[(J, \oplus K), K'] = \mathbf{1}[\oplus K' = \oplus K] \cdot M_2[J, K'],$$
$$W_2[(J, \oplus I), (J', \oplus K)] = \mathbf{1}[J = J'] \cdot \mathsf{Size}^s(\oplus I, \oplus J, \oplus K) \cdot \alpha[\oplus I \oplus J \oplus K].$$

Note that given the row $(J, \oplus I)$ and the column $(J, \oplus K)$, we can compute $\mathsf{Size}^s(\oplus I, \oplus J, \oplus K)$ and $\alpha[\oplus I \oplus J \oplus K]$. A similar calculation to before shows that $\|R_1\|, \|R_2'\| \le 1$ and

$$\|W_2\| \le \sqrt{\binom{n}{s_3} \cdot \binom{t}{s_1}\binom{t}{s_2}\binom{t}{s_4}} \cdot \sqrt{\binom{n}{s_1} \cdot \binom{t}{s_2}\binom{t}{s_3}\binom{t}{s_4}} = (n/t)^{(s_1+s_3)/2} \cdot t^\ell.$$

Let us try to define the other two decompositions $u^\dagger W_1 R_1' R_2' v$ and $u^\dagger R_1 R_2 W_2 v$ as depicted in Figure 5.7. Suppose we could define W_1 and W_2 such that

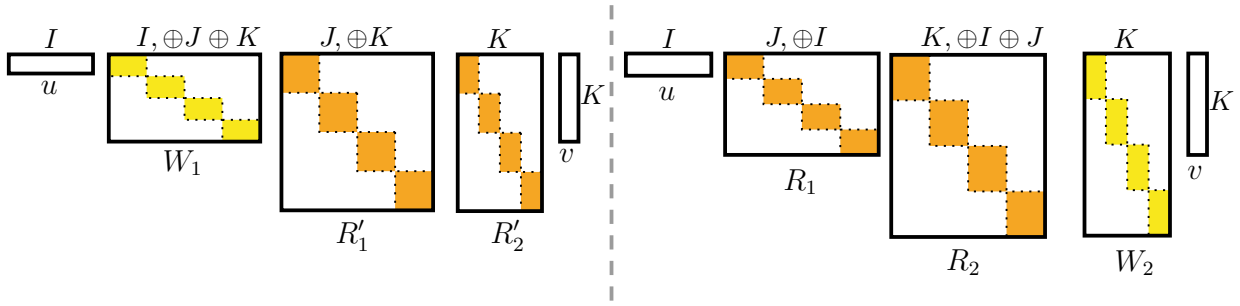


Figure 5.7: Expressing $g(s)$ as $u^\dagger W_1 R_1' R_2' v$ and $u^\dagger R_1 R_2 W_2 v$ respectively.

$$\begin{aligned} W_1[I,(I',\oplus J\oplus K)] &= \mathbf{1}\left[I=I'\right]\cdot \mathsf{Size}^s(\oplus I,\oplus J,\oplus K)\cdot \alpha[\oplus I\oplus J\oplus K],\\ W_3[K,(K',\oplus I\oplus K)] &= \mathbf{1}\left[K=K'\right]\cdot \mathsf{Size}^s(\oplus I,\oplus J,\oplus K)\cdot \alpha[\oplus I\oplus J\oplus K]. \end{aligned} \tag{5.9}$$

Then, a calculation similar to the previous case would give the desired operator norm bounds on W_1 and W_3 as in (5.8). The problem is that we cannot define matrices W_1, W_3 that satisfy (5.9). We explain this issue for W_1. Given a row I and a column $(I, \oplus J \oplus K)$, we cannot compute $\mathsf{Size}^s(\oplus I, \oplus J, \oplus K)$. After all, we only have the information about $\oplus I$ and $\oplus J \oplus K$, and hence the matrix W_1 can only enforce the constraints that $|\oplus I \setminus (\oplus J \oplus K)| = s_1 + s_4$ and $|(\oplus J \oplus K) \setminus \oplus I| = s_2 + s_3$, but it cannot enforce $|\oplus J \setminus (\oplus I \cup \oplus K)| = s_2$ or $|\oplus K \setminus (\oplus I \cup \oplus J)| = s_3$. In particular, if we only define W_1 to enforce the constraints that it is able to enforce, we will end up counting terms corresponding to I', J', K' which satisfy $\mathsf{Size}^{s'}(\oplus I, \oplus J, \oplus K)$ for s' with $s_2' \neq s_2$ and $s_3' \neq s_3$. In this case, instead of estimating the target $g(s)$, we would be over-counting. We need two new ideas here.

1. First we need to provide W_1 some additional information. One might hope that with a little extra information, W_1 can enforce Size^s, but this turns out to be false. Giving this information will increase the operator norms by too much. Instead, the idea is to provide some information that enforces a variant of the Size^s constraint.
2. This variant will allow us to bound a different function $h(s)$. This function is still an over-counting of $g(s)$, but the important point is that it is a predictable over-counting, that is, $h(s) = \sum_{s'} P[s, s']g(s')$ for some invertible matrix P such that P^{-1} has bounded norm. Therefore, we can derive bounds on $g(s)$ using bounds on $h(s)$.

We first explain step (2). Let $L(I, J, K) = \alpha[\oplus I \oplus J \oplus K] \cdot u[I]M_1[I, J]M_2[J, K]v[K]$. While we would like to bound the expression

$$g(s) := \sum_{I,J,K\in[n]^t} L(I,J,K)\cdot \mathsf{Size}^s(I,J,K),$$

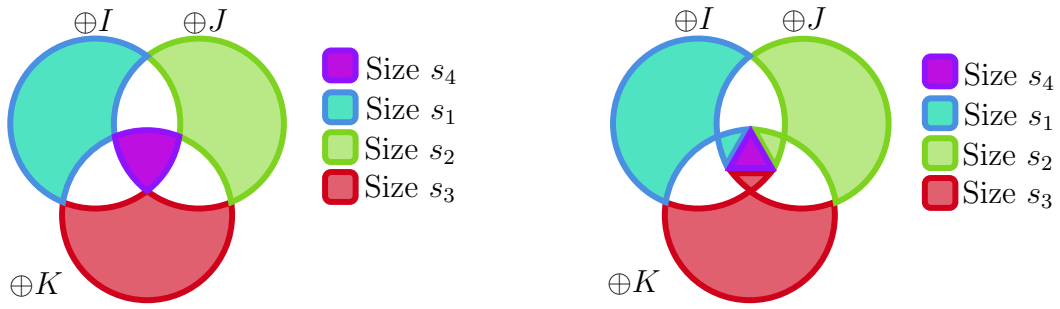
what we can bound turns out to be the expression

$$h(s) := \sum_{I,J,K\in[n]^t} L(I,J,K) \cdot \sum_{\substack{A,B,C,D\in[n]^t \\ A,B,C,D \text{ are disjoint} \\ A\cup B\cup C\cup D=\oplus I\oplus J\oplus K}} \mathsf{Subset}^s(A,B,C,D),$$

where $\mathsf{Subset}^s(A,B,C,D)$ is the indicator function of the constraint that

$$A \subseteq \oplus I, |A| = s_1, \quad B \subseteq \oplus J, |B| = s_2, \quad C \subseteq \oplus K, |C| = s_3, \quad D \subseteq \oplus I \cap \oplus J \cap \oplus K, |D| = s_4. \tag{5.10}$$

This is depicted in Figure 5.8.



(a) What we want to sum over

(b) What we end up summing over

Figure 5.8: The summation in $g(s)$ versus $h(s)$.

Observe that one of the terms in $h(s)$ is $A = \oplus I \setminus (\oplus J \cup \oplus K), B = \oplus J \setminus (\oplus I \cup \oplus K), C = \oplus K \setminus (\oplus I \cup \oplus J)$ and $D = \oplus I \cap \oplus J \cap \oplus K$. Hence, $h(s)$ consists of $g(s)$ plus some additional terms. For example, elements from D can be moved to either A, B, or C and still satisfy the constraints in (5.10). However, we can express

$$h(s) = \sum_{s'} P[s,s']g(s')$$

for a structured matrix P. This matrix is invertible and has bounded $\|P^{-1}\|_1$. Therefore, our goal of bounding $\|g\|_1$ reduces to bounding $\|h\|_1$ as $h = Pg$. This is done in step (1) which we now explain.

We now explain how to bound $h(s)$. We will blow up the matrices in the decomposition $u^\dagger W_1 R_1' R_2' v$ to include information about $A, B, C, D \subseteq [n]$. We will also introduce new matrices Q_1, Q_1', Q_2', Q_3' to enumerate A, B, C, D and verify that they satisfy the Subset^s constraints in (5.10). Consider the expression $u^\dagger Q_1 W_1 Q_1' R_1' Q_2' R_2' Q_3' v$, where the matrices are as depicted in Figure 5.9.

The matrices W_1, R_2', R_3' perform the same role as before and in addition, propagate information about the sets A, B, C, D. The matrices Q_1, Q_1', Q_2', Q_3' impose constraints on A, B, C, D as well as add and delete information as required. In more detail,

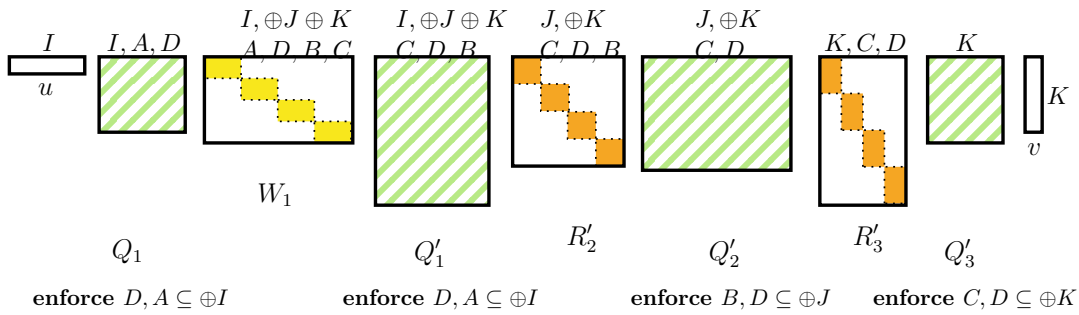


Figure 5.9: Expressing $h(s) = u^\dagger Q_1 W_1 Q_1' R_1' Q_2' R_2' Q_3' v$.

1. Q_1 propagates I and introduces A, D such that that $A, D \subseteq \oplus I$, $|A| = s_1$ and $|D| = s_4$. Given I, there are at most $\binom{t}{s_1} \cdot \binom{t}{s_4}$ possibilities for (A, D) and it follows that $\|Q_1\| \leq \sqrt{t^{s_1} \cdot t^{s_4}}$.
2. W_1 enforces $A \cup B \cup C \cup D = \oplus I \oplus J \oplus K$ and the size constraints on B, C. It also applies $\alpha[\oplus I \oplus J \oplus K]$. For each I, A, D, there are at most $\binom{n}{s_2} \cdot \binom{n}{s_3}$ possibilities for (B, C) and once we fix A, B, C, D and I, we also fix $\oplus J \oplus K = \oplus I \oplus (A \cup B \cup C \cup D)$. So $\|W_1\| \leq \sqrt{n^{s_2} \cdot n^{s_3}}$.
3. Q_3' back-propagates K and introduces C, D such that $C, D \subseteq \oplus K$, $|C| = s_3$, and $|D| = s_4$. Given K, there are at most $\binom{t}{s_3} \cdot \binom{t}{s_4}$ possibilities for (C, D) and hence $\|Q_3'\| \leq \sqrt{t^{s_3} \cdot t^{s_4}}$.
4. R_3' back-propagates $\oplus K, C, D$ and introduces J. It also applies the operator M_2. As before, $\|R_3'\| \leq 1$.
5. Q_2' back-propagates $J, \oplus K, C, D$, introduces B, and enforces that $B, D \subseteq \oplus J$ and $|B| = s_2$. Given J, there are at most $\binom{t}{s_2}$ possibilities for B, hence, $\|Q_2'\| \leq \sqrt{t^{s_2}}$.
6. R_2' back-propagates $D, B, C, \oplus J \oplus K$, introduces I, and applies the operator M_1. As before, $\|R_2'\| \leq 1$.
7. Q_1' back-propagates $D, B, C, \oplus J \oplus K, I$, introduces A, and enforces $D, A \subseteq \oplus I$ and $|A| = s_1$. Given I, there at most $\binom{t}{s_1}$ possibilities A, hence $\|Q_1'\| \leq \sqrt{t^{s_1}}$.

Combining all these bounds gives us an upper bound on $h(s)$ of

$$\sqrt{n^{s_2+s_3}} \cdot t^{s_4+s_1+(s_2+s_3)/2} = (n/t)^{(s_2+s_3)/2} \cdot t^\ell.$$

By a symmetric argument, we blow up the matrices in the decomposition $u^\dagger R_1 R_2 W_3 v$ to include information about $A, B, C, D \subseteq [n]$ and get

$$h(s) \leq (n/t)^{(s_1+s_2)/2} \cdot t^\ell.$$

Combining the three upper bounds on $h(s)$ we get $h(s) \leq (n/t)^{\ell/3} \cdot t^\ell$.

5.4 Fourier Growth of the Quantum Query Model

One way to think of O_x is to view the input x as a truth table of length $(n+1)$ where x_0 is fixed to 1. In this sense, the oracle query can be unified as $O_x|i\rangle = x_i|i\rangle$ for all $i \in \{0,\dots,n\}$. Meanwhile, for our purposes, it is desired to obtain Fourier growth bounds for downwards closed families. That is, the Fourier growth bounds should hold for the function even after fixing variables to values. This is usually not an issue regarding complexity measures, but the quantum query model is not evidently downwards closed. Therefore we prove the following more general theorem.

Theorem 5.4.1. *Let $n,t \geq 1$ and $m \geq 0$ and $d \geq 2$ be integers. Define $A = [n]^t \times [m]$. Let $u,v \in \mathbb{C}^A$ be two unit vectors. Let $M_1,\dots,M_{d-1} \in \mathbb{C}^{A\times A}$ be matrices satisfying $\|M_i\| \leq 1$ for each $i \in [d-1]$. Define $f\colon \{\pm1\}^{[n]} \to \mathbb{C}$ as*

$$f(x) = u^\dagger \left(O_x^{\otimes t} \otimes I_m\right) M_1 \left(O_x^{\otimes t} \otimes I_m\right) M_2 \cdots M_{d-1} \left(O_x^{\otimes t} \otimes I_m\right) v.$$

Let $\rho \in \{\pm1,*\}^{[n]}$ be an arbitrary restriction[3] and $\widetilde{n} = |\rho^{-1}(*)|$. Then for any $\ell \geq 0$, we have

$$L_\ell(f|_\rho) = \sum_{S\subseteq\rho^{-1}(*),|S|=\ell} \left|\widehat{f|_\rho}(S)\right| \leq 2^{\kappa(d,\ell)} \cdot t^\ell \cdot \max\left\{1,(\widetilde{n}/t)^{\frac{1}{2}\lfloor(d-1)\ell/d\rfloor}\right\},$$

where $\kappa(d,\ell) = O(d\ell)\cdot\min\left\{2^{d\ell},\ell^{2^d}\right\}$.

Before proving Theorem 5.4.1, we first summarize its application to the Fourier growth of quantum query algorithms, via the conversion stated in Chapter 2.

