An Optimal Separation of Randomized and Quantum Query Complexity

Alexander A. Sherstov
University of California, Los Angeles
Los Angeles, California, USA
sherstov@cs.ucla.edu

Andrey A. Storozhenko
University of California, Los Angeles
Los Angeles, California, USA
storozhenko@cs.ucla.edu

Pei Wu
University of California, Los Angeles
Los Angeles, California, USA
pwu@cs.ucla.edu

ABSTRACT

We prove that for every decision tree, the absolute values of the Fourier coefficients of given order $t \geqslant 1$ sum to at most $(cd/t)^{t/2}(1+\log n)^{(t-1)/2}$, where n is the number of variables, d is the tree depth, and $c > 0$ is an absolute constant. This bound is essentially tight and settles a conjecture due to Tal (arxiv 2019; FOCS 2020). The bounds prior to our work degraded rapidly with t, becoming trivial already at $t = \sqrt{d}$.

As an application, we obtain, for every integer $k \geqslant 1$, a partial Boolean function on n bits that has bounded-error quantum query complexity at most $\lceil k/2 \rceil$ and randomized query complexity $\tilde{\Omega}(n^{1-1/k})$. This separation of bounded-error quantum versus randomized query complexity is best possible, by the results of Aaronson and Ambainis (STOC 2015). Prior to our work, the best known separation was polynomially weaker: $O(1)$ versus $\Omega(n^{2/3-\varepsilon})$ for any $\varepsilon > 0$ (Tal, FOCS 2020).

As another application, we obtain an essentially optimal separation of $O(\log n)$ versus $\Omega(n^{1-\varepsilon})$ for bounded-error quantum versus randomized communication complexity, for any $\varepsilon > 0$. The best previous separation was polynomially weaker: $O(\log n)$ versus $\Omega(n^{2/3-\varepsilon})$ (implicit in Tal, FOCS 2020).

CCS CONCEPTS

• **Theory of computation → Quantum complexity theory**; **Problems, reductions and completeness**; *Probabilistic computation.*

KEYWORDS

quantum-classical separations, query complexity, communication complexity, forrelation, Fourier analysis of Boolean functions, Fourier weight of decision trees

ACM Reference Format:
Alexander A. Sherstov, Andrey A. Storozhenko, and Pei Wu. 2021. An Optimal Separation of Randomized and Quantum Query Complexity. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC '21), June 21–25, 2021, Virtual, Italy.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3406325.3451019

STOC '21, June 21–25, 2021, Virtual, Italy

ACM ISBN 978-1-4503-8053-9/21/06.
https://doi.org/10.1145/3406325.3451019

1 INTRODUCTION

Understanding the relative power of quantum and classical computing is of basic importance in theoretical computer science. This question has been studied most actively in the *query model*, which is tractable enough to allow unconditional lower bounds yet rich enough to capture most of the known quantum algorithms. Illustrative examples include the quantum algorithms of Deutsch and Jozsa [13], Bernstein and Vazirani [6], Grover [16], and Shor's period-finding [24]. In the query model, the task is to evaluate a fixed function f on an unknown n-bit input x. In the classical setting, query algorithms are commonly referred to as *decision trees*. A decision tree accesses the input one bit at a time, choosing the bits to query in adaptive fashion. The objective is to determine $f(x)$ by querying as few bits as possible. The minimum number of queries needed to determine $f(x)$ in the worst case is called the *query complexity of* f. The quantum model is a far-reaching generalization of the classical decision tree whereby all bits can be queried in superposition with a single query. The catch is that the outcomes of those queries are then also in superposition, and it is not clear a priori whether quantum query algorithms are more powerful than decision trees. The focus of our paper is on the *bounded-error* regime, where the query algorithm (quantum or classical) is allowed to err with small constant probability on any given input.

The comparative power of randomized and quantum query algorithms has been studied for more than two decades. In pioneering work, Deutsch and Jozsa [13] gave a quantum query algorithm that solves, with a single query, a problem on n bits that any deterministic decision tree needs at least $n/2$ queries to solve. Unfortunately, this separation does not apply to the more subtle, bounded-error setting. This was addressed in follow-up work by Simon [25], who exhibited a problem with bounded-error quantum query complexity $O(\log^2 n)$ and randomized query complexity $\Omega(\sqrt{n})$. These are striking examples of the computational advantages afforded by the quantum model.

1.1 Forrelation and Rorrelation

The above results leave us with a fundamental question: what is the largest possible separation between bounded-error quantum and randomized query complexity, for a problem with n-bit input? This question was raised by Buhrman et al. [9] and, a decade later, by Aaronson and Ambainis [1], who presented it as being essential to understanding the phenomenon of quantum speedups. Toward this goal, the authors of [1] obtained both positive and negative results. They showed, for every constant t, that every quantum algorithm with t queries can be converted to a randomized decision tree of cost $O(n^{1-1/2t})$. In particular, this rules out an $O(1)$ versus $\Omega(n)$ separation. In the opposite direction, Aaronson and Ambainis

exhibited a problem that can be solved to bounded error with a single quantum query but has randomized query complexity $\tilde{\Omega}(\sqrt{n})$. They left open the challenge of obtaining a separation of $O(1)$ versus $\Omega(n^{\alpha})$ for some $\alpha > 1/2$.

In more detail, Aaronson and Ambainis [1] introduced and studied the *k-fold forrelation problem*. The input to the problem is a k-tuple of vectors $x_1, x_2, \ldots, x_k \in \{-1,1\}^n$, where n is a power of 2. Define

$$\phi_{n,k}(x_1, x_2, \ldots, x_k) = \frac{1}{n}\mathbf{1}^{\mathsf{T}} D_{x_1} H D_{x_2} H D_{x_3} H \cdots H D_{x_k} \mathbf{1}, \quad (1)$$

where $\mathbf{1}$ is the all-ones vector, H the Hadamard transform matrix of order n, and D_{x_i} the diagonal matrix with the vector x_i on the diagonal. Since each of the linear transformations $H, D_{x_1}, D_{x_2}, \ldots, D_{x_n}$ preserves Euclidean length, we have $|\phi_{n,k}(x_1, x_2, \ldots, x_k)| \leqslant 1$. Given $x_1, x_2, \ldots, x_k$, the forrelation problem is to tell apart the cases $|\phi_{n,k}(x_1, x_2, \ldots, x_k)| \leqslant \alpha$ and $\phi_{n,k}(x_1, x_2, \ldots, x_k) \geqslant \beta$, where the problem parameters $0 < \alpha < \beta < 1$ are suitably chosen constants. Equation (1) directly gives a quantum algorithm that solves the forrelation problem with bounded error and query cost k, where the k queries correspond to the k diagonal matrices. The cost can be further reduced to $\lceil k/2 \rceil$ by viewing (1) as the *inner product* of two vectors obtained by $\lceil k/2 \rceil$ and $\lfloor k/2 \rfloor$ applications, respectively, of diagonal matrices [1]. Aaronson and Ambainis complemented this with an $\tilde{\Omega}(\sqrt{n})$ lower bound on the randomized query complexity of the forrelation problem for $k = 2$, hence the 1 versus $\tilde{\Omega}(\sqrt{n})$ separation mentioned above.