Corollary 5.4.2 (Formal Version of Theorem 5.2.3)**.** *Assume $\mathcal{A}$ is a query algorithm given oracle access O_x and uses arbitrarily many auxiliary qubits. Assume $\mathcal{A}$ makes r rounds of queries where each round consists of $t \leq n$ parallel queries. Let $f\colon \{\pm1\}^{[n]} \to [0,1]$ be its acceptance probability, i.e., $f(x) = \mathbf{Pr}[\mathcal{A}$ accepts $x]$. Then*

$$L_\ell(f) \leq 2^{\kappa(r,\ell)} \cdot t^\ell \cdot (n/t)^{\frac{1}{2}\left\lfloor\frac{(2r-1)\ell}{2r}\right\rfloor},$$

where $\kappa(r,\ell) = O(r\ell)\cdot\min\left\{2^{2r\ell},\ell^{4^r}\right\}$.

Moreover, this bound holds when some bits of x are fixed in advance.

Now we proceed to the proof of Theorem 5.4.1. Note that

$$f(x) = \sum_{\substack{\alpha_1,\dots,\alpha_d\in[m]\\ I_1,\dots,I_d\in[n]^t}} \left(u[(I_1,\alpha_1)]v[(I_d,\alpha_d)] \prod_{i\in[d-1]} M_i[(I_i,\alpha_i),(I_{i+1},\alpha_{i+1})]\right) \cdot \prod_{j\in[d],k\in[t]} x[I_j(k)].$$

[3] $f|_\rho$ is a sub-function on $\widetilde{n}$ variables of f by fixing x_i to $\rho(i)$ for $i \notin \rho^{-1}(*)$.

By rearranging coordinates, we assume without loss of generality $\rho^{-1}(*) = [\widetilde{n}]$, i.e., ρ fixes all but the first $\widetilde{n}$ bits of x. Now we expand $f|_\rho$ as

$$f|_\rho(x) = \sum_{\substack{\alpha_1,\ldots,\alpha_d\in[m]\\ I_1,\ldots,I_d\in[n]^t}} \left(u[(I_1,\alpha_1)]v[(I_d,\alpha_d)] \prod_{i\in[d-1]} M_i[(I_i,\alpha_i),(I_{i+1},\alpha_{i+1})] \right) \cdot \left(\prod_{\substack{j\in[d],k\in[t]\\ I_j(k)>\widetilde{n}}} \rho[I_j(k)] \right) \left(\prod_{\substack{j\in[d],k\in[t]\\ I_j(k)\le\widetilde{n}}} x[I_j(k)] \right).$$

Recall that $A = [n]^t \times [m]$ is the space of parallel queries and ancillary qubits. Now we define matrices $\widetilde{M}_1, \ldots, \widetilde{M}_{d-1} \in \mathbb{C}^{A\times A}$ as

$$\widetilde{M}_i[(I_i,\alpha_i),(I_{i+1},\alpha_{i+1})] = M_i[(I_i,\alpha_i),(I_{i+1},\alpha_{i+1})] \cdot \prod_{k\in[t],I_i(k)>\widetilde{n}} \rho[I_i(k)]$$

then define vectors $\widetilde{u},\widetilde{v}\in\mathbb{C}^A$ as $\widetilde{u} = u$ and

$$\widetilde{v}[(I_d,\alpha_d)] = v[(I_d,\alpha_d)] \cdot \prod_{k\in[t],I_d(k)>\widetilde{n}} \rho[I_d(k)].$$

Therefore we have

$$f|_\rho(x) = \sum_{\substack{\alpha_1,\ldots,\alpha_d\in[m]\\ I_1,\ldots,I_d\in[n]^t}} \left(\widetilde{u}[(I_1,\alpha_1)]\widetilde{v}[(I_d,\alpha_d)] \prod_{i\in[d-1]} \widetilde{M}_i[(I_i,\alpha_i),(I_{i+1},\alpha_{i+1})] \right) \cdot \prod_{\substack{j\in[d],k\in[t]\\ I_j(k)\le\widetilde{n}}} x[I_j(k)].$$

In addition, each $\widetilde{M}_i, \widetilde{u}, \widetilde{v}$ is the original M_i, u, v left multiplied by a ±1-diagonal matrix. By the norm guarantees of M_i, u, v, this means

$$\|\widetilde{u}\| = \|\widetilde{v}\| = 1 \quad \text{and} \quad \left\|\widetilde{M}_i\right\| \le 1. \tag{5.11}$$

Note that any index appearing twice in the multi-set $\{I_j(k)\}_{j\in[d],k\in[t]}$ cancels due to $(\pm1)^2 = 1$. We can compute each Fourier coefficient $\widehat{f|_\rho}(S)$ as

$$\widehat{f|_\rho}(S) = \sum_{\substack{(I_1,\alpha_1),\ldots,(I_d,\alpha_d)\in A\\ \oplus I_1\oplus\cdots\oplus I_d = S}} \widetilde{u}[(I_1,\alpha_1)]\widetilde{v}[(I_d,\alpha_d)] \prod_{i\in[d-1]} \widetilde{M}_i[(I_i,\alpha_i),(I_{i+1},\alpha_{i+1})],$$

where, from now on, we use $\oplus T_1 \oplus T_2 \oplus \cdots \subseteq [\widetilde{n}]$ to denote the set of indices in $[\widetilde{n}]$ that appear odd times in the multi-set consisting of indices from $T_1, T_2, \ldots$.

Now we introduce $a(S) = \mathsf{Phase}^{-1}\left(\widehat{f|_\rho}(S)\right)$ to denote the inverse of the phase of $\widehat{f|_\rho}(S)$, then

$$\begin{aligned} & L_\ell(f|_\rho) && (5.12)\\ &= \sum_{S\subseteq[\widetilde{n}],|S|=\ell} a(S)\cdot \widehat{f|_\rho}(S) \\ &= \sum_{\substack{(I_1,\alpha_1),\ldots,(I_d,\alpha_d)\in A\\ |\oplus I_1\oplus\cdots\oplus I_d|=\ell}} \underbrace{a(\oplus I_1\oplus\cdots\oplus I_d)\cdot \widetilde{u}[(I_1,\alpha_1)]\widetilde{v}[(I_d,\alpha_d)]\prod_{i\in[d-1]}\widetilde{M_i}[(I_i,\alpha_i),(I_{i+1},\alpha_{i+1})]}_{L(I_1,\alpha_1,\ldots,I_d,\alpha_d)}. && (5.13)\end{aligned}$$

For the analysis purpose, we will partition the binary strings $\{0,1\}^d$ and for this we introduce some notation. For each $b, b' \in \{0,1\}^d$, we say $b' \geq b$ if b' is no smaller than b entrywise; and we say $b' > b$ if $b' \geq b$ and $b' \neq b$. For $i \in [d]$, we write $i \in b$ if $b_i = 1$, and $i \notin b$ if $b_i = 0$.

Let $B = \left\{b \in \{0,1\}^d \,\middle|\, \|b\|_1 \equiv 1 \mod 2\right\}$ be the set of strings of odd Hamming weights, according to which we will partition $\oplus I_1 \oplus \cdots \oplus I_d$ into parts based on the membership in each $\oplus I_i$. Formally, for each $I = (I_1, \ldots, I_d) \in ([n]^t)^d$ and $b \in B$, define $I^{(b)} \subseteq [\widetilde{n}]$ to be the set of indices in $[\widetilde{n}]$ that appears in and only in those $\oplus I_i$ satisfying $b_i = 1$. Formally,

$$I^{(b)} = \left(\bigcap_{i\in b} \oplus I_i\right) \setminus \left(\bigcup_{i\notin b} \oplus I_i\right).$$

We emphasize that the intersection $\cap$ and union $\cup$ operators are applied to the inner sets $\oplus I_i$. Due to the construction, $|B| = 2^{d-1}$ and $\oplus I_1 \oplus \cdots \oplus I_d$ equals the (disjoint) union of all $I^{(b)}$'s.

Recall the definition of $L(I_1, \alpha_1, \ldots, I_d, \alpha_d)$ from (5.13). Now for each $s = (s^{(b)})_{b\in B} \in \mathbb{N}^B$ satisfying $\|s\|_1 = \ell$, we write the contribution of all the $(I_1, \alpha_1, \ldots, I_d, \alpha_d)$ consistent with s as

$$g(s) = \sum_{\substack{(I_1,\alpha_1),\ldots,(I_d,\alpha_d)\in A\\ \left|I^{(b)}\right|=s^{(b)},\forall b\in B}} L(I_1, \alpha_1, \ldots, I_d, \alpha_d). \tag{5.14}$$

Then we can express $L_\ell(f|_\rho)$ equivalently as

$$L_\ell(f|_\rho) = \sum_{s\in\mathbb{N}^B,\|s\|_1=\ell} g(s). \tag{5.15}$$

Here we are grouping $(I_1, \ldots, I_d)$ based on the sizes of the intersections and bounding the contribution from each group separately. We now count the number of possible sizes of intersection patterns. By a balls-into-bins counting, there are only

$$D := \binom{\ell + |B| - 1}{|B| - 1} = \binom{\ell + 2^{d-1} - 1}{2^{d-1} - 1} = O_{d,\ell}(1) \tag{5.16}$$

many possible s in the summation of (5.15).

Thus, our goal becomes bounding $\|g\|_1 = \sum_{\|s\|_1=\ell} |g(s)|$ and to do this we would like to bound each $g(s)$. However, as described in the proof overview, what we can bound turns out to be a function $h(s)$ where $h(s) = \sum_{\|s'\|_1=\ell} g(s') \cdot P[s, s']$ for some matrix P. We will now describe this function h and the matrix P. Lemma 5.4.3 will prove an upper bound on each $|h(s)|$ and we will use this lemma and properties about P to show the desired bound on $\|g\|_1$.

For each $s = (s^{(b)})_{b\in B} \in \mathbb{N}^B$ satisfying $\|s\|_1 = \ell$, define

$$h(s) = \sum_{\substack{(I_1,\alpha_1),\ldots,(I_d,\alpha_d)\in A \\ |\oplus I_1\oplus\cdots\oplus I_d|=\ell}} \sum_{\substack{J^{(b)}\subseteq[\widetilde{n}] \text{ of size } s^{(b)},\forall b\in B \\ J^{(b)}\text{'s are pairwise disjoint} \\ J^{(b)}\subseteq\bigcup_{b'\geq b} I^{(b')},\forall b\in B}} L(I_1,\alpha_1,\ldots,I_d,\alpha_d). \tag{5.17}$$

See Section 5.3 for a concrete example for the relation between h and g.

Each $h(\cdot)$ will be reformulated as a product of matrices that we can bound.

Lemma 5.4.3. $|h(s)| \leq t^\ell \cdot \max\left\{1, (\widetilde{n}/t)^{\frac{1}{2}\lfloor (d-1)\ell/d\rfloor}\right\}$.

The proof of Lemma 5.4.3 is deferred to the end of this section. Now we continue the task of bounding $L_\ell(f|_\rho)$ assuming Lemma 5.4.3. To relate $h(\cdot)$ with $g(\cdot)$, we count for any fixed $(I_1,\alpha_1),\ldots,(I_d,\alpha_d) \in A$ satisfying $|\oplus I_1 \oplus \cdots \oplus I_d| = \ell$, the number of possible $(J^{(b)})_{b\in B}$. By the condition $J^{(b)} \subseteq \bigcup_{b'\geq b} I^{(b')}, \forall b \in B$, we enumerate $J^{(b)}$ in the decreasing order of $\|b\|_1, b \in B$. Then the number of possibilities for each $J^{(b)}$ is exactly

$$\binom{\sum_{b'\geq b}\left|I^{(b')}\right| - \sum_{b'>b}\left|J^{(b')}\right|}{|J^{(b)}|} = \binom{\sum_{b'\geq b}\left|I^{(b')}\right| - \sum_{b'>b} s^{(b')}}{s^{(b)}},$$

where we fix $s = (s^{(b)})_{b\in B}$ and each $J^{(b)}$ has size $s^{(b)}$. Therefore the total number of choices is the telescoping product

$$\prod_{b\in B}\binom{\sum_{b'\geq b}\left|I^{(b')}\right| - \sum_{b'>b} s^{(b')}}{s^{(b)}},$$

which allows us to rewrite $h(s)$ as

$$\begin{aligned} h(s) &= \sum_{\substack{(I_1,\alpha_1),\ldots,(I_d,\alpha_d)\in A \\ |\oplus I_1\oplus\cdots\oplus I_d|=\ell}} L(I_1,\alpha_1,\ldots,I_d,\alpha_d)\cdot\prod_{b\in B}\binom{\sum_{b'\geq b}\left|I^{(b')}\right| - \sum_{b'>b} s^{(b')}}{s^{(b)}} \quad \text{(recall (5.17))} \\ &= \sum_{s'\in\mathbb{N}^B,\|s'\|_1=\ell}\prod_{b\in B}\binom{\sum_{b'\geq b} s'^{(b')} - \sum_{b'>b} s^{(b')}}{s^{(b)}} \cdot \sum_{\substack{(I_1,\alpha_1),\ldots,(I_d,\alpha_d)\in A \\ \left|I^{(b')}\right|=s'^{(b')},\forall b'\in B}} L(I_1,\alpha_1,\ldots,I_d,\alpha_d) \end{aligned}$$

$$= \sum_{s' \in \mathbb{N}^B, \|s'\|_1 = \ell} g(s') \cdot \prod_{b \in B} \binom{\sum_{b' \geq b} s'^{(b')} - \sum_{b' > b} s^{(b')}}{s^{(b)}} \qquad \text{(recall (5.14))}$$

$$=: \sum_{s' \in \mathbb{N}^B, \|s'\|_1 = \ell} g(s') \cdot P[s, s']. \qquad (5.18)$$

Therefore, viewing h and g as two vectors, they satisfy the relation $h = Pg$ where P is the coefficient matrix defined above and the dimension of P is $D \times D$ by (5.16). Recall the ℓ_1 norm of a matrix from Fact 2.0.3. The following lemma studies the properties of P itself.

Lemma 5.4.4. *P is an invertible matrix over $\mathbb{C}$ and $\|P^{-1}\|_1 \leq D \cdot \binom{\ell \cdot 2^{d-1}}{\ell}^D$.*

Then with the same linear algebraic notation, (5.15) completes the proof:

$$L_\ell(f|_\rho) = (1^D)^\top g \leq \|g\|_1 = \left\|P^{-1}h\right\|_1 \leq \left\|P^{-1}\right\|_1 \|h\|_1 \leq D \cdot \binom{\ell \cdot 2^{d-1}}{\ell}^D \cdot \|h\|_1 \qquad \text{(by Lemma 5.4.4)}$$

$$\leq D \cdot \binom{\ell \cdot 2^{d-1}}{\ell}^D \cdot D \cdot \|h\|_\infty$$

$$\leq D^2 \cdot \binom{\ell \cdot 2^{d-1}}{\ell}^D \cdot t^\ell \cdot \max\left\{1, (\widetilde{n}/t)^{\frac{1}{2}\lfloor (d-1)\ell/d \rfloor}\right\} \qquad \text{(by Lemma 5.4.3)}$$

$$\leq 2^{O(d\ell D)} \cdot t^\ell \cdot \max\left\{1, (\widetilde{n}/t)^{\frac{1}{2}\lfloor (d-1)\ell/d \rfloor}\right\}.$$

Finally by (5.16), we note that

$$D = \begin{cases} \binom{\ell + 2^{d-1} - 1}{\ell} \leq \binom{2^d}{\ell} \leq 2^{d\ell} & \ell \leq 2^{d-1} - 1, \\ \binom{\ell + 2^{d-1} - 1}{2^{d-1} - 1} \leq \binom{2\ell}{2^{d-1} - 1} \leq (2\ell)^{2^{d-1}-1} \leq \ell^{2^d} & \ell \geq 2^{d-1}. \end{cases}$$

This gives the desired bounds in Corollary 5.4.2.

Now we prove Lemma 5.4.4.

Proof of Lemma 5.4.4. We extend the partial order $>$ on elements in B to an arbitrary total order, denoted as $\gg$.[4] Let $T = \left\{s \in \mathbb{N}^B \,\middle|\, \|s\|_1 = \ell\right\}$ and let $\ggg$ be the lexicographical order on T induced by $\gg$ on B, i.e., $s \ggg s'$ iff there exists some $b \in B$ such that $s^{(b)} > s'^{(b)}$ and $s^{(b')} = s'^{(b')}$ holds for all $b' \gg b$.

Now we show that the matrix P, with rows and columns sorted according to $\ggg$, is lower-triangular with ones on the diagonal. This proves that P is invertible and $\det(P) = 1$.

[4] For example, one can think of $\gg$ as the decreasing order in the Hamming weight, and lexicographical order within the same Hamming weight.

Let $s \ggg s'$ be arbitrary elements from T and let $b^* \in B$ be such that $s^{(b^*)} > s'^{(b^*)}$ and $s^{(b')} = s'^{(b')}$ holds for all $b' \gg b^*$. Then we have

$$\begin{aligned}
P[s,s'] &= \prod_{b\in B} \binom{\sum_{b'\geq b} s'^{(b')} - \sum_{b'>b} s^{(b')}}{s^{(b)}} && \text{(recall (5.18))} \\
&= \binom{s'^{(b^*)} + \sum_{b'>b^*} s'^{(b')} - \sum_{b'>b^*} s^{(b')}}{s^{(b^*)}} \cdot \prod_{b\neq b^*} \binom{\sum_{b'\geq b} s'^{(b')} - \sum_{b'>b} s^{(b')}}{s^{(b)}} \\
&= \binom{s'^{(b^*)} + \sum_{b'\gg b^*} s'^{(b')} - \sum_{b'\gg b^*} s^{(b')}}{s^{(b^*)}} \cdot \prod_{b\neq b^*} \binom{\sum_{b'\geq b} s'^{(b')} - \sum_{b'>b} s^{(b')}}{s^{(b)}} \\
& && \text{(since } \gg \text{ is extended from } >) \\
&= 0 \cdot \prod_{b\neq b^*} \binom{\sum_{b'\geq b} s'^{(b')} - \sum_{b'>b} s^{(b')}}{s^{(b)}} = 0. \\
& && \text{(since } s^{(b^*)} > s'^{(b^*)} \text{ and } s^{(b')} = s'^{(b')} \text{ for all } b' \gg b^*)
\end{aligned}$$

The diagonal values can be calculated similarly:

$$P[s,s] = \prod_{b\in B} \binom{\sum_{b'\geq b} s^{(b')} - \sum_{b'>b} s^{(b')}}{s^{(b)}} = \prod_{b\in B} \binom{s^{(b)}}{s^{(b)}} = 1.$$

Now we bound $\|P^{-1}\|_1$. Let $P_{-s,-s'}$ be matrix P removing the s-th row and the s'-th column. Then the matrix inversion formula (See e.g., [Wik23]) gives

$$\left|P^{-1}[s,s']\right| = \left|\frac{\det\left(P_{-s',-s}\right)}{\det(P)}\right| = \left|\det\left(P_{-s',-s}\right)\right| \leq \operatorname{per}(P) \leq \|P\|_1^D,$$

where $\operatorname{per}(\cdot)$ denotes the permanent and the last inequality uses the fact that the dimension of P is D. Thus

$$\left\|P^{-1}\right\|_1 \leq D \cdot \max_{s,s'\in T} \left|P^{-1}[s,s']\right| \leq D \cdot \|P\|_1^D \tag{5.19}$$

and it suffices to bound $\|P\|_1 = \max_{s'\in T} \sum_{s\in T} |P[s,s']|$. Fix the maximizer s', we have

$$\begin{aligned}
\|P\|_1 &= \sum_{s\in T} \prod_{b\in B} \binom{\sum_{b'\geq b} s'^{(b')} - \sum_{b'>b} s^{(b')}}{s^{(b)}} \leq \sum_{s\in T} \prod_{b\in B} \binom{\sum_{b'\geq b} s'^{(b')}}{s^{(b)}} \\
&\leq \sum_{s\in T} \prod_{b\in B} \binom{\ell}{s^{(b)}} && \text{(since } \|s'\|_1 = \ell) \\
&= \text{the coefficient of } x^\ell \text{ in } \left((1+x)^\ell\right)^{|B|} && \text{(since } \|s\|_1 = \ell) \\
&= \binom{\ell \cdot 2^{d-1}}{\ell}, && \text{(since } |B| = 2^{d-1})
\end{aligned}$$

which completes the proof by plugging into (5.19). □

Finally we prove Lemma 5.4.3 which bounds $h(\cdot)$ entrywise. For convenience, we recall the definition of $h(s)$ and $L(I_1, \alpha_1, \ldots, I_d, \alpha_d)$ from (5.17) and (5.13):

$$h(s) = \sum_{\substack{(I_1,\alpha_1),\ldots,(I_d,\alpha_d)\in A \\ |\oplus I_1 \oplus \cdots \oplus I_d| = \ell}} \sum_{\substack{J^{(b)} \subseteq [\widetilde{n}] \text{ of size } s^{(b)}, \forall b \in B \\ J^{(b)}\text{'s are pairwise disjoint} \\ J^{(b)} \subseteq \bigcup_{b' \geq b} I^{(b')}, \forall b \in B}} L(I_1, \alpha_1, \ldots, I_d, \alpha_d)$$

and

$$L(I_1, \alpha_1, \ldots, I_d, \alpha_d) = a(\oplus I_1 \oplus \cdots \oplus I_d) \cdot \widetilde{u}[(I_1, \alpha_1)]\widetilde{v}[(I_d, \alpha_d)] \prod_{i \in [d-1]} \widetilde{M}_i[(I_i, \alpha_i), (I_{i+1}, \alpha_{i+1})].$$

Proof of Lemma 5.4.3. Let $r \in [d]$ be an index to be optimized later. We will write $h(s)$ as a product of matrices:

$$h(s) = \overline{u}^\dagger Q_1 R_1 Q_2 R_2 \cdots Q_{r-1} R_{r-1} Q_r W Q'_r R'_r Q'_{r+1} R'_{r+1} \cdots Q'_{d-1} R'_{d-1} Q'_d \overline{v}, \tag{5.20}$$

where

- Q_i enforces constraints on $J^{(b)}$s and propagates information about them forward,
- Q'_i enforces constraints on $J^{(b)}$s and propagates information about them backward,
- R_i implements the action of $\widetilde{M}_i$ and propogates information about $\oplus I_1 \oplus \cdots \oplus I_i$ forward,
- R'_i implements the action of $\widetilde{M}_i$ and propogates information about $\oplus I_{i+1} \oplus \cdots \oplus I_{d-1}$ backward,
- W is a sign matrix constructed to multiply by the phases $a(\cdot)$, as well as aggregate information about $J^{(b)}, \oplus I_1 \oplus \ldots \oplus I_i$ and $\oplus I_{i+1} \oplus \ldots \oplus I_{d-1}$.
- the vector $\overline{u}$ (resp., $\overline{v}$) is simply the vector u (resp., v) padded with zeros to fit with the dimension of Q_1 (resp., Q'_d).

In the following, we use the symbol $\perp$ to denote the value is unassigned. For any $b \in \{0,1\}^d$ and $i \in [d]$, we use $b_{\leq i}$ to denote string $(b_1, b_2, \ldots, b_i)$, and define similarly for $b_{<i}, b_{\geq i}, b_{>i}$.

We index the rows of matrix $\square \in \{Q_i, R_i, W, Q'_i, R'_i\}$ by $(I_\square, \alpha_\square) \in A$, $S_\square \subseteq [\widetilde{n}]$, and $J^{(b)}_\square \in 2^{[\widetilde{n}]} \cup \{\perp\}$ for all $b \in B$; and its columns are similarly indexed by $(I'_\square, \alpha'_\square)$, $S'_\square$, and ${J'_\square}^{(b)}$.

Likewise, we index the coordinates of vector $\diamondsuit \in \{\overline{u}, \overline{v}\}$ by $(I_\diamondsuit, \alpha_\diamondsuit)$, $S_\diamondsuit$, and $J^{(b)}_\diamondsuit$. In particular for the vectors, we assign

$$\diamondsuit[(I_\diamondsuit, \alpha_\diamondsuit, S_\diamondsuit, J^{(b)}_\diamondsuit)] = \begin{cases} \diamondsuit[(I_\diamondsuit, \alpha_\diamondsuit)] & S_\diamondsuit = \emptyset \text{ and } J^{(b)}_\diamondsuit = \perp, \forall b \in B, \\ 0 & \text{otherwise.} \end{cases}$$

Despite the dimension of the vectors $\overline{u}, \overline{v}$ being increased, they are simply padded by zeros. Therefore the norm is preserved from (5.11):

$$\|\overline{u}\| = \|\overline{v}\| = 1. \tag{5.21}$$

Now we turn to the matrices.

Construction of the Blow-Up Matrix Q_i. Each $\square = Q_i$ is a zero-one matrix where the entry is assigned one iff $I'_\square = I_\square$, $\alpha'_\square = \alpha_\square$, $S'_\square = S_\square$, and for each $b \in B$,

1. if $i \in b$, then ${J'_\square}^{(b)} \subseteq \oplus I_\square$,
2. a) If $b_{\le i} = 0^{i-1}1$, then $J_\square^{(b)} = \perp$ and $\left|{J'_\square}^{(b)}\right| = s^{(b)}$,
   b) If $b_{\le i} \neq 0^{i-1}1$, then $J_\square^{(b)} = {J'_\square}^{(b)}$,

The intuition behind this expression is the following. We need to ensure two conditions, namely, (1) for all i, we have $J^{(b)} \subseteq \oplus I_i$ if $b \ni i$, and, (2) $\left|J^{(b)}\right| = s^{(b)}$ for all b. Condition (1) will be checked by the matrix Q_i in Item 1. Condition (2) will be checked by the matrix Q_i in Item 2a, where i is the first non-zero coordinate in b. In contrast, Item 2b is to inherent Condition (2) from previous blow-up matrices. It will turn out that these conditions are enough to guarantee that the sets $J^{(b)}$ are pairwise disjoint (as shown in (5.32) and (5.33)).

We now upper bound the operator norm of Q_i. On the one hand, each column of $\square$ has at most one non-zero entry since the row index is a refinement of the column index. On the other hand, each row of $\square$ only has the possible freedom to select ${J'_\square}^{(b)}$ if $b_{\le i} = 0^{i-1}1$ in Item 2a, each of which amounts to at most $\binom{|\oplus I_\square|}{s^{(b)}} \le \binom{t}{s^{(b)}}$ options. Therefore by Fact 2.0.3, we have

$$\|Q_i\| \le \sqrt{\prod_{b: b_{\le i} = 0^{i-1}1} \binom{t}{s^{(b)}}}. \tag{5.22}$$

Construction of the Blow-Up Matrix Q'_i. Each $\square = Q'_i$ is a zero-one blow-up matrix similar to Q_i, with the role of the columns and rows exchanged: The entry is assigned one iff $I_\square = I'_\square$, $\alpha_\square = \alpha'_\square$, $S_\square = S'_\square$, and for each $b \in B$,

1. if $i \in b$, then ${J_\square}^{(b)} \subseteq \oplus I_\square$, and
2. a) if $b_{\ge i} = 10^{d-i}$, then ${J'_\square}^{(b)} = \perp$ and $\left|{J_\square}^{(b)}\right| = s^{(b)}$,
   b) if $b_{\ge i} \neq 10^{d-i}$, then ${J'_\square}^{(b)} = J_\square^{(b)}$.

By the same argument for (5.22), we have

$$\|Q'_i\| \le \sqrt{\prod_{b: b_{\ge i} = 10^{d-i}} \binom{t}{s^{(b)}}}. \tag{5.23}$$

Construction of the Operator Matrices R_i, R'_i. Each $\square = R_i$ is constructed to implement the action of $\widetilde{M}_i$ as well as to propagate forward information about $\oplus I_1 \oplus \cdots \oplus I_i$ and this information will be captured by $S'_\square$. Each entry of R_i is either $\widetilde{M}_i[(I_\square, \alpha_\square), (I'_\square, \alpha'_\square)]$ or zero, where the former case requires $S'_\square = \oplus S_\square \oplus I_\square$ and ${J'_\square}^{(b)} = J_\square^{(b)}$ for all $b \in B$.

To bound operator norm, we view row index $(I_\square, \alpha_\square, S_\square, \{J_\square^{(b)}\})$ as $(I_\square, \alpha_\square, T_\square, \{J_\square^{(b)}\})$ where $T_\square = \oplus S_\square \oplus I_\square$. Note that this is indeed a bijection since $S_\square = \oplus T_\square \oplus I_\square$. Moreover in the new indexing way, the entry is $\widetilde{M}_i[(I_\square, \alpha_\square), (I'_\square, \alpha_\square)]$ iff $S'_\square = T_\square$ and ${J'_\square}^{(b)} = J_\square^{(b)}$, which means $\square = R_i$ is a block diagonal matrix with block indexed by $(T_\square, \{J_\square^{(b)}\})$. Since each block is a sub-matrix of $\widetilde{M}_i$, by Fact 2.0.2 and Fact 2.0.1 the operator norm is preserved from (5.11):

$$\|R_i\| \le 1. \tag{5.24}$$

Each $\square = R'_i$ is similarly constructed to implement $\widetilde{M}_i$, as well as to propagate information about $\oplus I_{i+1} \oplus \cdots \oplus I_{d-1}$ using $S_\square$: Its entry is either $\widetilde{M}_i[(I_\square, \alpha_\square), (I'_\square, \alpha'_\square)]$ or zero, where the former case requires $S_\square = \oplus S'_\square \oplus I'_\square$ and $J_\square^{(b)} = {J'_\square}^{(b)}$ for all $b \in B$. By the same argument, we have

$$\|R'_i\| \le 1. \tag{5.25}$$

Construction of the Sign Matrix W. The final piece is to incorporate phases $a(\cdot)$ in the matrix $\square = W$. To this end, the entry is assigned $a(\oplus S_\square \oplus I_\square \oplus S'_\square)$ if (otherwise the entry is assigned zero)

1. $I_\square = I'_\square$, $\alpha_\square = \alpha'_\square$, and $|\oplus S_\square \oplus I_\square \oplus S'_\square| = \ell$,
2. for each $b \in B$,
   a) if $b_{\le r} = 0^r$, then $J_\square^{(b)} = \perp$, ${J'_\square}^{(b)} \ne \perp$, and $\left|{J'_\square}^{(b)}\right| = s^{(b)}$,
   b) else if $b_{\ge r} = 0^{d-r+1}$, then ${J'_\square}^{(b)} = \perp$, $J_\square^{(b)} \ne \perp$, and $\left|J_\square^{(b)}\right| = s^{(b)}$,
   c) else (i.e., $b_{\le r} \ne 0^r$ and $b_{\ge r} \ne 0^{d-r+1}$), then $J_\square^{(b)} = {J'_\square}^{(b)} \subseteq [\widetilde{n}]$ of size $s^{(b)}$,
3. $\oplus S_\square \oplus I_\square \oplus S'_\square = \bigcup_{b: b_{\le r} = 0^r} J'^{(b)} \cup \bigcup_{b: b_{\le r} \ne 0^r} J^{(b)}$.