Building on the work of Aaronson and Ambainis [1], last year Tal [28] gave an improved separation of $O(1)$ versus $\Omega(n^{2/3-\varepsilon})$ for bounded-error quantum and randomized query complexities, for any constant $\varepsilon > 0$. For this, Tal replaced (1) with the more general quantity

$$\phi_{n,k,U}(x_1, x_2, \ldots, x_k) = \frac{1}{n}\mathbf{1}^{\mathsf{T}} D_{x_1} U D_{x_2} U D_{x_3} U \cdots U D_{x_k} \mathbf{1}, \quad (2)$$

where U is an arbitrary but fixed orthogonal matrix. On input $x_1, x_2, \ldots, x_k \in \{-1,1\}^n$, the author of [28] considered the problem of distinguishing between the cases $|\phi_{n,k,U}(x_1, x_2, \ldots, x_k)| \leqslant 2^{-k-1}$ and $\phi_{n,k,U}(x_1, x_2, \ldots, x_k) \geqslant 2^{-k}$. This problem is referred to in [28] as the *k-fold rorrelation problem with respect to* U. The quantum algorithm of Aaronson and Ambainis, adapted to the arbitrary choice of U, solves this new problem with $\lceil k/2 \rceil$ queries and advantage $\Omega(2^{-k})$ over random guessing, which counts as a bounded-error algorithm for any constant k. On the other hand, Tal [28] proved that the randomized query complexity of k-fold rorrelation for uniformly random U is $\Omega(n^{2(k-1)/(3k-1)}/k \log n)$ with high probability. While this is weaker than Aaronson and Ambainis's bound for $k = 2$, setting k to a large constant gives a separation of $O(1)$ versus $\Omega(n^{2/3-\varepsilon})$ for bounded-error quantum and randomized query complexity for any constant $\varepsilon > 0$.

1.2 Our Results

Prior to our paper, Tal's separation of $O(1)$ versus $\Omega(n^{2/3-\varepsilon})$ was the strongest known, and Aaronson and Ambainis's challenge of obtaining an $O(1)$ versus $\Omega(n^{1-\varepsilon})$ separation remained open. The main contribution of our work is to resolve this question.

Separations for Partial Functions. In what follows, we let $f_{n,k,U}$ denote the k-fold rorrelation problem with respect to U. We prove:

Theorem 1.1. *Let n and k be positive integers, with $k \leqslant \frac{1}{3}\log n - 1$. Let $U \in \mathbb{R}^{n \times n}$ be a uniformly random orthogonal matrix. Then with probability $1 - o(1)$,*

$$R_{\frac{1}{2}-\gamma}(f_{n,k,U}) = \Omega\left(\frac{\gamma^2}{k} \cdot \frac{n^{1-\frac{1}{k}}}{(\log n)^{2-\frac{1}{k}}}\right) \quad (3)$$

for all $0 \leqslant \gamma \leqslant 1/2$.

For $k = 2$, this lower bound is the same as Aaronson and Ambainis's lower bound for the forrelation problem (which is $f_{n,2,H}$ in our notation). For $k = 3$ already, Theorem 1.1 is a polynomial improvement on all previous work, including Tal's recent result [28]. Theorem 1.1 is essentially tight for all k, both even and odd, due to the matching upper bound $O_k(n^{1-1/k})$ of Aaronson and Ambainis [1] for bounded block-multilinear polynomials of degree k. Since $f_{n,k,U}$ has an efficient quantum protocol for every U (see Section 5.2 for details), we obtain the following corollary:

Corollary 1.2. *Let $\varepsilon > 0$ be given. Then there is a partial Boolean function f on $\{-1,1\}^n$ with*

$$Q_{1/3}(f) = O(1),$$
$$R_{1/3}(f) = \Omega(n^{1-\varepsilon}).$$

This separation of bounded-error quantum and randomized query complexities is best possible for all f due to Aaronson and Ambainis's aforementioned result that every quantum protocol with k queries can be simulated by a randomized query algorithm of cost $O(n^{1-1/2k})$. In particular, Corollary 1.2 shows that the rorrelation problem separates quantum and randomized query complexity optimally, of all problems f. The following incomparable corollary can be obtained by taking $k = k(n)$ in Theorem 1.1 to be an arbitrarily slow-growing function, e.g., $k = \log\log\log n$:

Corollary 1.3. *Let $\alpha : \mathbb{N} \to \mathbb{N}$ be any monotone function with $\alpha(n) \to \infty$ as $n \to \infty$. Then there is a partial Boolean function f on $\{-1,1\}^n$ with $Q_{1/3}(f) \leqslant \alpha(n)$ and $R_{1/3}(f) \geqslant n^{1-o(1)}$.*

Again, this quantum-classical separation is best possible since [1] rules out the possibility of an $O(1)$ versus $n^{1-o(1)}$ gap.

A satisfying probability-theoretic interpretation of our results is that the phenomenon of quantum-classical gaps is a common one. More precisely, our results show that the set of orthogonal matrices U for which $f_{n,k,U}$ does *not* exhibit a best-possible quantum-classical separation has Haar measure 0. Prior to our work, this was unknown for any integer $k > 2$.

Separation for Total Functions. Our results so far pertain to *partial* Boolean functions, whose domain of definition is a proper subset of the Boolean hypercube. For total Boolean functions, such large quantum-classical gaps are not possible. In a seminal paper, Beals et al. [5] prove that the bounded-error quantum query complexity of a total function f is always polynomially related to the randomized query complexity of f. A natural question to ask is how large this polynomial gap can be. Grover's search [16] shows that the n-bit OR function has bounded-error quantum query complexity $\Theta(\sqrt{n})$ and randomized complexity $\Theta(n)$. For a long time, this quadratic

separation was believed to be the largest possible. In a surprising result, Aaronson et al. [2] proved the existence of a total function f with $R_{1/3}(f) = \tilde{\Omega}(Q_{1/3}(f)^{2.5})$. This was improved by Tal [28] to $R_{1/3}(f) \geqslant Q_{1/3}(f)^{8/3-o(1)}$. We give a polynomially stronger separation:

Theorem 1.4. *There is a function $f\colon \{-1,1\}^n \to \{0,1\}$ with*

$$R_{1/3}(f) \geqslant Q_{1/3}(f)^{3-o(1)}.$$

Theorem 1.4 follows automatically by combining our Corollary 1.3 with the "cheatsheet" framework of Aaronson et al. [2]. Specifically, they prove that any partial function f on n bits that exhibits an $n^{o(1)}$ versus $n^{1-o(1)}$ separation for bounded-error quantum versus randomized query complexity, can be automatically converted into a total function with $R_{1/3}(f) \geqslant Q_{1/3}(f)^{3-o(1)}$. A recent paper of Aaronson et al. [3] conjectures that $R_{1/3}(f) = O(Q_{1/3}(f)^3)$ for every total function f, which would mean that our separation in Theorem 1.4 is essentially optimal. The best current upper bound is $R_{1/3}(f) = O(Q_{1/3}(f)^4)$ due to [3], derived there from the breakthrough result of Huang [17] on the sensitivity conjecture.

Separations for Communication Complexity. Using standard reductions, our quantum-classical query separations imply analogous separations for communication complexity. In more detail, let f be a (possibly partial) Boolean function on $\{-1,1\}^n$. For any communication problem $g\colon \{-1,1\}^m \times \{-1,1\}^m \to \{-1,1\}$, we let $f \circ g$ denote the (possibly partial) communication problem on $(\{-1,1\}^m)^n \times (\{-1,1\}^m)^n$ given by $(f \circ g)(x,y) = f(g(x_1,y_1),\dots,g(x_n,y_n))$. Buhrman, Cleve, and Wigderson [7] proved that any quantum query algorithm for f gives a quantum communication protocol for $f \circ g$ with the same error and approximately the same cost. Quantitatively,

$$Q^{\mathrm{cc}}_{\varepsilon}(f \circ g) \leqslant Q_{\varepsilon}(f) \cdot O(m + \log n), \tag{4}$$

where $Q^{\mathrm{cc}}_{\varepsilon}$ denotes ε-error quantum communication complexity. Reversing this inequality has seen a great deal of work, mainly in the classical setting. A well-studied function g in this line of research is the inner product function $\mathrm{IP}_m\colon \{-1,1\}^m \times \{-1,1\}^m \to \{-1,1\}$, given by $\mathrm{IP}_m(u,v) = \bigoplus_{i=1}^m (u_i \wedge v_i)$. In particular, Chattopadhyay, Filmus, Koroth, Meir, and Pitassi [10, Theorem 1] prove that

$$R^{\mathrm{cc}}_{1/3}(f \circ \mathrm{IP}_{c\log n}) = \Omega(R_{1/3}(f) \log n) \tag{5}$$

for every (possibly partial) function f on $\{-1,1\}^n$, where $R^{\mathrm{cc}}_{\varepsilon}$ denotes ε-error randomized communication complexity and $c > 1$ is an absolute constant. In light of this connection between query complexity and communication complexity, our main results have the following consequences.

Theorem 1.5. *Let $\varepsilon > 0$ be given. Then there is a partial Boolean function F on $\{-1,1\}^N \times \{-1,1\}^N$ with*

$$Q^{\mathrm{cc}}_{1/3}(F) = O(\log N),$$
$$R^{\mathrm{cc}}_{1/3}(F) = \Omega(N^{1-\varepsilon}).$$

Proof. Take f as in Corollary 1.2 and define $N = cn\log n$ and $F = f \circ \mathrm{IP}_{c\log n}$. Then the communication bounds follow from (4) and (5), respectively. □

Theorem 1.5 is essentially optimal and a polynomial improvement on previous work. The best previous quantum-classical separation for communication complexity was $O(\log N)$ versus $\Omega(N^{2/3-\varepsilon})$, implicit in Tal [28] and preceded in turn by other exponential separations [14, 21, 22]. Similarly, our Corollary 1.3 translates in a black-box manner to communication complexity:

Theorem 1.6. *Let $\alpha\colon \mathbb{N} \to \mathbb{N}$ be any monotone function with $\alpha = \omega(1)$. Then there is a partial Boolean function F on $\{-1,1\}^N \times \{-1,1\}^N$ with*

$$Q^{\mathrm{cc}}_{1/3}(F) \leqslant \alpha(N)\log N,$$
$$R^{\mathrm{cc}}_{1/3}(F) \geqslant N^{1-o(1)}.$$

Proof. Take f as in Corollary 1.3 and define $N = cn\log n$ and $F = f \circ \mathrm{IP}_{c\log n}$. Then the communication bounds follow from (4) and (5), respectively. □

Finally, we obtain the following result for *total* functions.

Theorem 1.7. *There is a function $F\colon \{-1,1\}^N \times \{-1,1\}^N \to \{0,1\}$ with*

$$R^{\mathrm{cc}}_{1/3}(F) \geqslant Q^{\mathrm{cc}}_{1/3}(F)^{3-o(1)}.$$

Proof. The cheatsheet framework [2] ensures that the quantum and classical query complexities of f in Theorem 1.4 are polynomial in the number of variables n. With this in mind, we proceed as before, setting $N = cn\log n$ and $F = f \circ \mathrm{IP}_{c\log n}$ and applying (4) and (5). □

Again, Theorem 1.7 is a polynomial improvement on previous work, the best previous result being a power of $8/3$ separation implicit in [28].

Fourier Weight of Decision Trees. It is straightforward to verify that a uniformly random input $x \in (\{-1,1\}^n)^k$ is with high probability a *negative* instance of the rorrelation problem $f_{n,k,U}$. With this in mind, Tal [28] proves his lower bound for rorrelation by constructing a probability distribution $\mathscr{D}_{n,k,U}$ that generates *positive* instances of $f_{n,k,U}$ with nontrivial probability yet is indistinguishable from the uniform distribution by a decision tree T of cost $n^{2/3-O(1/k)}$. His notion of indistinguishability is based on the Fourier spectrum. Specifically, Tal [28] shows that: (i) the *sum* of the absolute values of the Fourier coefficients of T of given order ℓ does not grow too fast with ℓ; and (ii) the *maximum* Fourier coefficient of $\mathscr{D}_{n,k,U}$ of order ℓ decays exponentially fast with ℓ. In Tal's paper, the bound for (ii) is essentially optimal, whereas the bound for (i) is far from tight. The sum of the absolute values of the order-ℓ Fourier coefficients of a decision tree T, which we refer to as the *ℓ-Fourier weight of T*, is shown in [28] to be at most

$$c^{\ell}\sqrt{d^{\ell}(1+\log kn)^{\ell-1}}, \tag{6}$$

where d is the depth of the tree and $c \geqslant 1$ is an absolute constant. This bound is strong for any constant ℓ but degrades rapidly as ℓ grows. In particular, for $\ell = \sqrt{d}$ already, (6) is weaker than the trivial bound $\binom{d}{\ell}$. This is a major obstacle since the indistinguishability proof requires strong bounds for every ℓ. This obstacle is the reason why Tal's analysis gives the randomized query lower

bound $n^{2/3-O(1/k)}$ as opposed to the optimal $\tilde{\Omega}(n^{1-1/k})$. Tal conjectured that the ℓ-Fourier weight of a depth-d decision tree is in fact bounded by $c^\ell \sqrt{\binom{d}{\ell}(1+\log kn)^{\ell-1}}$, which is a factor of $\sqrt{\ell!}$ improvement on (6) and essentially optimal. We prove his conjecture:

Theorem 1.8. *Let $T\colon \{-1,1\}^n \to \{0,1\}$ be a function computable by a decision tree of depth d. Then*

$$\sum_{\substack{S\subseteq\{1,2,\ldots,n\}:\\ |S|=\ell}} |\hat{T}(S)| \leqslant c^\ell \sqrt{\binom{d}{\ell}(1+\log n)^{\ell-1}}, \quad \ell = 1,2,\ldots,n,$$

where $c \geqslant 1$ is an absolute constant.

It is well known and easy to show that Theorem 1.8 is essentially tight, even for *nonadaptive* decision trees [19, Theorem 5.19]. The actual statement that we prove is more precise and takes into account the density parameter $\mathbf{P}[T(x) \neq 0]$; see Theorem 4.12 for details. With Theorem 1.8 in hand, all our main results (Theorem 1.1 and its corollaries) follow immediately by combining the new bound on the Fourier weight of decision trees with Tal's near-optimal bounds on the individual Fourier coefficients of $\mathscr{D}_{n,k,U}$.

Theorem 1.8 is of interest in its own right, independent of its use in this paper to obtain optimal quantum-classical separations. The study of the Fourier spectrum has a variety of applications in theoretical computer science, including circuit complexity, learning theory, pseudorandom generators, and quantum computing. Even prior to Tal's work, the ℓ-Fourier weight of decision trees was studied for $\ell = 1$ by O'Donnell and Servedio [20], who proved the tight $O(\sqrt{d})$ bound and used it to give a polynomial-time learning algorithm for monotone decision trees. Fourier weight has been studied for various other classes of Boolean functions, including bounded-depth circuits, branching programs, low-degree polynomials over finite fields, and functions with bounded sensitivity; see the recent papers [11, 12, 15, 26, 27] and the references therein.