The analysis of $\|W\|$ is similar to the one of $\|Q_i\|$. Each row of $\square$ is allowed to select ${J'_\square}^{(b)}$ if $b_{\le r} = 0^r$ in Item 2a, each of which has at most $\binom{\widetilde{n}}{s^{(b)}}$ options. Let $\overline{S} = \bigcup_{b: b_{\le r} = 0^r} J'^{(b)} \cup \bigcup_{b: b_{\le r} \ne 0^r} J^{(b)}$, which is fixed after enumerating ${J'_\square}^{(b)}$'s. By Item 3, we have $S'_\square = \oplus S_\square \oplus I_\square \oplus \overline{S}$ which is also fixed. Since each $a(\cdot)$ is a phase which has unit norm, we have

$$\|W\|_\infty \le \prod_{b: b_{\le r} = 0^r} \binom{\widetilde{n}}{s^{(b)}}.$$

Similarly, we can bound $\|W\|_1 \le \prod_{b: b_{\ge r} = 0^{d-r+1}} \binom{\widetilde{n}}{s^{(b)}}$. Therefore by Fact 2.0.3, we have

$$\|W\| \le \sqrt{\prod_{b: b_{\le r} = 0^r} \binom{\widetilde{n}}{s^{(b)}} \cdot \prod_{b: b_{\ge r} = 0^{d-r+1}} \binom{\widetilde{n}}{s^{(b)}}}. \tag{5.26}$$

Optimizing Bounds. To conclude the proof of Lemma 5.4.3, it suffices to verify (5.20) and optimize the choice of $r \in [d]$. We will deal with the former later, and focus on the bounds first.

Assuming (5.20), we have

$$\begin{aligned}
|h(s)| &\le \|\overline{u}\| \, \|Q_1\| \, \|R_1\| \cdots \|Q_{r-1}\| \, \|R_{r-1}\| \, \|Q_r\| \, \|W\| \, \|Q'_r\| \, \|R'_r\| \cdots \left\|Q'_{d-1}\right\| \left\|R'_{d-1}\right\| \, \|Q'_d\| \, \|\overline{v}\| \\
&\le \sqrt{\prod_{i=1}^{r} \prod_{b:b_{\le i}=0^{i-1}1} \binom{t}{s^{(b)}} \cdot \prod_{i=r}^{d} \prod_{b:b_{\ge i}=10^{d-i}} \binom{t}{s^{(b)}} \cdot \prod_{b:b_{\le r}=0^r} \binom{\widetilde{n}}{s^{(b)}} \cdot \prod_{b:b_{\ge r}=0^{d-r+1}} \binom{\widetilde{n}}{s^{(b)}}} \\
&\qquad\qquad \text{(by (5.21), (5.22), (5.23), (5.24), (5.25), and (5.26))} \\
&= \sqrt{\prod_{b:b_{\le r}\neq 0^r} \binom{t}{s^{(b)}} \cdot \prod_{b:b_{\ge r}\neq 0^{d-r+1}} \binom{t}{s^{(b)}} \cdot \prod_{b:b_{\le r}=0^r} \binom{\widetilde{n}}{s^{(b)}} \cdot \prod_{b:b_{\ge r}=0^{d-r+1}} \binom{\widetilde{n}}{s^{(b)}}} \\
&\le (\sqrt{t})^{\sum_{b:b_{\le r}\neq 0^r} s^{(b)} + \sum_{b:b_{\ge r}\neq 0^{d-r+1}} s^{(b)}} (\sqrt{\widetilde{n}})^{\sum_{b:b_{\le r}=0^r} s^{(b)} + \sum_{b:b_{\ge r}=0^{d-r+1}} s^{(b)}} \\
&= t^{\ell} \cdot (\sqrt{\widetilde{n}/t})^{\sum_{b:b_{\le r}=0^r} s^{(b)} + \sum_{b:b_{\ge r}=0^{d-r+1}} s^{(b)}} \qquad \text{(since } \|s\|_1 = \ell\text{)} \\
&=: t^{\ell} \cdot (\sqrt{\widetilde{n}/t})^{e_r}.
\end{aligned}$$

If $t \ge \widetilde{n}$, then $|h(s)| \le t^{\ell}$ since $e_r \ge 0$. Now consider the case $t \le \widetilde{n}$. For each $b \in B$, define

$$z(b) = \max\{r \in \mathbb{N} \,|\, j \notin b, \forall j \le r\} \quad \text{and} \quad z'(b) = \min\{r \in \mathbb{N} \,|\, j \notin b, \forall j \ge r\}.$$

Since $\|b\|_1 \ge 1$, we have $0 \le z(b) < d$, $0 < z'(b) \le d+1$, and $z(b) \le z'(b) - 2$. Therefore

$$\begin{aligned}
\sum_{r=1}^{d} e_r &= \sum_{r=1}^{d} \sum_{b:b_{\le r}=0^r} s^{(b)} + \sum_{r=1}^{d} \sum_{b:b_{\ge r}=0^{d-r+1}} s^{(b)} = \sum_{b\in B} s^{(b)} \cdot (z(b) + d + 1 - z'(b)) \\
&\le \sum_{b\in B} s^{(b)} \cdot (d-1) = (d-1)\ell, \qquad \text{(since } z(b) \le z'(b) - 2 \text{ and } \|s\|_1 = \ell\text{)}
\end{aligned}$$

which by averaging argument implies there exists a choice $r \in [d]$ such that $e_r \le \lfloor (d-1)\ell/d \rfloor$. This particular choice of r allows us to bound

$$|h(s)| \le t^{\ell} \cdot (\widetilde{n}/t)^{\frac{1}{2}\lfloor (d-1)\ell/d \rfloor}$$

as desired.

Verifying (5.20). Finally we verify the multiplication in (5.20) is consistent with the definition of $h(s)$ in (5.17).

For each $i \in [r]$, define vector $y^{(i)} = \left(\overline{u}^{\dagger} Q_1 R_1 Q_2 \cdots R_{i-1} Q_i\right)^{\dagger}$. For any fixed $I_1, \ldots, I_i$, define the following indicator functions.

$$\mathsf{Size}^{(i)}_{<}(\{J^{(b)}\}) := 1\left[\begin{array}{ll} J^{(b)} = \perp & \text{if } b_{\le i} = 0^i \\ \left|J^{(b)}\right| = s^{(b)} & \text{if } b_{\le i} \neq 0^i \end{array}\right]$$

$$\mathsf{Subset}^{(i)}_{<}(\{J^{(b)}\}) := 1\left[\oplus I_j \supseteq J^{(b)}, \forall j \le i, b \ni j\right]$$

We also define an indicator functions that captures the constraints of Q_i.

$$\mathsf{Q}^{(i)}(\{J^{(b)}, J'^{(b)}\}) := 1\left[\begin{array}{ll} J^{(b)} = \perp, J'^{(b)} \subseteq \oplus I \text{ has size } s^{(b)} & b_{\le i} = 0^{i-1}1 \\ J'^{(b)} = J^{(b)} \subseteq \oplus I & b_{\le i-1} \ne 0^{i-1}, b_i = 1 \\ J'^{(b)} = J^{(b)} & b_i = 0 \end{array}\right].$$

Claim 5.4.5. *The $(I, \alpha, S, \{J^{(b)}\})$-th entry of $y^{(i)}$ equals*

$$\sum_{\substack{(I_1,\alpha_1),\ldots,(I_i,\alpha_i)\in A \\ I_i = I, \alpha_i = \alpha \\ \oplus I_1 \oplus \cdots \oplus I_{i-1} = S}} \mathsf{Size}^{(i)}_{<}(\{J^{(b)}\}) \cdot \mathsf{Subset}^{(i)}_{<}(\{J^{(b)}\}) \cdot \widetilde{u}[(I_1, \alpha_1)] \prod_{j=1}^{i-1} \widetilde{M}_j[(I_j, \alpha_j), (I_{j+1}, \alpha_{j+1})] \quad (5.27)$$

We proof Claim 5.4.5 by induction. We now verify the base case $i = 1$. In this case, the expression in Claim 5.4.5 reduces to $1\left[\begin{array}{ll} J^{(b)} = \perp & \text{if } b_1 = 0 \\ \left|J^{(b)}\right| = s^{(b)} & \text{if } b_1 = 1 \end{array}\right] \cdot \sum_{\oplus I_1 \supseteq J^{(b)}, \forall b \ni 1} \widetilde{u}[(I, \alpha)]$. We have $y^{(1)} = \left(\overline{u}^\dagger Q_1\right)^\dagger$, and the $(I, \alpha, S, \{J^{(b)}\})$-th entry of $y^{(1)}$ equals

$$\begin{aligned} &= \sum_{(I', \alpha', S', \{J'^{(b)}\})} \overline{u}[(I', \alpha', S', \{J'^{(b)}\})] \cdot Q_1[(I', \alpha', S', \{J'^{(b)}\}), (I, \alpha, S, \{J^{(b)}\})] \\ &= \sum_{(I', \alpha', \emptyset, \{\perp\})} u[(I', \alpha')] \cdot Q_1[(I', \alpha', \emptyset, \{\perp\}), (I, \alpha, S, \{J^{(b)}\})] \qquad \text{(by the definition of } \overline{u}) \\ &= u[(I, \alpha)] \cdot 1[S = \perp] \cdot 1\left[\begin{array}{ll} J^{(b)} \subseteq \oplus I, \left|J^{(b)}\right| = s^{(b)} & \text{if } b \ni 1, \\ = \perp & \text{if } b \not\ni 1 \end{array}\right]. \qquad \text{(by the definition of } Q_1) \end{aligned}$$

This proves the base case of $i = 1$ for Claim 5.4.5. We now handle the inductive case $i \ge 2$. We have $y^{(i)} = \left(\left(y^{(i-1)}\right)^\dagger R_{i-1} Q_i\right)^\dagger$ and hence, the $(I, \alpha, S, \{J^{(b)}\})$-th entry of $y^{(i)}$ equals the sum over all possible $(I', \alpha', S', \{J'^{(b)}\})$ and $(I'', \alpha'', S'', \{J''^{(b)}\})$ of the product of the following three terms:

1. $y^{(i-1)}[(I', \alpha', S', \{J'^{(b)}\})]$,
2. $R_{i-1}[(I', \alpha', S', \{J'^{(b)}\}), (I'', \alpha'', S'', \{J''^{(b)}\})]$, and
3. $Q_i[(I'', \alpha'', S'', \{J''^{(b)}\}), (I, \alpha, S, \{J^{(b)}\})]$.

It is easy to see from the definition of R_{i-1} that (2) is non-zero only if $S'' = \oplus I' \oplus S'$ and $J''^{(b)} = J'^{(b)}$. Similarly, it follows from the definition of Q_i that (3) is non-zero only if $I'' = I$, $\alpha'' = \alpha$, and $S'' = S$. We now use the inductive hypothesis to express the $(I, \alpha, S, \{J^{(b)}\})$-th entry of $y^{(i)}$ as the sum over all possible $(I', \alpha', S', \{J'^{(b)}\})$ where $S = \oplus I' \oplus S'$ of the product of the following three terms:

1. $y^{(i-1)}[(I', \alpha', S', \{J'^{(b)}\})]$, which by induction can be expressed as

$$\sum_{\substack{(I_1,\alpha_1),\ldots,(I_{i-1},\alpha_{i-1})\in A \\ I_{i-1}=I',\alpha_{i-1}=\alpha' \\ \oplus I_1\oplus\cdots\oplus I_{i-2}=S'}} \mathsf{Size}_{<}^{(i-1)}(\{J'^{(b)}\}) \cdot \mathsf{Subset}_{<}^{(i-1)}(\{J'^{(b)}\}) \cdot \widetilde{u}[(I_1,\alpha_1)] \prod_{j=1}^{i-2} \widetilde{M}_j[(I_j,\alpha_j),(I_{j+1},\alpha_{j+1})].$$

2. $R_{i-1}[(I', \alpha', S', \{J'^{(b)}\}), (I, \alpha, S, \{J'^{(b)}\})]$, which is equal to $\widetilde{M}_{i-1}[(I', \alpha'), (I, \alpha)]$.
3. $Q_i[(I, \alpha, S, \{J'^{(b)}\}), (I, \alpha, S, \{J^{(b)}\})]$, which is equal to $\mathsf{Q}^{(i-1)}(\{J'^{(b)}, J^{(b)}\})$

We now combine the indicator functions in (1) and (3) by a case analysis. It is not too difficult to show that

$$\begin{aligned} &\mathsf{Size}_{<}^{(i-1)}(\{J'^{(b)}\}) \cdot \mathsf{Q}^{(i-1)}(\{J'^{(b)}, J^{(b)}\}) \\ &= \mathsf{Size}_{<}^{(i)}(\{J^{(b)}\}) \cdot 1 \begin{bmatrix} J'^{(b)} = J^{(b)} = \perp & \text{if } b_{\le i} = 0^i \\ J'^{(b)} = \perp, J^{(b)} \subseteq \oplus I & \text{if } b_{\le i} = 0^{i-1}1 \\ J'^{(b)} = J^{(b)} \subseteq \oplus I & \text{if } b_{\le i} \neq 0^{i-1}, b_i = 1 \\ J'^{(b)} = J^{(b)} & \text{if } b_{\le i} \neq 0^{i-1}, b_i = 0 \end{bmatrix}. \end{aligned}$$

Furthermore, setting $I_i = I$, we have

$$1 \begin{bmatrix} J'^{(b)} = J^{(b)} = \perp & \text{if } b_{\le i} = 0^i \\ J'^{(b)} = \perp, J^{(b)} \subseteq \oplus I & \text{if } b_{\le i} = 0^{i-1}1 \\ J'^{(b)} = J^{(b)} \subseteq \oplus I & \text{if } b_{\le i} \neq 0^{i-1}, b_i = 1 \\ J'^{(b)} = J^{(b)} & \text{if } b_{\le i} \neq 0^{i-1}, b_i = 0 \end{bmatrix} \cdot \mathsf{Subset}_{<}^{(i-1)}(\{J'^{(b)}\}) = \mathsf{Subset}_{<}^{(i)}(\{J^{(b)}\}).$$

Putting this together with the above facts completes the proof of Claim 5.4.5.

The $(I, \alpha, S, \{J^{(b)}\})$-th entry of $y'^{(i)} := Q'_i R'_i \cdots Q'_{d-1} R'_{d-1} Q'_d \overline{v}$ can be analyzed analogously as

$$\sum_{\substack{(I_i,\alpha_i),\ldots,(I_d,\alpha_d)\in A \\ I_i=I,\alpha_i=\alpha \\ \oplus I_{i+1}\oplus\cdots\oplus I_d=S}} \mathsf{Size}_{>}^{(i)}(\{J^{(b)}\}) \cdot \mathsf{Subset}_{>}^{(i)}(\{J^{(b)}\}) \cdot \widetilde{v}[(I_d,\alpha_d)] \prod_{j=i}^{d-1} \widetilde{M}_j[(I_j,\alpha_j),(I_{j+1},\alpha_{j+1})]. \tag{5.28}$$

where any fixed $I_1, \ldots, I_i$, we define the following indicator functions.

$$\mathsf{Size}_{>}^{(i)}(\{J^{(b)}\}) := 1 \begin{bmatrix} J^{(b)} = \perp & \text{if } b_{\ge i} = 0^{d-i+1} \\ \left|J^{(b)}\right| = s^{(b)} & \text{if } b_{\ge i} \neq 0^{d-i+1} \end{bmatrix}$$

$$\mathsf{Subset}_{>}^{(i)}(\{J^{(b)}\}) := 1\left[\oplus I_j \supseteq J^{(b)}, \forall j \ge i, b \ni j\right]$$

Hence the RHS of (5.20) equals $\left(y^{(r)}\right)^\dagger W y'^{(r)}$ and evaluates to

$$\sum_{\substack{(I,\alpha,S,\{J^{(b)}\})\\(I',\alpha',S',\{J'^{(b)}\})}} y^{(r)}[(I,\alpha,S,\{J^{(b)}\})]W[(I,\alpha,S,\{J^{(b)}\}),(I',\alpha',S',\{J'^{(b)}\})]y'^{(r)}[(I',\alpha',S',\{J'^{(b)}\})]$$

$$= \sum_{\substack{I,\alpha,S,S'\\ \{J^{(b)}\},\{J'^{(b)}\}\\ |\oplus S\oplus I\oplus S'|=\ell}} y^{(r)}[(I,\alpha,S,\{J^{(b)}\})]W[(I,\alpha,S,\{J^{(b)}\}),(I,\alpha,S',\{J'^{(b)}\})]y'^{(r)}[(I,\alpha,S',\{J'^{(b)}\})]$$

(by condition Item 1 of the definition of W)

$$= \sum_{\substack{I,\alpha,S,S'\\ \{J^{(b)}\},\{J'^{(b)}\}\\ |\oplus S\oplus I\oplus S'|=\ell}} W[(I,\alpha,S,\{J^{(b)}\}),(I,\alpha,S',\{J'^{(b)}\})]$$

$$\cdot \sum_{\substack{(I_1,\alpha_1),\ldots,(I_r,\alpha_r)\in A\\ I_r=I,\alpha_r=\alpha\\ \oplus I_1\oplus\cdots\oplus I_{r-1}=S}} \mathsf{Size}^{(r)}_{<}(\{J^{(b)}\})\cdot\mathsf{Subset}^{(r)}_{<}(\{J^{(b)}\})\cdot\widetilde{u}[(I_1,\alpha_1)]\prod_{i=1}^{r-1}\widetilde{M}_i[(I_i,\alpha_i),(I_{i+1},\alpha_{i+1})]$$

(by Claim 5.4.5)

$$\cdot \sum_{\substack{(I_r,\alpha_r),\ldots,(I_d,\alpha_d)\in A\\ I_r=I,\alpha_r=\alpha\\ \oplus I_{r+1}\oplus\cdots\oplus I_d=S'}} \mathsf{Size}^{(r)}_{>}(\{J'^{(b)}\})\cdot\mathsf{Subset}^{(r)}_{>}(\{J'^{(b)}\})\cdot\widetilde{v}[(I_d,\alpha_d)]\prod_{i=r}^{d-1}\widetilde{M}_i[(I_i,\alpha_i),(I_{i+1},\alpha_{i+1})]$$

(by (5.28))

$$= \sum_{\substack{(I_1,\alpha_1),\ldots,(I_d,\alpha_d)\in A\\ |\oplus I_1\oplus\cdots\oplus I_d|=\ell}} \widetilde{u}[(I_1,\alpha_1)]\widetilde{v}[(I_d,\alpha_d)]\prod_{i=1}^{d-1}\widetilde{M}_i[(I_i,\alpha_i),(I_{i+1},\alpha_{i+1})] \tag{5.29}$$

$$\cdot \sum_{\{J^{(b)}\},\{J'^{(b)}\}} W[(I_r,\alpha_r,\oplus I_1\oplus\cdots\oplus I_{r-1},\{J^{(b)}\}),(I_r,\alpha_r,\oplus I_{r+1}\oplus\cdots\oplus I_d,\{J'^{(b)}\})] \tag{5.30}$$

$$\cdot\,\mathsf{Size}^{(r)}_{<}(\{J^{(b)}\})\cdot\mathsf{Subset}^{(r)}_{<}(\{J^{(b)}\})\cdot\mathsf{Size}^{(r)}_{>}(\{J'^{(b)}\})\cdot\mathsf{Subset}^{(r)}_{>}(\{J'^{(b)}\}). \tag{5.31}$$

Notice that $b_{\le r}$ and $b_{\ge r}$ cannot both be zeros for $b\in B$. Thus conditions Items 2a to 2c of W show that we can enumerate $J''^{(b)}\subseteq[\widetilde{n}]$ of size $s^{(b)}$ and then let $J^{(b)}, J'^{(b)}$ be $\perp$ or $J''^{(b)}$ based on b. After this, (5.31) simply becomes the indicator of $J''^{(b)}\subseteq\bigcap_{i\in b}\oplus I_i$, and condition Item 3 of W becomes $\oplus I_1\oplus\cdots\oplus I_d=\bigcup_{b\in B}J''^{(b)}$. That is, (5.30) and (5.31) when combined, is equal to

$$a(\oplus I_1\oplus\cdots\oplus I_d)\cdot\sum_{\substack{J''^{(b)}\subseteq[\widetilde{n}]\text{ of size }s^{(b)},\forall b\in B\\ J''^{(b)}\subseteq\bigcap_{i\in b}\oplus I_i}}\left[\oplus I_1\oplus\cdots\oplus I_d=\bigcup_{b\in B}J''^{(b)}\right].$$

Now recall the definition of $L(I_1, \alpha_1, \ldots, L_d, \alpha_d)$ from (5.13) and combine (5.29). The RHS of (5.20) equals

$$\sum_{\substack{(I_1,\alpha_1),\ldots,(I_d,\alpha_d)\in A \\ |\oplus I_1 \oplus \cdots \oplus I_d| = \ell \\ J''^{(b)} \subseteq [\widetilde{n}] \text{ of size } s^{(b)}}} \left[J''^{(b)} \subseteq \bigcap_{i \in b} \oplus I_i \quad \wedge \quad \oplus I_1 \oplus \cdots \oplus I_d = \bigcup_{b \in B} J''^{(b)} \right] \cdot L(I_1, \alpha_1, \ldots, I_d, \alpha_d). \tag{5.32}$$

Finally it suffices to show this is equivalent to the summation in (5.17) which we restate here:

$$\sum_{\substack{(I_1,\alpha_1),\ldots,(I_d,\alpha_d)\in A \\ |\oplus I_1 \oplus \cdots \oplus I_d| = \ell \\ J^{(b)} \subseteq [\widetilde{n}] \text{ of size } s^{(b)}}} \left[J^{(b)} \subseteq \bigcup_{b' \geq b} I^{(b')} \quad \wedge \quad J^{(b)}\text{'s are pairwise disjoint} \right] \cdot L(I_1, \alpha_1, \ldots, I_d, \alpha_d), \tag{5.33}$$

where we recall that $I^{(b')} = \left(\bigcap_{i \in b'} \oplus I_i\right) \backslash \left(\bigcup_{i \notin b'} \oplus I_i\right)$. To this end, we fix $(I_1, \alpha_1), \ldots, (I_d, \alpha_d) \in A$ satisfying $|\oplus I_1 \oplus \cdots \oplus I_d| = \ell$ and show each possible $\{J''^{(b)}\}$ from (5.32) is also counted as $\{J^{(b)}\}$ in (5.33), and vice versa.

From (5.32) to (5.33). By the definition of $I^{(b')}$, we know $\oplus I_i \cap I^{(b')} = \emptyset$ whenever $i \notin b'$. Since $J''^{(b)} \subseteq \bigcap_{i \in b} \oplus I_i$, we have $J''^{(b)} \cap I^{(b')} \neq \emptyset$ implies $b' \geq b$. Note that $\oplus I_1 \oplus \cdots \oplus I_d = \bigcup_b I^{(b)}$. Therefore $\bigcup_b J''^{(b)} = \bigcup_b J^{(b)}$, and thus $J''^{(b)} \subseteq \bigcup_{b' \geq b} I^{(b')}$ as desired in (5.33). On the other hand, $\sum_b |J''^{(b)}| = \|s\|_1 = \ell = |\oplus I_1 \oplus \cdots \oplus I_d|$. Thus $\oplus I_1 \oplus \cdots \oplus I_d = \bigcup_b J''^{(b)}$ implies that $J''^{(b)}$'s are pairwise disjoint as desired in (5.33).

From (5.33) to (5.32). Since $\oplus I_1 \oplus \cdots \oplus I_d = \bigcup_b I^{(b)}$, we have $\bigcup_b J^{(b)} \subseteq \bigcup_b \bigcup_{b' \geq b} I^{(b')} = \bigcup_b I^{(b')} = \oplus I_1 \oplus \cdots \oplus I_d$. On the other hand, $\sum_b |J^{(b)}| = \|s\|_1 = \ell = |\oplus I_1 \oplus \cdots \oplus I_d|$. Thus $J^{(b)}$'s being pairwise disjoint implies that $\oplus I_1 \oplus \cdots \oplus I_d = \bigcup_b J^{(b)}$ as desired in (5.32). By the definition of $I^{(b')}$, we know $I^{(b')} \subseteq \bigcap_{i \in b'} \oplus I_i$. Therefore $J^{(b)} \subseteq \bigcup_{b' \geq b} I^{(b')} \subseteq \bigcup_{b' \geq b} \bigcap_{i \in b'} \oplus I_i = \bigcap_{i \in b} \oplus I_i$ as desired in (5.32). □

5.5 Tightness of the Non-Adaptive Case

Recall the definition of k-fold Forrelation problem from Definition 2.0.10. Here we show the tightness of our Fourier growth bounds for the non-adaptive parallel query algorithms using two-fold Forrelation function:

$$\mathrm{forr}_2(x_1, x_2) = \frac{1}{n} \sum_{i,j \in [n]} x_1(i) \cdot H_{i,j} \cdot x_2(j).$$

Since H is the orthonormal Hadamard matrix, each $H_{i,j}$ is $\pm 1/\sqrt{n}$. As a result, forr_2 is a degree-2 homogeneous function with

$$L_2(\mathrm{forr}_2) = \sqrt{n}. \tag{5.34}$$

Let Q be the quantum query algorithm from Fact 2.0.11 for $k = 2$. Then its acceptance probability function $f(x) = \mathbf{Pr}[Q \text{ accepts } x]$ equals $(1 + \mathrm{forr}_2(x))/2$.

Now, for a fixed positive odd number s, let $\mathrm{Maj}_s \colon \{\pm 1\}^s \to \{\pm 1\}$ be the majority function on s bits. Define $\mathrm{forr}_2 \circ \mathrm{Maj}_s \colon \{\pm 1\}^{2sn} \to \mathbb{R}$ by replacing input bits of forr_2 with majorities on disjoint sets of s bits:

$$\mathrm{forr}_2 \circ \mathrm{Maj}_s(y) = \frac{1}{n} \sum_{i,j \in [n]} \mathrm{Maj}_s(y_{1,i}) \cdot H_{i,j} \cdot \mathrm{Maj}_s(y_{2,j}), \tag{5.35}$$

where $y = (y_{1,1}, \ldots, y_{1,n}, y_{2,1}, \ldots, y_{2,n})$ and each $y_{1,i}, y_{2,j} \in \{\pm 1\}^s$.

We substitute the quantum query of Q on x by s parallel queries on y. This produces a non-adaptive quantum query algorithm $\overline{Q}$ with s parallel queries, and its acceptance probability function is

$$\overline{f}(y) = \frac{1 + \mathrm{forr}_2 \circ \mathrm{Maj}_s(y)}{2}.$$

Now for a fixed positive integer L, consider executing $\overline{Q}$ in parallel on L disjoint inputs and taking the parity of the results. This is a non-adaptive quantum query algorithm Q' with $t = sL$ parallel queries, and its acceptance probability function is

$$f'(z) = \frac{1}{2} + \frac{1}{2} \prod_{k \in [L]} \mathrm{forr}_2 \circ \mathrm{Maj}_s(y^k), \tag{5.36}$$

where $z = (y^1, \ldots, y^L)$ and each $y^k = (y^k_{1,1}, \ldots, y^k_{1,n}, y^k_{2,1}, \ldots, y^k_{2,n}) \in \{\pm 1\}^{2sn}$.

We lower bound the level-ℓ Fourier weight of f' with $\ell = 2L$. To this end, we observe that $\widehat{\mathrm{Maj}_s}(\emptyset) = 0$ and recall that forr_2 is degree-2 homogeneous. Since f' is essentially the product of L disjoint copies of $\mathrm{forr}_2 \circ \mathrm{Maj}_s$ (see (5.36)) and each $\mathrm{forr}_2 \circ \mathrm{Maj}_s$ is a sum of products of two disjoint Maj_s (see (5.35)), the level-ℓ Fourier coefficients of f' comes from expanding the products of level-1 Fourier weight of Maj_s, weighed by the level-2 Fourier coefficients of forr_2. Therefore

$$\begin{aligned} L_\ell(f') &= \frac{1}{2} \cdot \left(L_2(\mathrm{forr}_2) \cdot L_1(\mathrm{Maj}_s)^2\right)^L \\ &= \Omega\left(\sqrt{n} \cdot s\right)^L && \text{(by (5.34) and } L_1(\mathrm{Maj}_s) = \Theta(\sqrt{s})\text{)} \end{aligned}$$

Recall that $\ell = 2L$, $t = sL$, and f' is a function on $n = 2sLn$ input bits. This implies

$$L_\ell(f') \geq \Omega_\ell\left(n^{\ell/4} \cdot t^{\ell/4}\right),$$

matching the bound in Remark 5.2.4.

5.6 Quantum Query with Classical Preprocessing

In this section, we show that our Fourier analytic approach can be generalized to handle a more general setting, where the quantum query algorithm is allowed to first perform many

classical queries as a preprocessing phase. More precisely, we prove Lemma 5.6.1 analogous to Theorem 5.2.3.

Lemma 5.6.1. *Let $\mathcal{A}$ be an algorithm on n-bit inputs:*

- Classical Preprocessing Phase. *First $\mathcal{A}$ performs at most d classical queries.*
- Quantum Parallel Query Phase. *Then based on the results of the previous phase, $\mathcal{A}$ executes a quantum query algorithm $\mathcal{B}$ with arbitrarily many auxiliary qubits and r adaptive rounds of $t \le n$ parallel quantum queries per round.*

Define $f\colon \{\pm 1\}^n \to [0,1]$ by $f(x) = \mathbf{Pr}[\mathcal{A}$ accepts $x]$. Then

$$L_\ell(f) \le O_{r,\ell}\left((d\cdot t)^\ell \cdot \left(\sqrt{n/t}\right)^{\lfloor (1-\frac{1}{2r})\ell \rfloor}\right).$$

Moreover, this bound holds when some bits of x are fixed in advance.

By a similar calculation as in the proof of Theorem 5.2.1, Lemma 5.6.1 strengthens Theorem 5.2.1 as the following theorem.

Theorem 5.6.2. *For any constant $r \ge 2$, the $2r$-fold Forrelation problem on n-bit inputs*

1. *can be solved with advantage 2^{-10r} by r adaptive rounds of queries with one quantum query per round, yet*
2. *any algorithm with $n^{1/(2r)}$ classical preprocessing queries and $r-1$ adaptive quantum query rounds requires $\widetilde{\Omega}(n^{1/r^2})$ parallel quantum queries to approximate it.*

Now we prove Lemma 5.6.1 which is a simple black-box reduction to Theorem 5.2.3.

Proof of Lemma 5.6.1. We view the classical preprocessing phase of $\mathcal{A}$ as a decision tree $\mathcal{D}$ of depth at most d, where each leaf z of $\mathcal{D}$ selects a quantum query algorithm $\mathcal{B}_z$. In addition, we identify each z as a partial assignment in $\{\pm 1, *\}^n$ where the ± 1 values correspond to classical queries and their outcome, and $*$'s correspond to bits that are not queried in this phase. In particular, there are at most d non-$*$ values for each z and we use $z^{-1}(*) \subseteq [n]$ to denote the entries of these non-$*$ values.