1.3 Limitations of Previous Analyses

In this part, we overview Tal's bound on the ℓ-Fourier weight of decision trees. To build intuition, it is helpful to first examine the case $\ell = 1$, due to O'Donnell and Servedio [20] and Tal [28]. For simplicity, consider a perfect tree T of depth d with leaves labeled 0 and 1, where the i-th variable queried in each path is x_i. Throughout this discussion, we identify a decision tree with the function that it computes, and use the same variable for both. By negating the variables if necessary, we may assume that $\hat{T}(i) \geqslant 0$. In particular,

$$\sum_{i=1}^n |\hat{T}(i)| = \mathop{\mathbf{E}}_x \left[T(x) \sum_{i=1}^d x_i \right].$$

This gives a new perspective on $\sum |\hat{T}(i)|$ in terms of the random experiment whereby one picks a random root-to-leaf path, sums all the variables in that path, and multiplies the result by the label of the leaf. The expected value of this experiment equals $\sum |\hat{T}(i)|$. It is clear that this value is maximized when the leaves labeled 1 correspond to paths with large sums. With this observation [28], one can verify that

$$\sum_{i=1}^n |\hat{T}(i)| = O\left(p\sqrt{d \ln \frac{e}{p}}\right), \tag{7}$$

where $p = \mathbf{P}[T(x) \neq 0]$ is the fraction of nonzero leaves, which we refer to as the *density of T*. By linearity, the same argument applies even to adaptive trees.

Tal's analysis for $\ell \geqslant 2$ is a natural inductive generalization of the above argument. Let T be an arbitrary tree in variables $x_1, x_2, \ldots, x_n$. Let V_i denote the set of internal nodes in T labeled by the variable x_i. The key notion is that of the *contraction of T with respect to x_i*, which is a tree denoted by T_i with real-valued labels at the leaves. This tree T_i is formed by the following two-step process: (i) for each path that does not query x_i, set the leaf label to 0; and (ii) for each $v \in V_i$, replace the subtree T_v rooted at v by a single leaf labeled by the Fourier coefficient $\hat{T}_v(i)$. The n contractions of T give rise to the decomposition

$$\sum_{|S|=\ell} |\hat{T}(S)| \leqslant \sum_{i=1}^n \sum_{|S|=\ell-1} |\hat{T}_i(S)|, \tag{8}$$

which is the foundation of Tal's inductive argument. The real-valued labels of the T_i present no difficulty since one can replace each such label by its binary expansion and thus write T_i as a linear combination of trees with binary labels. The key parameter in Tal's inductive proof is density, and it needs to be maintained carefully for each of the trees involved. Since the contractions of T can overlap in complicated ways, it becomes increasingly difficult to accurately keep track of the densities. This translates into progressively larger losses at each step of the inductive argument. Cumulatively, the argument incurs an extraneous factor of $\sqrt{\ell!}$ in the final bound. Despite considerable efforts, we were not able to find a way forward within this framework.

1.4 Our Approach

To obtain the near-optimal bound in Theorem 1.8, we adopt a completely different approach. At a high level, we partition $\sum_{|S|=\ell} |\hat{T}(S)|$ into well-structured parts. We discuss the partitioning strategy first, and then our analysis of each part in the partition.

The Partition. Let T be a perfect tree of depth d. We think of the vertices at any given depth as forming a *layer*, and we number the layers of T consecutively 1 through d. Consider a grouping of the layers into ℓ disjoint blocks $I_1, I_2, \ldots, I_\ell \subseteq \{1,2,\ldots,d\}$, where each block consists of consecutive layers from T, and the union $I_1 \cup I_2 \cup \cdots \cup I_\ell$ may be a proper subset of $\{1,2,\ldots,d\}$. As a canonical example, we could partition the layers into ℓ blocks of roughly equal size. Viewed as a function, T is the sum of the characteristic functions of the root-to-leaf paths, each such path weighted by the corresponding leaf. If one alters this sum by keeping, for each path, only those Fourier coefficients that have exactly one variable in each block, the result is a real-valued function which we denote by $T|_{I_1 * I_2 * \cdots * I_\ell}$. Here we define $I_1 * I_2 * \cdots * I_\ell = \{S \in \binom{[d]}{\ell} : |S \cap I_i| = 1 \text{ for each } i\}$, and we refer to any such family of sets in $\binom{[d]}{\ell}$ as an *elementary family*. Our challenge is to find an efficient partition of $\binom{[d]}{\ell}$ into elementary families $\mathscr{E}_1, \mathscr{E}_2, \ldots, \mathscr{E}_N$. Then

$$T|_{\binom{[d]}{\ell}} = \sum_{i=1}^N T|_{\mathscr{E}_i}, \tag{9}$$

and we can bound the Fourier weight of the degree-ℓ homogeneous part of T by bounding that of $T|_{\mathscr{E}_i}$ for each i. For the proof of

Theorem 1.8, we need a partition that achieves

$$\sum_{i=1}^{N} \sqrt{|\mathscr{E}_i|} \leqslant C^{\ell} \sqrt{\binom{d}{\ell}} \tag{10}$$

for an absolute constant $C \geqslant 1$. Such a partition would be essentially extremal due to the trivial lower bound $\sum \sqrt{|\mathscr{E}_i|} \geqslant \binom{d}{\ell}^{1/2}$ for every partition of $\binom{[d]}{\ell}$. Unfortunately, with elementary families defined as above, such a partition does not exist! For the sake of simplicity, we ignore this complication altogether in the remainder of this discussion. In the actual proof, we resolve this issue by allowing elementary families to contain up to two variables per block. This makes the rest of the proof more delicate, but still suffices for the purposes of proving Theorem 1.8. We give a first-principles combinatorial construction of a partition with (10) in Section 3.

Analysis of Individual Parts. For any elementary family $\mathscr{E} = I_1 * I_2 * \cdots * I_\ell$, we prove that $T|_{\mathscr{E}}$ has Fourier weight

$$\sqrt{|\mathscr{E}| \cdot O(\log n)^{\ell-1}}. \tag{11}$$

Along with (9) and (10), this immediately implies Theorem 1.8. In this overview, we will focus on the special case

$$|I_1| = |I_2| = \cdots = |I_\ell| = \frac{d}{\ell}.$$

Our bound (11) uses a generalization of decision trees where the leaves can be labeled by polynomials. With this generalization, we can further define tree addition, as well as tree multiplication by polynomials. This provides a powerful framework for decomposing trees and expressing them as conical combinations of simpler trees. To see how this generalization comes into play, consider the subtree T_v rooted at some node v in the *first* layer of I_ℓ. By the structure of $T|_{\mathscr{E}}$, the only relevant aspect of T_v is its degree-1 homogeneous part. Therefore, T_v can be *replaced* with its degree-1 homogeneous part. Now, let T' be the decision tree obtained by contracting every node v in the first layer of I_ℓ into a leaf labeled by the polynomial $\sum_{i=1}^{n} \hat{T}_v(i) x_i$. We show that analyzing the Fourier weight of $T|_{I_1 * I_2 * \cdots * I_\ell}$ is equivalent to analyzing that of T' with respect to the smaller elementary family $I_1 * I_2 * \cdots * I_{\ell-1}$. The latter is a delicate task, and our solution involves three stages.

(i) In the first stage, we group leaves v in T' according to the density α_v of the original subtree T_v. Observe that

$$\sum_{i=1}^{n} |\hat{T}_v(i)| \leqslant c' \alpha_v \sqrt{\frac{d}{\ell} \ln \frac{e}{\alpha_v}}$$

for some constant $c' \geqslant 1$. We decompose $T' = \sum_{j=0}^{\infty} T'_j$, where T'_j keeps a leaf v if $\alpha_v \in (3^{-j-1}, 3^{-j}]$ and replaces it with 0 otherwise.

(ii) In the second stage, we further decompose T'_j as follows. Let β_j be the fraction of nonzero leaves in T'_j, and let m be the maximum Fourier weight of a nonzero leaf v of T'_j. We then express T'_j as the conical combination $T'_j = \sum_{r=1}^{\infty} c_r T'_{j,r}$ such that: $\sum c_r = m$; each nonzero leaf of $T'_{j,r}$ is labeled with some variable or its negation; and the fraction of nonzero leaves in each $T'_{j,r}$ is β_j.