For each z, define $g_z(x) = \mathbf{Pr}[\mathcal{B}_z$ accepts $x|z]$ as the acceptance probability function of $\mathcal{B}_z$ conditioned that x is consistent with z on entries in $z^{-1}(*)$. By Theorem 5.2.3, we have

$$L_k(g_z) \le O_{r,k}\left(t^k \cdot \left(\sqrt{n/t}\right)^{\lfloor (1-\frac{1}{2r})k \rfloor}\right) \quad \text{for each } k \ge 0 \tag{5.37}$$

For each $S \subseteq [n]$ of size ℓ, define $a_S = \mathsf{sgn}(\widehat{f}(S))$ as the sign of the Fourier coefficients at level ℓ. Then we have

$$L_\ell(f) = \mathop{\mathbb{E}}_{\boldsymbol{x}}\left[f(\boldsymbol{x})\sum_{|S|=\ell} a_S\cdot \boldsymbol{x}_S\right] = \mathop{\mathbb{E}}_{\boldsymbol{z}}\left[\mathop{\mathbb{E}}_{\boldsymbol{x}}\left[f(\boldsymbol{x})\sum_{|S|=\ell} a_S\cdot \boldsymbol{x}_S \,\middle|\, \boldsymbol{z}\right]\right]$$

$$= \mathop{\mathbb{E}}_{\boldsymbol{z}} \left[\mathop{\mathbb{E}}_{\boldsymbol{x}} \left[g_{\boldsymbol{z}}(\boldsymbol{x}) \sum_{|S|=\ell} a_S \cdot \boldsymbol{x}_S \,\middle|\, \boldsymbol{z} \right] \right]$$

($\boldsymbol{z}$ is sampled by a random root-to-leaf path in $\mathcal{D}$)

$$= \mathop{\mathbb{E}}_{\boldsymbol{z}} \left[\mathop{\mathbb{E}}_{\boldsymbol{x}} \left[g_{\boldsymbol{z}}(\boldsymbol{x}) \sum_{\substack{T_1 \subseteq \boldsymbol{z}^{-1}(*) \\ T_2 \subseteq [n] \setminus \boldsymbol{z}^{-1}(*) \\ |T_1|+|T_2|=\ell}} a_{T_1 \cup T_2} \cdot \boldsymbol{z}_{T_1} \cdot \boldsymbol{x}_{T_2} \,\middle|\, \boldsymbol{z} \right] \right] \qquad \text{(by the definition of } \boldsymbol{z})$$

$$= \mathop{\mathbb{E}}_{\boldsymbol{z}} \left[\sum_{T_1 \subseteq \boldsymbol{z}^{-1}(*), |T_1| \le \ell} \boldsymbol{z}_{T_1} \cdot \mathop{\mathbb{E}}_{\boldsymbol{x}} \left[g_{\boldsymbol{z}}(\boldsymbol{x}) \sum_{\substack{T_2 \subseteq [n] \setminus \boldsymbol{z}^{-1}(*) \\ |T_2| = \ell - |T_1|}} a_{T_1 \cup T_2} \cdot \boldsymbol{x}_{T_2} \right] \right]$$

(by the definition of $g_{\boldsymbol{z}}$)

$$\le \mathop{\mathbb{E}}_{\boldsymbol{z}} \left[\sum_{T_1 \subseteq \boldsymbol{z}^{-1}(*), |T_1| \le \ell} L_{\ell - |T_1|}(g_{\boldsymbol{z}}) \right] \qquad \text{(by the definition of } L_{\ell-k} \text{ and since } \boldsymbol{z}_{T_1} \in \{\pm 1\})$$

$$\le \sum_{k=0}^{\ell} d^k \cdot \mathop{\mathbb{E}}_{\boldsymbol{z}} \left[L_{\ell-k}(g_{\boldsymbol{z}}) \right] \qquad \text{(since } |\boldsymbol{z}^{-1}(*)| \le d\text{)}$$

$$\le \sum_{k=0}^{\ell} d^k \cdot O_{r,k} \left(t^{\ell-k} \cdot \left(\sqrt{n/t} \right)^{\left\lfloor \left(1 - \frac{1}{2r}\right)(\ell-k) \right\rfloor} \right) \qquad \text{(by (5.37))}$$

$$= O_{r,\ell} \left((d \cdot t)^{\ell} \cdot \left(\sqrt{n/t} \right)^{\left\lfloor \left(1 - \frac{1}{2r}\right)\ell \right\rfloor} \right) \qquad \square$$

Chapter 6

Other Projects

The theme of this thesis is quantum advantages via Fourier growth analysis of Boolean functions, which includes the works described in detail above. There are several other projects that I was involved in and completed during the PhD study, which do not fit within scope. Nevertheless, I give short descriptions on them here in chronological order.

Sketching Algorithm for Edit Distance. This is the work [JNW21] joint with Ce Jin and Jelani Nelson.

The edit distance between two binary strings x and y is defined to be the minimal number of *changes* on x to get y, where each *change* can be deletion, insertion, and substitution. This metric captures DNA edits and document similarities.

We consider the setting where x and y are stored apart by Alice and Bob respectively; and they aim to compute the edit distance (and recover a corresponding edit sequence) by some efficient communication with a referee Carol. More formally, Alice computes a bit-string sketch s_x based on her string x; and Bob computes s_y from y; then Carol should be able to compute the edit distance between x and y solely from s_x and s_y, with high probability. In real life, Carol can be seen as a central server that stores sketches of documents for comparisons of similarity, and Alice Bob are users to upload their documents.

One natural approach for this is to set $s_x = x$ and $s_y = y$. It turns out that in many cases this is not necessary. In our work, we show that if the edit distance between x and y is guaranteed to be small, say, at most k. Then Alice and Bob can efficiently prepare sketches of length roughly k^3 where n is an upper bound on the lengths of x, y. Our bound quantitatively improved the previous k^8 bound by Belazzougui and Zhang [BZ16], and was subsequently improved by Kociumaka, Porat, and Starikovskaya [KPS22] to k^2 and by Koucký and Saks [KS24] to k.

Sample Solutions to Local-Lemma-Type k-CNFs. This is the work [HSW21] joint with Kun He and Xiaoming Sun.

A k-CNF is a conjunction of clauses, where each clause is a disjunction of exactly k variables or their negations. It is well-known that the satisfiability of k-CNFs is NP-complete

even for $k = 3$. However in some cases, the k-CNFs in question have some additional assumption, for example, every variable appears in at most d clauses for d being relatively small. This is what we call local-lemma-type k-CNFs, as the celebrated Lovász local lemma [Erd75] guarantees satisfiability if $d \lesssim 2^k$ and the famous Moser-Tardos algorithm [MT10] efficiently finds a solution in this regime.

Our focus is to sample a uniformly random solution of these local-lemma-type k-CNFs. This is closely related to the partition function in statistical physics and inference of graphical models [Moi19]. One natural approach for this sampling task is to do rejection sampling, which unfortunately fails as the solution space is exponentially small. Another tempting algorithm to try is the metropolis algorithm, which however also fails due to large distances between solutions. We analyzed a Markov chain algorithm that is intuitively metropolis algorithm under projection, and showed how to sample a uniform solution in polynomial time as long as $d \lesssim 2^{0.175k}$. This bound was further improved to $d \lesssim 2^{0.2k}$ by He, Wang, and Yin [HWY23a]. Our algorithm was also the first perfect sampler that samples a *perfect* uniformly random solution, and the idea extended easily to larger alphabets for sampling random coloring of local-lemma-type hypergraphs.

Sample Solutions to Random k-CNFs. This is the work [HWY23b] joint with Kun He and Kuan Yang.

Pertinent to the local-lemma-type k-CNFs above, here we consider random k-CNFs, where each clause is itself uniformly random. This model is particularly interesting as each variable appears limited number of times *in average* but a small fraction of variables will appear for many times *with high probability*. Hence, the Lovász local lemma and Moser-Tardos algorithm, served as a worst-case analysis, do not apply here and the satisfiability of random k-CNFs is highly non-trivial [DSS15] and relates to phase transitions in statistical physics. Nevertheless, let d be the average degree of variables; then for $d \lesssim 2^k$, a random k-CNF is satisfiable with high probability and a solution can be found efficiently [CO10].

Our focus is again trying to sample a solution uniformly at random. Just like the local-lemma-type k-CNFs, the solution space of random k-CNFs is also highly fragmented and prohibits many natural samplers. We analyzed a recursive sampling algorithm inspired by [HWY23a] to sample bit by bit of a random solution, and showed that it runs in polynomial time as long as $d \lesssim 2^{k/3}$. This was a significant improvement over the $d \lesssim 2^{k/300}$ from previous work [GGGY21] and our $2^{k/3}$ threshold was, for the first time, better than the threshold of worst-case local-lemma-type k-CNFs for efficiently sampling solutions. Our result was recently improved by Chen, Lonkar, Wang, Yang, and Yin [CLW+24] to $d \lesssim 2^k$, which, up to low order terms, matches the satisfiability threshold of random k-CNFs [DSS15].

Differential Private Algorithm for The Counting Problem on Trees. This is the work [GKK+23] joint with Badih Ghazi, Pritish Kamath, Ravi Kumar, and Pasin Manurangsi.

Consider the scenario where Google wants to analyze the popularity of each webpage by the number of user clicks on it. This can be abstracted as a rooted tree with some non-negative integers weight on each leaf node, and the task is to compute for each internal node the total weight of leaves below it. Here, the tree structure represents the hierarchical nature of domains and sub-domains, and the value for each internal node represents the total number of user clicks accumulated on it, by visiting its sub-domains. This is, of course, a simple algorithmic task that only needs a linear-time bottom-up dynamic programming. The extra demand here is privacy.

The simple algorithm that outputs each node value is not considered private, as each leaf weight change will influence all its ancestors and can be detected easily. In real life, this is saying that each user's behavior, though subtle, can be detected. A private algorithm should be resilient to such perturbation, i.e., the algorithm's output should not be too sensitive to any individual change. This is formalized in a mathematically rigorous way as *differential privacy* [DKM+06]. In a nutshell, the privacy is guaranteed by blurring the exact output values with some random, but still controlled, shifts. The contribution of our work is differential private algorithms for the above problem with (optimal) trade-offs regarding privacy and accuracy.

Parameterized PCP and Hardness of Approximation. This is the works [GLR+24, GLR+25] joint with Venkatesan Guruswami, Bingkai Lin, Xuandi Ren, and Yican Sun.

The satisfiability of Boolean formula is a well-known **NP**-complete problem, meaning that, given a solution x of the input formula ϕ, we can verify its correctness in polynomial time; however we do not expect to find such a solution in polynomial time, unless $\mathbf{P} = \mathbf{NP}$. The celebrated PCP theorem [AS98, ALM+98] shows that the verification of $\phi(x) = \mathsf{True}$ can be further sped up, if we moderately increase the size of the proof. More precisely, let V be a verifier of the satisfiability of ϕ. If ϕ is satisfiable, then we can write a proof π of binary bits and polynomial length, such that V always accepts the proof π; however if ϕ is not satisfiable, then any proposed proof π will fail to convince V with high probability. Moreover, V only reads a constant number of random locations in π. Of course a natural V will be the one directly verifying if π is a solution of ϕ, but this is inefficient as it necessarily needs to read every bit of π. Indeed the PCP theorem is highly non-trivial.

One motivation for proving such a PCP theorem is to prove hardness of approximation. For example, it is well-known that the clique number is **NP**-hard to compute [Kar09], but what if we are satisfied by a constant approximation of the clique number. With the help of the PCP theorem above, it was shown that this relaxation remains **NP**-hard [FGL+91].

The contribution of our works is to study the trade-off between the alphabet size and proof length in the PCP theorem. More concretely, we show that for any parameter k, the PCP theorem holds with proofs of alphabet size $2^{n/k}$ and length $k^{1+o(1)}$, where n is the size of the input ϕ. This parameterized version recovers the classic PCP theorem when $k = \Theta(n)$, but is more flexible to use. Indeed, it implies that, under the standard assumption *Exponential Time Hypothesis* [IP01], finding a clique of size $k/10$ in a graph promised to have

size-k cliques still requires time $N^{k^{1-o(1)}}$, where N is the size of the input graph. Note that N^k is the runtime of the brute-force algorithm, and hence this bound is essentially optimal.

(Post-Quantum) Security of Cryptographic Salting. This is the work [DLW24] joint with Fangqi Dong and Qipeng Liu.

Imagine that Alice and Bob want to use some public channel for communication, and the adversary Eve aims to steal their secret. As a public-key protocol, the messages between Alice and Bob are broadcasted and the security of the protocol heavily relies on some cryptographic primitive such as hash function f. It is usually the case that Eve succeeds if f is broken, e.g., Eve discovers some $x \neq y$ producing a collision $f(x) = f(y)$.

While the primitive f can be highly complicated to prevent an efficient way of finding collisions, Eve, in real life, can study f with unlimited time and efforts, before executing an efficient attack. This is because both the protocol and the primitive are announced in public in advance. Under this consideration, Eve can trivially store some particular collision as a witness for breaking f. In fact, it is not hard to see that any fixed hash function becomes vulnerable against adversaries with a preprocessing phase like this.

To regain security, Morris and Thompson [MT79] proposed a natural fix, called *cryptographic salting*, as follows. Instead of announcing a single hash function, the public protocol includes many different hash functions $f_1, \ldots, f_K$ and Alice and Bob will later pick a uniform random k from $\{1, 2, \ldots, K\}$ and use f_k for their communication. This heuristically prevents attacks if K is sufficiently large, as a space-bounded adversary cannot store collisions of every hash function possibility now. The contribution of our work is to rigorously analyze this enhanced protocol against both classical and quantum adversaries. Our bound quantitatively improves the prior works [CDGS18, Liu23]; in fact, our result achieves asymptotically optimal security parameter in most settings and generalizes beyond hash functions and colision finding problems.

Optimal Clifford+T Circuit for Preparing Quantum States. This is the work [GKW24] joint with David Gosset and Robin Kothari.

Quantum state preparation is a fundamental subroutine in many quantum algorithms, metrologies, communication, and more. In this problem, we are given a classical description of some n-qubit quantum state $|\psi\rangle$ and aim to construct it from scratch. More formally, we start with the all-zero state, then apply some amount of quantum gates, to convert the state into a quantum state that is close enough to the target $|\psi\rangle$. A natural and standard choice of quantum gates is the Clifford+T gate set, consisting of Clifford gates (Hadamard gate, phase gate, and CNOT gate) and T gate. Interestingly, T gate is considered more expensive than Clifford gates for various reasons from classical simulation [GOT98], quantum error correction [BK05], and quantum magic [VMGE14]. Hence, reducing the number of T gates used in the state preparation scheme becomes a vital task.

While an n-qubit state has roughly 2^n degrees of freedom, surprisingly it can always be prepared in a Clifford+T circuit with roughly $\sqrt{2^n \cdot n \log n}$ T gates due to Low, Kliuchnikov,

Schaeffer [LKS24]. They also proved a lower bound of $\sqrt{2^n}$, almost matching the upper bound. The main contribution of our work is to bring down the upper bound to $\sqrt{2^n}$, and, more generally, to achieve an optimal trade-off regarding the dimension n, the approximation error ε, and the number of T gates in the state preparation circuit. Our techniques also have interesting consequences and savings for implementing certain unitaries.

Sampling Power of Shallow Boolean Circuits. This is the works [KOW24, KOW25] joint with Daniel M. Kane and Anthony Ostuni.

A Boolean circuit consists of AND, OR, and negation gates with bounded fan-in, as well as designated input gates and output gates. The depth of a Boolean circuit is the maximum length of a computation path from input to output. We consider shallow Boolean circuits that have constant (or slightly superconstant) depth. This computation model is typically not interesting, as every output bit depends only on constant number of input bits; hence it cannot compute any function that needs to aggregate information of many input bits. For example, the parity function on n bits is obviously not computable here.

However things become interesting as we move to sampling tasks, where we view input bits as independent unbiased coins and aim to produce some target output distribution. In this setting, we can actually sample a distribution naturally induced by the parity function: uniform n-bit strings with even Hamming weight. To see this, assume $r_1, \ldots, r_n$ are unbiased coins, then the joint distribution of $(r_1 \oplus r_2, r_2 \oplus r_3, \ldots, r_n \oplus r_1)$ is uniform over even strings. If we could compute the parity function, this distribution can be sampled in a more direct way by $(r_1, \ldots, r_{n-1}, r_1 \oplus r_2 \oplus \cdots \oplus r_{n-1})$, which is, of course, not feasible in our model.

It is now natural to ask if there are other examples like above: can we use shallow circuits to sample a uniform n-bit string conditioned on some Hamming weight constraint, despite that the constraint function itself is not computable by shallow circuits. Our work shows that unfortunately the above example is the only miracle, confirming a conjecture by Filmus, Leigh, Riazanov, and Sokolov [FLRS23]. Our result also leads to interesting applications in data structure lower bounds [Vio12], quantum-classical separation [BGK18], and learning theory [Dia16].

Bibliography

[AA18] Scott Aaronson and Andris Ambainis. Forrelation: A problem that optimally separates quantum from classical computing. *SIAM J. Comput.*, 47(3):982–1038, 2018. doi:10.1137/15M1050902. 5, 12, 13, 16, 58, 119, 120

[AAB+19] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando GSL Brandao, David A Buell, et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 2019. 1

[Aar10] Scott Aaronson. BQP and the polynomial hierarchy. In *STOC*, pages 141–150, 2010. URL: http://doi.acm.org/10.1145/1806689.1806711, doi:10.1145/1806689.1806711. 5, 57

[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. 58

[ACC+23] Atul Singh Arora, Andrea Coladangelo, Matthew Coudron, Alexandru Gheorghiu, Uttam Singh, and Hendrik Waldner. Quantum depth in the random oracle model. In Barna Saha and Rocco A. Servedio, editors, *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023, Orlando, FL, USA, June 20-23, 2023*, pages 1111–1124, USA, 2023. ACM. doi:10.1145/3564246.3585153. 121

[AdW14] Andris Ambainis and Ronald de Wolf. How low can approximate degree and quantum query complexity be for total boolean functions? *Comput. Complex.*, 23(2):305–322, 2014. URL: https://doi.org/10.1007/s00037-014-0083-2, doi:10.1007/S00037-014-0083-2. 119

[AGL+23] Dorit Aharonov, Xun Gao, Zeph Landau, Yunchao Liu, and Umesh Vazirani. A polynomial-time classical algorithm for noisy random circuit sampling. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 945–957, 2023. 1

[Agr20] Rohit Agrawal. Coin theorems and the fourier expansion. *Chic. J. Theor. Comput. Sci.*, 2020, 2020. URL: http://cjtcs.cs.uchicago.edu/articles/2020/4/contents.html. 3, 55

[ALM+98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM (JACM)*, 45(3):501–555, 1998. 151

[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. URL: https://doi.org/10.1214%2F19-aihp1041, doi:10.1214/19-aihp1041. 68

[AS98] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: A new characterization of np. *Journal of the ACM (JACM)*, 45(1):70–122, 1998. 151

[BB20] Shalev Ben-David and Eric Blais. A tight composition theorem for the randomized query complexity of partial functions: Extended abstract. In *FOCS*, pages 240–246. IEEE, 2020. 17

[BCW98] Harry Buhrman, Richard Cleve, and Avi Wigderson. Quantum vs. classical communication and computation. In *STOC*, pages 63–68. ACM, 1998. 57

[BGGS22] Sergey Bravyi, David Gosset, Daniel Grier, and Luke Schaeffer. Classical algorithms for forrelation. *CoRR*, 2022. URL: https://arxiv.org/abs/2102.06963, arXiv:2102.06963. 5, 120

[BGK18] Sergey Bravyi, David Gosset, and Robert König. Quantum advantage with shallow circuits. *Science*, 362(6412):308–311, 2018. 153

[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. 4

[BK05] Sergey Bravyi and Alexei Kitaev. Universal quantum computation with ideal clifford gates and noisy ancillas. *Physical Review A—Atomic, Molecular, and Optical Physics*, 71(2):022316, 2005. 152

[Bon70] Aline Bonami. Étude des coefficients de fourier des fonctions de $l^p(g)$. *Annales de l'institut Fourier*, 20(2):335–402, 1970. URL: http://eudml.org/doc/74019. 11

[Bor75] Christer Borell. The brunn-minkowski inequality in gauss space. *Inventiones mathematicae*, 30(2):207–216, 1975. 9

[BP18] Joseph Briggs and Wesley Pegden. Extremal collections of k-uniform vectors. *arXiv preprint arXiv:1801.09609*, 2018. 42

[BS21] Nikhil Bansal and Makrand Sinha. k-forrelation optimally separates quantum and classical query complexity. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 1303–1316, USA, 2021. ACM. doi:10.1145/3406325.3451040. 3, 4, 5, 13, 16, 17, 58, 119

[BTW15] Eric Blais, Li-Yang Tan, and Andrew Wan. An inequality for the fourier spectrum of parity decision trees. *CoRR*, abs/1506.01055, 2015. URL: http://arxiv.org/abs/1506.01055, arXiv:1506.01055. 4, 16, 24, 29, 53

[Bur19] Paul Burchard. Lower bounds for parallel quantum counting. *arXiv preprint arXiv:1910.04555*, 2019. 117

[BV97] Ethan Bernstein and Umesh Vazirani. Quantum complexity theory. *SIAM Journal on Computing*, 26(5):1411–1473, 1997. doi:10.1137/S0097539796300921. 4, 117

[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. doi:10.1109/FOCS.2010.10. 55

[BZ16] Djamal Belazzougui and Qin Zhang. Edit distance: Sketching, streaming, and document exchange. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 51–60. IEEE, 2016. 149

[CDGS18] Sandro Coretti, Yevgeniy Dodis, Siyao Guo, and John Steinberger. Random oracles and non-uniformity. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 227–258. Springer, 2018. 152

[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. 57, 58, 114

[CGL+21] Eshan Chattopadhyay, Jason Gaitonde, Chin Ho Lee, Shachar Lovett, and Abhishek Shetty. Fractional pseudorandom generators from any fourier level. In Valentine Kabanets, editor, *36th Computational Complexity Conference, CCC 2021, July 20-23, 2021, Toronto, Ontario, Canada (Virtual Conference)*, volume 200 of *LIPIcs*, pages 10:1–10:24, USA, 2021. Schloss Dagstuhl - Leibniz-Zentrum für Informatik. URL: https://doi.org/10.4230/LIPIcs.CCC.2021.10, doi:10.4230/LIPICS.CCC.2021.10. 3

[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. 55

[CH23] Nai-Hui Chia and Shih-Han Hung. Non-interactive classical verification of quantum depth: A fine-grained characterization. *IACR Cryptol. ePrint Arch.*, page 1911, 2023. URL: https://eprint.iacr.org/2023/1911. 121

[CHHL19] Eshan Chattopadhyay, Pooya Hatami, Kaave Hosseini, and Shachar Lovett. Pseudorandom generators from polarizing random walks. *Theory Comput.*, 15:1–26, 2019. URL: https://doi.org/10.4086/toc.2019.v015a010, doi:10.4086/TOC.2019.V015A010. 3, 4

[CHLT19] Eshan Chattopadhyay, Pooya Hatami, Shachar Lovett, and Avishay Tal. Pseudorandom generators from the second fourier level and applications to AC0 with parity gates. In *ITCS*, volume 124 of *LIPIcs*, pages 22:1–22:15. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019. 3, 4, 60

[Cho19] Adrian Cho. Ibm casts doubt on google's claims of quantum supremacy. *Science*, 2019. 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. 3, 4

[CKLM19] Arkadev Chattopadhyay, Michal Koucký, Bruno Loff, and Sagnik Mukhopadhyay. Simulation theorems via pseudo-random properties. *Comput. Complex.*, 28(4):617–659, 2019. 57

[CLW+24] Zongchen Chen, Aditya Lonkar, Chunyang Wang, Kuan Yang, and Yitong Yin. Counting random k-sat near the satisfiability threshold. *arXiv preprint arXiv:2411.02980*, 2024. 150

[CLZ22] Yilei Chen, Qipeng Liu, and Mark Zhandry. Quantum algorithms for variants of average-case lattice problems via filtering. In *Annual international conference on the theory and applications of cryptographic techniques*, pages 372–401. Springer, 2022. 1

[CM20] Matthew Coudron and Sanketh Menda. Computations with greater quantum depth are strictly more powerful (relative to an oracle). In Konstantin Makarychev, Yury Makarychev, Madhur Tulsiani, Gautam Kamath, and Julia Chuzhoy, editors, *Proccedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, Chicago, IL, USA, June 22-26, 2020*, pages 889–901, USA, 2020. ACM. doi:10.1145/3357713.3384269. 121

[CO10] Amin Coja-Oghlan. A better algorithm for random k-sat. *SIAM Journal on Computing*, 39(7):2823–2864, 2010. 150

[CPTS21] Gil Cohen, Noam Peri, and Amnon Ta-Shma. Expander random walks: A fourier-analytic approach. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1643–1655, 2021. 3, 19

[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. doi:10.1137/120861072. 56, 65

[CS16] Gil Cohen and Igor Shinkar. The complexity of DNF of parities. In *ITCS*, pages 47–58. ACM, 2016. 15

[CW00] Richard Cleve and John Watrous. Fast parallel circuits for the quantum fourier transform. In *41st Annual Symposium on Foundations of Computer Science, FOCS 2000, 12-14 November 2000, Redondo Beach, California, USA*, pages 526–536, USA, 2000. IEEE Computer Society. doi:10.1109/SFCS.2000.892140. 117

[Dia16] Ilias Diakonikolas. Learning structured distributions. *Handbook of Big Data*, 267:10–1201, 2016. 153

[DJ92] David Deutsch and Richard Jozsa. Rapid solution of problems by quantum computation. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 439(1907):553–558, 1992. doi:10.1098/rspa.1992.0167. 4, 117

[DKM+06] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *Advances in cryptology-EUROCRYPT 2006: 24th annual international conference on the theory and applications of cryptographic techniques, st. Petersburg, Russia, May 28-June 1, 2006. proceedings 25*, pages 486–503. Springer, 2006. 151

[DLW24] Fangqi Dong, Qipeng Liu, and Kewen Wu. Tight characterizations for preprocessing against cryptographic salting. In *Annual International Cryptology Conference*, pages 377–411. Springer, 2024. 152

[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. 57

[DSS15] Jian Ding, Allan Sly, and Nike Sun. Proof of the satisfiability conjecture for large k. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 59–68, 2015. 150

[EI22] Alexandros Eskenazis and Paata Ivanisvili. Learning low-degree functions from a logarithmic number of random queries. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 203–207, USA, 2022. ACM. doi:10.1145/3519935.3519981. 120

[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. 9

[Erd75] Paul Erdős. Problems and results on 3-chromatic hypergraphs and some related questions. *(No Title)*, page 609, 1975. 150

[FGL+91] Uriel Feige, Shafi Goldwasser, László Lovász, Shmuel Safra, and Mario Szegedy. Approximating clique is almost np-complete. In *[1991] Proceedings 32nd Annual Symposium of Foundations of Computer Science*, pages 2–12. IEEE Computer Society, 1991. 151

[FLRS23] Yuval Filmus, Itai Leigh, Artur Riazanov, and Dmitry Sokolov. Sampling and certifying symmetric functions. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, 2023. 153

[Gav20] Dmitry Gavinsky. Entangled simultaneity versus classical interactivity in communication complexity. *IEEE Trans. Inf. Theory*, 66(7):4641–4651, 2020. doi:10.1109/TIT.2020.2976074. 58

[GGGY21] Andreas Galanis, Leslie Ann Goldberg, Heng Guo, and Kuan Yang. Counting solutions to random cnf formulas. *SIAM Journal on Computing*, 50(6):1701–1738, 2021. 150

[GKK+23] Badih Ghazi, Pritish Kamath, Ravi Kumar, Pasin Manurangsi, and Kewen Wu. On differentially private counting on trees. In *50th International Colloquium on Automata, Languages, and Programming (ICALP 2023)*, pages 66–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2023. 150

[GKPW19] Mika Göös, Pritish Kamath, Toniann Pitassi, and Thomas Watson. Query-to-communication lifting for P NP. *Comput. Complex.*, 28(1):113–144, 2019. 57

[GKW24] David Gosset, Robin Kothari, and Kewen Wu. Quantum state preparation with optimal t-count. *arXiv preprint arXiv:2411.04790*, 2024. Appeared in QIP 2025. 152

[GLM+15] Mika Göös, Shachar Lovett, Raghu Meka, Thomas Watson, and David Zuckerman. Rectangles are nonnegative juntas. In *STOC*, pages 257–266. ACM, 2015. 57

[GLR+24] Venkatesan Guruswami, Bingkai Lin, Xuandi Ren, Yican Sun, and Kewen Wu. Parameterized inapproximability hypothesis under exponential time hypothesis. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 24–35, 2024. 151

[GLR+25] Venkatesan Guruswami, Bingkai Lin, Xuandi Ren, Yican Sun, and Kewen Wu. Almost optimal time lower bound for approximating parameterized clique, csp, and more, under eth. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, 2025. 151

[Göö15] Mika Göös. Lower bounds for clique vs. independent set. In *FOCS*, pages 1066–1076. IEEE Computer Society, 2015. 57

[GOT98] D GOTTESMAN. The heisenberg representation of quantum computers. In *Proc. XXII International Colloquium on Group Theoretical Methods in Physics, 1998*, pages 32–43, 1998. 152

[GPW15] Mika Göös, Toniann Pitassi, and Thomas Watson. Deterministic communication vs. partition number. In *FOCS*, pages 1077–1088. IEEE Computer Society, 2015. 57

[GPW20] Mika Göös, Toniann Pitassi, and Thomas Watson. Query-to-communication lifting for BPP. *SIAM J. Comput.*, 49(4), 2020. 57, 60, 113, 114

[Gro96] Lov K. Grover. A fast quantum mechanical algorithm for database search. In Gary L. Miller, editor, *Proceedings of the Twenty-Eighth Annual ACM Symposium on the Theory of Computing, Philadelphia, Pennsylvania, USA, May 22-24, 1996*, pages 212–219, USA, 1996. ACM. doi:10.1145/237814.237866. 4, 117

[GRT22] Uma Girish, Ran Raz, and Avishay Tal. Quantum versus randomized communication complexity, with efficient players. *Comput. Complex.*, 31(2):17, 2022. URL: https://doi.org/10.1007/s00037-022-00232-7, doi:10.1007/S00037-022-00232-7. 3, 4, 6, 18, 53, 54, 55, 57, 58, 97

[GRZ21] Uma Girish, Ran Raz, and Wei Zhan. Lower bounds for xor of forrelations. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, 2021. 4, 53

[GSTW16] Parikshit Gopalan, Rocco A. Servedio, Avishay Tal, and Avi Wigderson. Degree and sensitivity: tails of two distributions. *CoRR*, abs/1604.07432, 2016. URL: http://arxiv.org/abs/1604.07432, arXiv:1604.07432. 4

[GSTW23] Uma Girish, Makrand Sinha, Avishay Tal, and Kewen Wu. Fourier growth of communication protocols for XOR functions. In *64th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2023, Santa Cruz, CA, USA,*

November 6-9, 2023, pages 721–732, USA, 2023. IEEE. doi:10.1109/FOCS57990.2023.00047. 6

[GSTW24] Uma Girish, Makrand Sinha, Avishay Tal, and Kewen Wu. The power of adaptivity in quantum query algorithms. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1488–1497, 2024. 6

[GTW21] Uma Girish, Avishay Tal, and Kewen Wu. Fourier growth of parity decision trees. In Valentine Kabanets, editor, *36th Computational Complexity Conference, CCC 2021, July 20-23, 2021, Toronto, Ontario, Canada (Virtual Conference)*, volume 200 of *LIPIcs*, pages 39:1–39:36, USA, 2021. Schloss Dagstuhl - Leibniz-Zentrum für Informatik. URL: https://doi.org/10.4230/LIPIcs.CCC.2021.39, doi:10.4230/LIPICS.CCC.2021.39. 5, 6, 53, 59