(iii) In the final stage, we decompose $T'_{j,r}$ into n different trees according to the n variables: $T'_{j,r} = \sum_{i=1}^{n} T'_{j,r,i} \cdot x_i$. The tree $T'_{j,r,i}$ keeps only those leaves v that are labeled by $\pm x_i$, and the new label is exactly the sign of the variable x_i. Now $T'_{j,r,i} : \{-1,1\}^n \to \{-1,0,1\}$ has density β_j/n on average, and $T'_{j,r,i}|_{I_1 * I_2 * \cdots * I_{\ell-1}}$ can be analyzed using the inductive hypothesis.

Of the three stages, the first stage is the least natural but crucial. To see this, let $\ell = 2$ and consider the following extreme case: for all nonzero leaves v in T', the densities α_v are equal, $\alpha_v = \alpha$. Let p denote the density of T. Then there is some j such that $T' = T'_j$, and T'_j has density p/α. Consequently, $T'_{j,r,i}$ has density $p/(n\alpha)$ on average. The 1-Fourier weight of $T'_{j,r,i}$ for average i can be bounded by

$$c' \cdot \frac{p}{n\alpha} \sqrt{\frac{d}{2} \ln \frac{en\alpha}{p}}.$$

The Fourier weight of $T'|_{\{1,2,\ldots,d/2\} * \{d/2+1, d/2+2, \ldots, d\}}$ can then be bounded by

$$\begin{aligned} c' \cdot \alpha \sqrt{\frac{d}{2} \ln \frac{e}{\alpha}} \cdot \sum_{i=1}^{n} c' \cdot \frac{p}{n\alpha} \sqrt{\frac{d}{2} \ln \frac{en\alpha}{p}} \\ = (c')^2 \cdot p \sqrt{\left(\frac{d}{2}\right)^2 \ln \frac{e}{\alpha} \cdot \ln \frac{en\alpha}{p}}. \end{aligned} \tag{12}$$

The corresponding bound for $\ell = 2$ that Tal obtains is

$$O\left(p \sqrt{d^2 \ln \frac{e}{p} \cdot \ln \frac{en}{p}}\right).$$

Comparing it with our bound (12) shows that for $\alpha \gg p$, our factor $\ln \frac{e}{\alpha}$ is substantially smaller than Tal's corresponding factor $\ln \frac{e}{p}$; while for α close to p, our factor $\ln \frac{en\alpha}{p}$ is substantially smaller than Tal's $\ln \frac{en}{p}$. This is the intuitive reason why the first stage allows us to avoid the $\sqrt{\ell!}$ loss. Its surprising power comes from the framework of elementary families set up at the beginning of the proof.

1.5 Independent Work by Bansal and Sinha

Independently and concurrently with our work, Bansal and Sinha [4] also obtained an optimal, $\lceil k/2 \rceil$ versus $\tilde{\Omega}(n^{1-1/k})$ separation of quantum and randomized query complexity. Their result uses completely different techniques and is incomparable with ours. In more detail, Bansal and Sinha [4] construct a function f with randomized query complexity

$$R_{\frac{1}{2}-\gamma}(f) = \Omega\left(\frac{\gamma^2}{k^{29}} \cdot \left(\frac{n}{\log(k+n)}\right)^{1-\frac{1}{k}}\right), \quad \forall \gamma \in [0, 1/2]. \tag{13}$$

This is essentially the same as our lower bound on randomized query complexity (Theorem 1.1):

$$R_{\frac{1}{2}-\gamma}(f_{n,k,U}) = \Omega\left(\frac{\gamma^2}{k} \cdot \frac{n^{1-\frac{1}{k}}}{(\log n)^{2-\frac{1}{k}}}\right), \qquad \forall \gamma \in [0, 1/2].$$

In both cases, the function in question has a quantum query algorithm with cost $\lceil k/2 \rceil$ and error $\frac{1}{2} - 2^{-\Theta(k)}$. In particular, for an arbitrary constant $k \geqslant 1$, the bounded-error quantum query complexity

is at most $\lceil k/2 \rceil$. (The original version of [4], released concurrently with our paper, had a poorer error parameter: $\frac{1}{2} - (\log n)^{-\Theta(k)}$. But the authors of [4] were able to improve it several weeks later to match our error parameter, $\frac{1}{2} - 2^{-\Theta(k)}$.)

The two approaches have incomparable strengths. To start with, Bansal and Sinha [4] prove their lower bound for an *explicit* function f (namely, the forrelation and rorrelation problems with a properly chosen gap parameter), as opposed to the uniformly random choice of $f_{n,k,U}$ in this paper.

On the other hand, our analysis has the advantage of determining the ℓ-Fourier weight of decision trees. This result is of independent interest beyond quantum computing, given the numerous recent applications of Fourier weight to learning theory and pseudorandom generators. We believe that our techniques may be relevant to other unresolved questions on the Fourier spectrum of Boolean functions. The work in [4], by contrast, does not imply any improved bounds on Fourier weight.

Another strength of our analysis is methodological. The proof in [4] uses advanced analytic machinery, whereas our approach is elementary and self-contained. Indeed, the only analytic fact used in this paper and Tal [28] is the p.d.f. of the multivariate normal distribution. With this simple toolkit, we obtain all the same optimal quantum-classical separations for query complexity and communication complexity as in [4].

2 PRELIMINARIES

2.1 General Notation

There are two common arithmetic encodings for the Boolean values: the traditional encoding false $\leftrightarrow 0$, true $\leftrightarrow 1$, and the Fourier-motivated encoding false $\leftrightarrow 1$, true $\leftrightarrow -1$. Throughout this manuscript, we use the former encoding for the range of a Boolean function and the latter for the domain. With this convention, Boolean functions are mappings $\{-1,1\}^n \to \{0,1\}$ for some n.

We denote the empty string as usual by ε. For an alphabet Σ and a natural number n, we let $\Sigma^{\leqslant n}$ denote the set of all strings over Σ of length up to n, so that $\Sigma^{\leqslant n} = \{\varepsilon\} \cup \Sigma \cup \Sigma^2 \cup \cdots \cup \Sigma^n$. For a string v over a given alphabet, we let $|v|$ denote the length of v. For a set S, we let $v|_S$ denote the substring of v indexed by the elements of S. In other words, $v|_S = v_{i_1} v_{i_2} \cdots v_{i_{|S|}}$ where $i_1 < i_2 < \cdots < i_{|S|}$ are the elements of S. In the same spirit, we define $v_{\leqslant i} = v_1 v_2 \ldots v_i$.

The power set of a set S is denoted by $\mathscr{P}(S)$. For a set S and a nonnegative integer k, we let $\binom{S}{k}$ denote the family of subsets of S that have cardinality exactly k:

$$\binom{S}{k} = \{S' \subseteq S : |S'| = k\}.$$

We further define

$$\mathscr{P}_{n,k} = \binom{\{1,2,\ldots,n\}}{k} = \{S \subseteq \{1,2,\ldots,n\} : |S| = k\}.$$

The following well-known bound [18, Proposition 1.4] is used in our proofs without further mention:

$$\left(\frac{n}{k}\right)^k \leqslant \binom{n}{k} \leqslant \left(\frac{en}{k}\right)^k, \qquad k = 1,2,\ldots,n, \tag{14}$$

where $e = 2.7182\ldots$ denotes Euler's number.