[HG22] Atsuya Hasegawa and François Le Gall. An optimal oracle separation of classical and quantum hybrid schemes. In Sang Won Bae and Heejin Park, editors, *33rd International Symposium on Algorithms and Computation, ISAAC 2022, December 19-21, 2022, Seoul, Korea*, volume 248 of *LIPIcs*, pages 6:1–6:14, USA, 2022. Schloss Dagstuhl - Leibniz-Zentrum für Informatik. URL: https://doi.org/10.4230/LIPIcs.ISAAC.2022.6, doi:10.4230/LIPICS.ISAAC.2022.6. 121

[HHL18] Hamed Hatami, Kaave Hosseini, and Shachar Lovett. Structure of protocols for XOR functions. *SIAM J. Comput.*, 47(1):208–217, 2018. 15, 52, 57, 60

[HS05] Peter Høyer and Robert Spalek. Lower bounds on quantum query complexity. *Bull. EATCS*, 87:78–103, 2005. 12

[HSW21] Kun He, Xiaoming Sun, and Kewen Wu. Perfect sampling for (atomic) lov\'asz local lemma. *arXiv preprint arXiv:2107.03932*, 2021. 149

[HWY23a] Kun He, Chunyang Wang, and Yitong Yin. Deterministic counting lovász local lemma beyond linear programming. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 3388–3425. SIAM, 2023. 150

[HWY23b] Kun He, Kewen Wu, and Kuan Yang. Improved bounds for sampling solutions of random cnf formulas. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 3330–3361. SIAM, 2023. 150

[INW94] Russell Impagliazzo, Noam Nisan, and Avi Wigderson. Pseudorandomness for network algorithms. In *STOC*, pages 356–364. ACM, 1994. 60

[IP01] Russell Impagliazzo and Ramamohan Paturi. On the complexity of k-sat. *Journal of Computer and System Sciences*, 62(2):367–375, 2001. 151

[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. 4, 120

[JMdW17] Stacey Jeffery, Frédéric Magniez, and Ronald de Wolf. Optimal parallel quantum query algorithms. *Algorithmica*, 79(2):509–529, 2017. URL: https://doi.org/10.1007/s00453-016-0206-z, doi:10.1007/S00453-016-0206-Z. 117, 118

[JNW21] Ce Jin, Jelani Nelson, and Kewen Wu. An improved sketching algorithm for edit distance. In *38th International Symposium on Theoretical Aspects of Computer Science*, 2021. 149

[JRS02] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. The quantum communication complexity of the pointer chasing problem: The bit version. In Manindra Agrawal and Anil Seth, editors, *FST TCS 2002: Foundations of Software Technology and Theoretical Computer Science, 22nd Conference Kanpur, India, December 12-14, 2002, Proceedings*, volume 2556 of *Lecture Notes in Computer Science*, pages 218–229, USA, 2002. Springer. doi:10.1007/3-540-36206-1_20. 121

[Kar09] Richard M Karp. Reducibility among combinatorial problems. In *50 Years of Integer Programming 1958-2008: from the Early Years to the State-of-the-Art*, pages 219–241. Springer, 2009. 151

[KLPY10] Pascal Koiran, Jürgen Landes, Natacha Portier, and Penghui Yao. Adversary lower bounds for nonadaptive quantum algorithms. *J. Comput. Syst. Sci.*, 76(5):347–355, 2010. URL: https://doi.org/10.1016/j.jcss.2009.10.007, doi:10.1016/J.JCSS.2009.10.007. 117

[KM93] Eyal Kushilevitz and Yishay Mansour. Learning decision trees using the fourier spectrum. *SIAM J. Comput.*, 22(6):1331–1348, 1993. 2, 3, 15

[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. 57

[KNTZ01] Hartmut Klauck, Ashwin Nayak, Amnon Ta-Shma, and David Zuckerman. Interaction in quantum communication and the complexity of set disjointness. In Jeffrey Scott Vitter, Paul G. Spirakis, and Mihalis Yannakakis, editors, *Proceedings on 33rd Annual ACM Symposium on Theory of Computing, July 6-8, 2001, Heraklion, Crete, Greece*, pages 124–133, USA, 2001. ACM. doi:10.1145/380752.380786. 121

[KOW24] Daniel M Kane, Anthony Ostuni, and Kewen Wu. Locality bounds for sampling hamming slices. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1279–1286, 2024. 153

[KOW25] Daniel M Kane, Anthony Ostuni, and Kewen Wu. Locally sampleable uniform symmetric distributions. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, 2025. 153

[KPS22] Tomasz Kociumaka, Ely Porat, and Tatiana Starikovskaya. Small-space and streaming pattern matching with k edits. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 885–896. IEEE, 2022. 149

[KQS15] Raghav Kulkarni, Youming Qiao, and Xiaoming Sun. On the power of parity queries in boolean decision trees. In *TAMC*, volume 9076 of *Lecture Notes in Computer Science*, pages 99–109. Springer, 2015. 15

[Kra10] Joshua Brown Kramer. On the most weight w vectors in a dimension k binary code. *Electron. J. Comb.*, 17(1), 2010. 42

[KS24] Michal Kouckỳ and Michael E Saks. Almost linear size edit distance sketch. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 956–967, 2024. 149

[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. doi:10.4230/LIPIcs.CCC.2019.7. 4

[Liu23] Qipeng Liu. Non-uniformity and quantum advice in the quantum random oracle model. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 117–143. Springer, 2023. 152

[LKS24] Guang Hao Low, Vadym Kliuchnikov, and Luke Schaeffer. Trading t gates for dirty qubits in state preparation and unitary synthesis. *Quantum*, 8:1375, 2024. 153

[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. 57, 114

[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 and 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, USA, 2022. Schloss Dagstuhl - Leibniz-Zentrum für Informatik. URL: https://doi.org/10.4230/LIPIcs.APPROX/RANDOM.2022.2, doi:10.4230/LIPICS.APPROX/RANDOM.2022.2. 4

[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. 57

[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. 55

[LV18] Chin Ho Lee and Emanuele Viola. The coin problem for product tests. *ACM Transactions on Computation Theory (TOCT)*, 10(3):1–10, 2018. 55

[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. 3, 4

[MN01] Cristopher Moore and Martin Nilsson. Parallel quantum computation and quantum codes. *SIAM J. Comput.*, 31(3):799–815, 2001. doi:10.1137/S0097539799355053. 117

[MNR11] Ashley Montanaro, Harumichi Nishimura, and Rudy Raymond. Unbounded-error quantum query complexity. *Theor. Comput. Sci.*, 412(35):4619–4628, 2011. doi:10.1016/j.tcs.2011.04.043. 18

[MO09] Ashley Montanaro and Tobias Osborne. On the communication complexity of XOR functions. *CoRR*, abs/0909.3392, 2009. 15, 52

[Moi19] Ankur Moitra. Approximate counting, the lovász local lemma, and inference in graphical models. *Journal of the ACM (JACM)*, 66(2):1–25, 2019. 150

[Mon10] Ashley Montanaro. Nonadaptive quantum query complexity. *Inf. Process. Lett.*, 110(24):1110–1113, 2010. doi:10.1016/j.ipl.2010.09.009. 12, 117

[MS20] Nikhil S. Mande and Swagato Sanyal. On parity decision trees for fourier-sparse boolean functions. In *FSTTCS*, volume 182 of *LIPIcs*, pages 29:1–29:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020. 15

[MT79] Robert Morris and Ken Thompson. Password security: A case history. *Communications of the ACM*, 22(11):594–597, 1979. 152

[MT10] Robin A Moser and Gábor Tardos. A constructive proof of the general lovász local lemma. *Journal of the ACM (JACM)*, 57(2):1–15, 2010. 150

[MVM+24] Alexis Morvan, B Villalonga, X Mi, S Mandra, A Bengtsson, PV Klimov, Z Chen, S Hong, C Erickson, IK Drozdov, et al. Phase transitions in random circuit sampling. *Nature*, 634(8033):328–333, 2024. 1

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM J. Comput.*, 22(4):838–856, 1993. 2, 3

[NY04] Harumichi Nishimura and Tomoyuki Yamakami. An algorithmic argument for nonadaptive query complexity lower bounds on advised quantum computation (extended abstract). In Jirí Fiala, Václav Koubek, and Jan Kratochvíl, editors, *Mathematical Foundations of Computer Science 2004, 29th International Symposium, MFCS 2004, Prague, Czech Republic, August 22-27, 2004, Proceedings*, volume 3153 of *Lecture Notes in Computer Science*, pages 827–838, USA, 2004. Springer. doi:10.1007/978-3-540-28629-5_65. 117

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. 2, 10, 11, 15, 17

[OS07] Ryan O'Donnell and Rocco A. Servedio. Learning monotone decision trees in polynomial time. *SIAM Journal on Computing*, 37(3):827–844, 2007. doi:10.1137/060669309. 4, 53

[OWZ+14] Ryan O'Donnell, John Wright, Yu Zhao, Xiaorui Sun, and Li-Yang Tan. A composition theorem for parity kill number. In *Computational Complexity Conference*, pages 144–154. IEEE Computer Society, 2014. 15

[PCZ22] Feng Pan, Keyang Chen, and Pan Zhang. Solving the sampling problem of the sycamore quantum circuits. *PRL*, 2022. 1

[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. 4

[Raz95] Ran Raz. Fourier analysis for probabilistic communication complexity. *Comput. Complex.*, 5(3/4):205–221, 1995. doi:10.1007/BF01206318. 52

[Reg25] Oded Regev. An efficient quantum factoring algorithm. *Journal of the ACM*, 72(1):1–13, 2025. 117

[RM99] Ran Raz and Pierre McKenzie. Separation of the monotone NC hierarchy. *Comb.*, 19(3):403–435, 1999. 57

[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. 57

[RS10] Alexander A. Razborov and Alexander A. Sherstov. The sign-rank of ac^0. *SIAM J. Comput.*, 39(5):1833–1855, 2010. 57

[RSV13] Omer Reingold, Thomas Steinke, and Salil P. Vadhan. Pseudorandomness for regular branching programs via fourier analysis. In *APPROX-RANDOM*, volume 8096 of *Lecture Notes in Computer Science*, pages 655–670. Springer, 2013. 4

[RT22] Ran Raz and Avishay Tal. Oracle separation of BQP and PH. *J. ACM*, 69(4):30:1–30:21, 2022. doi:10.1145/3530258. 2, 3, 4, 5, 119

[RY22] Anup Rao and Amir Yehudayoff. Anticoncentration and the exact gap-hamming problem. *SIAM Journal on Discrete Mathematics*, 36(2):1071–1092, 2022. arXiv:https://doi.org/10.1137/21M1435288, doi:10.1137/21M1435288. 56

[San19] Swagato Sanyal. Fourier sparsity and dimension. *Theory Comput.*, 15:1–13, 2019. 15

[She11] Alexander A. Sherstov. The pattern matrix method. *SIAM J. Comput.*, 40(6):1969–2000, 2011. 57

[She12] Alexander A. Sherstov. The communication complexity of gap hamming distance. *Theory Comput.*, 8(1):197–208, 2012. doi:10.4086/toc.2012.v008a008. 56

[Sho99] Peter W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Rev.*, 41(2):303–332, 1999. doi:10.1137/S0036144598347011. 4, 117

[Sim83] Hans Ulrich Simon. A tight omega(loglog n)-bound on the time for parallel ram's to compute nondegenerated boolean functions. In Marek Karpinski, editor, *Fundamentals of Computation Theory, Proceedings of the 1983 International FCT-Conference, Borgholm, Sweden, August 21-27, 1983*, volume 158 of *Lecture Notes in Computer Science*, pages 439–444, USA, 1983. Springer. doi:10.1007/3-540-12689-9_124. 118

[Sim97] Daniel R. Simon. On the power of quantum computation. *SIAM J. Comput.*, 26(5):1474–1483, October 1997. doi:10.1137/S0097539796298637. 4, 117

[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. doi:10.1145/28395.28404. 4

[SSW23] Alexander A. Sherstov, Andrey A. Storozhenko, and Pei Wu. An optimal separation of randomized and quantum query complexity. *SIAM J. Comput.*, 52(2):525–567, 2023. URL: https://doi.org/10.1137/22m1468943, doi:10.1137/22M1468943. 4, 5, 15, 16, 17, 23, 24, 53, 58, 60, 114, 120

[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. 9

[STlV17] Amir Shpilka, Avishay Tal, and Ben lee Volk. On the structure of boolean functions with small spectral norm. *Comput. Complex.*, 26(1):229–273, 2017. 15

[SVW17] Thomas Steinke, Salil P. Vadhan, and Andrew Wan. Pseudorandomness and fourier-growth bounds for width-3 branching programs. *Theory Comput.*, 13(1):1–50, 2017. URL: https://doi.org/10.4086/toc.2017.v013a012, doi:10.4086/TOC.2017.V013A012. 4

[SZ08] Yaoyun Shi and Zhiqiang Zhang. Communication complexities of xor functions. *arXiv preprint arXiv:0808.1762*, 2008. 52

[SZ09] Yaoyun Shi and Yufan Zhu. Quantum communication complexity of block-composed functions. *Quantum Inf. Comput.*, 9(5&6):444–460, 2009. 57

[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 für Informatik, 2017. 4

[Tal20] Avishay Tal. Towards optimal separations between quantum and randomized query complexities. In Sandy Irani, editor, *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 228–239, USA, 2020. IEEE. doi:10.1109/FOCS46700.2020.00030. 4, 5, 15, 16, 17, 23, 53, 60, 67, 114, 119, 120

[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. doi:10.1109/FOCS.2013.76. 15, 52

[vD98] Wim van Dam. Quantum oracle interrogation: Getting all information for almost half the price. In *39th Annual Symposium on Foundations of Computer Science, FOCS '98, November 8-11, 1998, Palo Alto, California, USA*, pages 362–367, USA, 1998. IEEE Computer Society. doi:10.1109/SFCS.1998.743486. 117

[Ver18] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018. 9

[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. URL: http://cjtcs.cs.uchicago.edu/articles/2012/1/contents.html. 56, 65

[Vio12] Emanuele Viola. The complexity of distributions. *SIAM Journal on Computing*, 41(1):191–218, 2012. 153

[VMGE14] Victor Veitch, SA Hamed Mousavian, Daniel Gottesman, and Joseph Emerson. The resource theory of stabilizer quantum computation. *New Journal of Physics*, 16(1):013009, 2014. 152

[Wat09] John Watrous. Quantum computational complexity. In Robert A. Meyers, editor, *Encyclopedia of Complexity and Systems Science*, pages 7174–7201. Springer, USA, 2009. doi:10.1007/978-0-387-30440-3_428. 121

[Wik23] Wikipedia. Adjugate matrix. http://en.wikipedia.org/w/index.php?title=Adjugate%20matrix&oldid=1131417820, 2023. [Online; accessed 13-February-2023]. 137

[Wu22] Xinyu Wu. A stochastic calculus approach to the oracle separation of BQP and PH. *Theory Comput.*, 18:1–11, 2022. URL: https://theoryofcomputing.org/articles/v018a017/. 3

[WYY17] Xiaodi Wu, Penghui Yao, and Henry S. Yuen. Raz-mckenzie simulation with the inner product gadget. *Electron. Colloquium Comput. Complex.*, 24:10, 2017. 57

[Zal99] Christof Zalka. Grover's quantum searching algorithm is optimal. *Phys. Rev. A*, 60:2746–2751, Oct 1999. URL: https://link.aps.org/doi/10.1103/PhysRevA.60.2746, doi:10.1103/PhysRevA.60.2746. 117, 118

[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. 52

[ZS09] Zhiqiang Zhang and Yaoyun Shi. Communication complexities of symmetric XOR functions. *Quantum Inf. Comput.*, 9(3&4):255–263, 2009. 15

[ZS10] Zhiqiang Zhang and Yaoyun Shi. On the parity complexity measures of boolean functions. *Theor. Comput. Sci.*, 411(26-28):2612–2618, 2010. 15