We adopt the standard notation $\mathbb{N} = \{0,1,2,3,\ldots\}$ and $\mathbb{Z}^+ = \{1,2,3,\ldots\}$ for the sets of natural numbers and positive integers, respectively. We adopt the extended real number system $\mathbb{R} \cup \{-\infty, \infty\}$ in all calculations. The functions $\ln x$ and $\log x$ stand for the natural logarithm of x and the logarithm of x to base 2, respectively. To avoid excessive use of parentheses, we follow the notational convention that $\ln a_1 a_2 \ldots a_k = \ln(a_1 a_2 \ldots a_k)$ for any factors $a_1, a_2, \ldots, a_k$. The binary entropy function $H\colon [0,1] \to [0,1]$ is given by

$$H(x) = x \log \frac{1}{x} + (1-x) \log \frac{1}{1-x}.$$

Basic calculus reveals that

$$H(x) \leqslant 1 - \frac{2}{\ln 2}\left(x - \frac{1}{2}\right)^2. \tag{15}$$

For nonempty sets $A, B \subseteq \mathbb{R}$, we write $A < B$ to mean that $a < b$ for all $a \in A$, $b \in B$. It is clear that this relation is a partial order on nonempty subsets of $\mathbb{R}$. We use the standard definition of the sign function:

$$\operatorname{sgn} x = \begin{cases} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } x > 0. \end{cases}$$

For a finite set X, we let $\mathbb{R}^X$ denote the family of real-valued functions on X. For $f, g \in \mathbb{R}^X$, we let $f \cdot g \in \mathbb{R}^X$ denote the pointwise product of f and g, with $(f \cdot g)(x) = f(x)g(x)$. We use the standard inner product $\langle f, g \rangle = \sum_{x \in X} f(x)g(x)$.

2.2 Fourier Transform

Consider the real vector space of functions $\{-1,1\}^n \to \mathbb{R}$. For $S \subseteq \{1,2,\ldots,n\}$, define $\chi_S\colon \{-1,1\}^n \to \{-1,1\}$ by $\chi_S(x) = \prod_{i \in S} x_i$. Then

$$\langle \chi_S, \chi_T \rangle = \begin{cases} 2^n & \text{if } S = T, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, $\{\chi_S\}_{S \subseteq \{1,2,\ldots,n\}}$ is an orthogonal basis for the vector space in question. In particular, every function $\phi\colon \{-1,1\}^n \to \mathbb{R}$ has a unique representation of the form

$$\phi = \sum_{S \subseteq \{1,2,\ldots,n\}} \hat{\phi}(S) \chi_S$$

for some reals $\hat{\phi}(S)$, where by orthogonality $\hat{\phi}(S) = 2^{-n} \langle \phi, \chi_S \rangle$. The reals $\hat{\phi}(S)$ are called the *Fourier coefficients of* ϕ, and the mapping $\phi \mapsto \hat{\phi}$ is the *Fourier transform of* ϕ. Put another way, every function $\phi : \{-1,1\}^n \to \mathbb{R}$ has a unique representation as a multilinear polynomial

$$\phi(x) = \sum_{S \subseteq \{1,2,\ldots,n\}} \hat{\phi}(S) \prod_{i \in S} x_i, \tag{16}$$

where the real numbers $\hat{\phi}(S)$ are the Fourier coefficients of f. The *order* of a Fourier coefficient $\hat{\phi}(S)$ is the cardinality $|S|$.

For $k = 0,1,2,\ldots,n$, we introduce the operator $L_k\colon \mathbb{R}^{\{-1,1\}^n} \to \mathbb{R}^{\{-1,1\}^n}$ that linearly sends a function $\phi\colon \{-1,1\}^n \to \mathbb{R}$ to the function $L_k\phi\colon \{-1,1\}^n \to \mathbb{R}$ given by

$$(L_k\phi)(x) = \sum_{S \in \mathscr{P}_{n,k}} \hat{\phi}(S) \chi_S(x).$$

We refer to $L_k\phi$ as the *degree-k homogeneous part of* ϕ.

For any polynomial $p \in \mathbb{R}[x_1, x_2, \ldots, x_n]$, we let $\|\!|p|\!\|$ denote the sum of the absolute values of the coefficients of p. One easily verifies the well-known fact that $\|\!|\cdot|\!\|$ is a norm on the polynomial ring $\mathbb{R}[x_1, x_2, \ldots, x_n]$. We identify a function $\phi\colon \{-1,1\}^n \to \mathbb{R}$ with its unique representation (16) as a multilinear polynomial, to the effect that

$$\|\!|\phi|\!\| = \sum_{S \subseteq \{1,2,\ldots,n\}} |\hat{\phi}(S)|$$

is the sum of the absolute values of the Fourier coefficients of ϕ.

Proposition 2.1. *For any functions $\phi, \psi\colon \{-1,1\}^n \to \mathbb{R}$ and reals a, b, one has*

$$\|\!|a\phi + b\psi|\!\| \leqslant |a|\,\|\!|\phi|\!\| + |b|\,\|\!|\psi|\!\|.$$

A proof of this proposition is available in the full version of our paper [23]. We will frequently use the norm $\|\!|\cdot|\!\|$ in conjunction with the operator L_k to refer to the sum of the absolute values of the Fourier coefficients of given order k:

$$\|\!|L_k\,\phi|\!\| = \sum_{S \in \mathscr{P}_{n,k}} |\hat{\phi}(S)|.$$

2.3 Generalized Decision Trees

Throughout this manuscript, we assume decision trees to be perfect binary trees, with each internal node having two children and all leaves having the same depth. This convention is without loss of generality since a decision tree computing a given function f can be made into a perfect binary tree for f of the same depth, by querying dummy variables as necessary. We denote the variables of a decision tree by $x_1, x_2, \ldots, x_n \in \{-1,1\}$, and identify the vertices of a decision tree in the natural manner with strings in $\{-1,1\}^*$. Thus, ε denotes the root of the tree, and a string $v \in \{-1,1\}^k$ denotes the vertex at depth k reached from the root by following the path $v_1 v_2 \ldots v_k$. Formally, a *decision tree* of depth d in Boolean variables $x_1, x_2, \ldots, x_n \in \{-1,1\}$ is a function T on $\{-1,1\}^{\leqslant d}$ with the following two properties.

(i) One has $T(v) \in \{1, 2, \ldots, n\}$ for every $v \in \{-1,1\}^{\leqslant d-1}$, with the interpretation that $T(v)$ is the index of the variable queried at the internal node found by following the path $v = v_1 v_2 v_3 \ldots$ from the root of the decision tree. We note that a variable cannot be queried twice on the same path, and therefore the d numbers $T(\varepsilon), T(v_1), T(v_1 v_2), \ldots, T(v_1 v_2 \ldots v_{d-1})$ are pairwise distinct for every $v \in \{-1,1\}^{d-1}$.

(ii) One has $T(v) \in \mathbb{R}[x_1, x_2, \ldots, x_n]$ for every $v \in \{-1,1\}^d$, with the interpretation that $T(v)$ is the label of the leaf reached by following the path $v = v_1 v_2 \ldots v_d$ from the root of the tree. Thus, every leaf is labeled with a real-valued polynomial in the input variables $x_1, x_2, \ldots, x_n$. At a given leaf $v \in \{-1,1\}^d$, the variables $x_{T(\varepsilon)}, x_{T(v_1)}, \ldots, x_{T(v_1 v_2 \ldots v_{d-1})}$ have been queried and therefore have fixed values. For this reason, we require $T(v)$ to be a real polynomial in variables other than $x_{T(\varepsilon)}, x_{T(v_1)}, \ldots, x_{T(v_1 v_2 \ldots v_{d-1})}$. We refer to a leaf $v \in \{-1,1\}^d$ as a *nonzero leaf* if $T(v)$ is not the zero polynomial. While we formally allow arbitrary real polynomials, the identity $x_i^2 = x_i$ effectively forces $T(v)$ for each $v \in \{-1,1\}^d$ to be multilinear.

Our formalism generalizes the traditional notion of a decision tree, where the leaf labels are restricted to the Boolean constants 0, 1.

Proposition 2.2. *Let T be a given decision tree of depth d. Then the function $f\colon \{-1,1\}^n \to \mathbb{R}$ computed by T is given by*

$$f(x) = \sum_{v \in \{-1,1\}^d} T(v) \cdot \prod_{i=1}^{d} \frac{1 + v_i x_{T(v_1 v_2 \ldots v_{i-1})}}{2}. \tag{17}$$

We emphasize that $T(v)$ here is a polynomial in $x_1, x_2, \ldots, x_n$ and not necessarily a constant value. In fact, the norm $\|\!|T(v)|\!\|$ for leaves v is a prominent quantity in this paper.

Proof. For an input $x \in \{-1,1\}^n$ and a leaf $v \in \{-1,1\}^d$, the product

$$\prod_{i=1}^{d} \frac{1 + v_i x_{T(v_1 v_2 \ldots v_{i-1})}}{2}$$

evaluates to 1 if the input x reaches the leaf v in T, and evaluates to 0 otherwise. Recall that any given input x reaches precisely one leaf v, and the output of the tree on x is defined to be the corresponding polynomial $T(v) \in \mathbb{R}[x_1, x_2, \ldots, x_n]$ evaluated at x. Thus, (17) evaluates to $T(v)$ where v is the leaf reached by x. □

For a decision tree T of depth d, we let $\mathrm{dns}(T)$ denote the fraction of leaves in T with nonzero labels:

$$\mathrm{dns}(T) = \mathop{\mathbf{P}}_{v \in \{-1,1\}^d}[T(v) \neq 0].$$

We refer to this quantity as the *density* of T. Another important complexity measure is the *degree of T*, denoted $\deg(T)$ and defined as the maximum of the degrees of the polynomials $T(v) \in \mathbb{R}[x_1, x_2, \ldots, x_n]$ for $v \in \{-1,1\}^d$. Recall that the zero polynomial 0 is considered to have degree $-\infty$. For an internal node $v \in \{-1,1\}^{\leqslant d-1}$, we let T_v denote the subtree of T rooted at v. Thus, T_v is the tree of depth $d - |v|$ given by $T_v(u) = T(vu)$ for all $u \in \{-1,1\}^{\leqslant d-|v|}$. The following fact is straightforward and well-known.

Fact 2.3. *Let T be a given decision tree of degree at most 0. Let $f\colon \{-1,1\}^n \to \mathbb{R}$ be the function computed by T. Then*

$$\mathop{\mathbf{P}}_{x \in \{-1,1\}^n}[f(x) \neq 0] = \mathrm{dns}(T).$$

Proof. Let d be the depth of T. Since T is a perfect binary tree, the fraction of inputs $x \in \{-1,1\}^n$ that reach any given leaf of T is exactly 2^{-d}. Therefore, the probability that a random input $x \in \{-1,1\}^n$ reaches a leaf with a nonzero label is precisely the fraction of leaves with nonzero labels, which is by definition $\mathrm{dns}(T)$. □

We will be working with special classes of trees described by several parameters. Specifically, we let $\mathscr{T}(n, d, p, k)$ denote the set of all trees in n Boolean variables $x_1, x_2, \ldots, x_n \in \{-1,1\}$ of depth d and density p such that for every leaf $v \in \{-1,1\}^d$, the label $T(v)$ is either the zero polynomial 0 or a homogeneous multilinear polynomial of degree k. We further define $\mathscr{T}^*(n, d, p, k)$ to be the set of all trees $T \in \mathscr{T}(n, d, p, k)$ that have the additional property that $T(v) \in \{0\} \cup \{\pm \prod_{i \in S} x_i : S \in \mathscr{P}_{n,k}\}$ for every leaf $v \in \{-1,1\}^d$. Thus, every nonzero leaf in a tree $T \in \mathscr{T}^*(n, d, p, k)$ is labeled with a signed monomial of degree k.

The Fourier spectrum of decision trees has been studied in several works, as discussed in the introduction. We will need the following special case of a result due to Tal [28, Theorem 7.5].

Theorem 2.4 (Tal). *Let $f\colon \{-1,1\}^n \to \{-1,0,1\}$ be given, $f \not\equiv 0$. Define $p = \mathbf{P}_{x\in\{-1,1\}^n}[f(x) \neq 0]$. Suppose that f can be computed by a depth-d decision tree. Then*

$$\|\!|L_1 f|\!\| \leqslant \binom{d}{1}^{1/2} Cp\sqrt{\ln\frac{e}{p}},$$
$$\|\!|L_2 f|\!\| \leqslant \binom{d}{2}^{1/2} C^2 p\sqrt{\ln\frac{e}{p}}\sqrt{\ln\frac{en}{p}},$$

where $C \geqslant 1$ is an absolute constant.

Tal states his result for functions $f\colon \{-1,1\}^n \to \{0,1\}$ rather than $f\colon \{-1,1\}^n \to \{-1,0,1\}$. But Theorem 2.4 follows immediately by writing $f = f^+ - f^-$, where $f^+, f^-\colon \{-1,1\}^n \to \{0,1\}$ are the positive and negative parts of f, and applying Tal's result separately to f^+ and f^-.

3 ELEMENTARY SET FAMILIES

As explained in the introduction, we obtain our Fourier weight bound by combining the Fourier coefficients of a decision tree into well-structured groups and bounding the sum of the absolute values in each group. In this section, we lay the combinatorial groundwork for this result by proving that $\mathscr{P}_{n,k}$ can be efficiently partitioned into what we call "elementary families."

For set families $\mathscr{A}, \mathscr{B} \subseteq \mathscr{P}(\mathbb{Z})$, we define $\mathscr{A} * \mathscr{B} = \{A \cup B : A \in \mathscr{A}, B \in \mathscr{B}\}$. We collect basic properties of this operation in the proposition below.

Proposition 3.1. *Let $\mathscr{A}, \mathscr{B}, \mathscr{C} \subseteq \mathscr{P}(\mathbb{Z})$ be given. Then:*

(i) $\mathscr{A} * \varnothing = \varnothing * \mathscr{A} = \varnothing$;
(ii) $\mathscr{A} * \{\varnothing\} = \{\varnothing\} * \mathscr{A} = \mathscr{A}$;
(iii) $(\mathscr{A} * \mathscr{B}) * \mathscr{C} = \mathscr{A} * (\mathscr{B} * \mathscr{C})$;
(iv) $\mathscr{A} * \mathscr{B} = \mathscr{B} * \mathscr{A}$;
(v) $(\mathscr{A} \cup \mathscr{B}) * \mathscr{C} = (\mathscr{A} * \mathscr{C}) \cup (\mathscr{B} * \mathscr{C})$.

Proof. All properties are immediate from the definition of the $*$ operation. □

We define an *integer interval* to be any finite set whose elements are consecutive integers, namely, $\{i, i+1, i+2, \ldots, j\}$ for some $i, j \in \mathbb{Z}$. As a special case, this includes the empty interval $\varnothing$. An *elementary family* is any family of the form

$$\mathscr{E} = \binom{I_1}{k_1} * \binom{I_2}{k_2} * \cdots * \binom{I_\ell}{k_\ell}, \tag{18}$$

where ℓ is a positive integer, $I_1, I_2, \ldots, I_\ell$ are pairwise disjoint integer intervals, and $k_1, k_2, \ldots, k_\ell \in \{0,1,2\}$. Trivial examples of elementary families are $\binom{\varnothing}{0} = \{\varnothing\}$ and $\binom{\varnothing}{1} = \varnothing$. Another example of an elementary family is the singleton family $\{A\}$ for any nonempty finite set $A \subseteq \mathbb{Z}$, using $\{A\} = \binom{\{a_1\}}{1} * \binom{\{a_2\}}{1} * \cdots * \binom{\{a_\ell\}}{1}$ where $a_1 < a_2 < \cdots < a_\ell$ are the distinct elements of A. We now define a partition measure that captures how efficiently a family can be partitioned into elementary families.

Definition 3.2 (Partition measure π). For any $\mathscr{A} \subseteq \mathscr{P}(\{1,2,\ldots,n\})$, define $\pi(\mathscr{A})$ to be the minimum

$$\sum_{i=1}^{N} |\mathscr{E}_i|^{1/2} \tag{19}$$

over all integers N and all elementary families $\mathscr{E}_1, \mathscr{E}_2, \ldots, \mathscr{E}_N$ that are pairwise disjoint and satisfy $\mathscr{E}_1 \cup \mathscr{E}_2 \cup \cdots \cup \mathscr{E}_N = \mathscr{A}$.

Straight from the definition, $\pi(\varnothing) = 0$ and $\pi(\{\varnothing\}) = 1$. More generally,

$$|\mathscr{A}|^{1/2} \leqslant \pi(\mathscr{A}) \leqslant |\mathscr{A}| \tag{20}$$

for every $\mathscr{A} \subseteq \mathscr{P}(\{1,2,\ldots,n\})$. The upper bound here corresponds to the trivial partition $\mathscr{A} = \bigcup_{A\in\mathscr{A}}\{A\}$. The lower bound holds because (19) is no smaller than $(\sum |\mathscr{E}_i|)^{1/2} = |\mathscr{A}|^{1/2}$.

Our analysis of the Fourier spectrum of decision trees relies on the partition measure of the family $\mathscr{P}_{n,k}$. To this end, we prove the following near-tight upper bound in the full version of this paper [23].

Theorem 3.3. *For some absolute constant $c \geqslant 1$ and all positive integers n and k,*

$$\pi(\mathscr{P}_{n,k}) \leqslant \frac{(2+\sqrt{2})^{k-1}c^{k-1}}{\sqrt{k}}\left(\frac{2n}{k}\right)^{k/2}.$$

4 FOURIER SPECTRUM OF DECISION TREES

This section is devoted to the proof of our main result on the Fourier spectrum of decision trees. Stated in its simplest terms, our result shows that for any function $f\colon \{-1,1\}^n \to \{-1,0,1\}$ computable by a decision tree of depth d, the sum of the absolute values of the Fourier coefficients of order k is at most

$$C^k\sqrt{\binom{d}{k}(1+\ln n)^{k-1}},$$

where $C \geqslant 1$ is an absolute constant that does not depend on n, d, k. Sections 4.1–4.3 focus on partitioning the Fourier spectrum of f into highly structured parts and analyzing each in isolation. Section 4.4 then recombines these pieces using the machinery of elementary families.

4.1 Slicing the Tree

Let T be a given decision tree of depth d in Boolean variables $x_1, x_2, \ldots, x_n$. For a set family $\mathscr{S} \subseteq \mathscr{P}(\{1,2,\ldots,d\})$, we define a real function $T|_{\mathscr{S}}\colon \{-1,1\}^n \to \mathbb{R}$ by

$$T|_{\mathscr{S}}(x) = \sum_{S\in\mathscr{S}}\sum_{v\in\{-1,1\}^d} T(v)\cdot 2^{-d}\prod_{i\in S} v_i x_{T(v_1v_2\ldots v_{i-1})}. \tag{21}$$

A straightforward but crucial observation is that $T|_{\mathscr{S}}$ is additive with respect to $\mathscr{S}$, in the following sense.

Proposition 4.1. *Let T be a depth-d decision tree. Let $\mathscr{S}', \mathscr{S}'' \subseteq \mathscr{P}(\{1,2,\ldots,d\})$ be set families with $\mathscr{S}' \cap \mathscr{S}'' = \varnothing$. Then*

$$T|_{\mathscr{S}'\cup\mathscr{S}''} = T|_{\mathscr{S}'} + T|_{\mathscr{S}''}.$$

Proof. Immediate by taking $\mathscr{S} = \mathscr{S}' \cup \mathscr{S}''$ in the defining equation (21). □

The relevance of (21) to the Fourier spectrum of decision trees is borne out by the following lemma.

Lemma 4.2. *Let T be a decision tree of depth d and degree at most 0, computing a function $f\colon \{-1,1\}^n \to \mathbb{R}$. Then*

$$L_k f = T|_{\mathscr{P}_{d,k}}, \qquad k = 0, 1, 2, \ldots, n.$$

Proof. By Proposition 2.2,

$$\begin{aligned} f(x) &= \sum_{v\in\{-1,1\}^d} T(v) \cdot \prod_{i=1}^{d} \frac{1 + v_i x_{T(v_1 v_2 \ldots v_{i-1})}}{2} \\ &= \sum_{v\in\{-1,1\}^d} T(v)\cdot 2^{-d} \sum_{S\subseteq\{1,2,\ldots,d\}} \prod_{i\in S} v_i x_{T(v_1 v_2 \ldots v_{i-1})} \\ &= \sum_{k=0}^{d} \sum_{S\in\mathscr{P}_{d,k}} \sum_{v\in\{-1,1\}^d} T(v)\cdot 2^{-d} \prod_{i\in S} v_i x_{T(v_1 v_2 \ldots v_{i-1})}. \end{aligned} \tag{22}$$

Since $\deg(T) \leqslant 0$, the coefficients $T(v)$ for $v \in \{-1,1\}^d$ are real numbers. Moreover, for any $v \in \{-1,1\}^d$ and $S \subseteq \{1,2,\ldots,d\}$, the definition of a decision tree ensures that $\prod_{i\in S} v_i x_{T(v_1 v_2 \ldots v_{i-1})}$ is a signed monomial of degree $|S|$. We conclude from (22) that the degree-k homogeneous part of f is

$$\begin{aligned} L_k f &= \sum_{S\in\mathscr{P}_{d,k}} \sum_{v\in\{-1,1\}^d} T(v)\cdot 2^{-d} \prod_{i\in S} v_i x_{T(v_1 v_2 \ldots v_{i-1})} \\ &= T|_{\mathscr{P}_{d,k}}. \end{aligned}$$

In particular, $L_k f = 0$ for $k \geqslant d+1$. ∎

4.2 Analytic Preliminaries

For positive integers m and k, define

$$\Lambda_{m,k}(p) = \begin{cases} 0 & \text{if } p = 0, \\ p\sqrt{\left(\frac{1}{k}\ln\frac{e^k m^{k-1}}{p}\right)^k} & \text{if } 0 < p \leqslant 1/m, \\ p\sqrt{\left(\ln\frac{e}{p}\right)(\ln em)^{k-1}} & \text{if } 1/m < p \leqslant 1. \end{cases}$$

Our bound for the Fourier spectrum of decision trees is in terms of this function. As preparation for our main result, we now collect the analytic properties of $\Lambda_{m,k}$ that we will need.

Lemma 4.3. *Let m and k be any positive integers. Then:*

(i) *$\Lambda_{m,k}$ is continuous on $[0,1]$;*
(ii) *$\Lambda_{m,k}$ is monotonically increasing on $[0,1]$;*
(iii) *$\Lambda_{m,k}$ is concave on $[0,1]$.*

We refer the reader to the full version of our paper [23] for the proof of Lemma 4.3 and of all other results in this section. The function $\Lambda_{m,k}$ arises as the solution to a natural optimization problem, which we now describe.

Lemma 4.4. *Let m and k be positive integers. Then for $0 < p \leqslant 1$, $\Lambda_{m,k}(p)$ equals*

$$p\max\left\{\prod_{i=1}^{k}\sqrt{\ln ex_i} : x_i \geqslant 1 \text{ and } x_1 x_2 \ldots x_i \leqslant \frac{m^{i-1}}{p} \text{ for all } i\right\}.$$

This optimization view of $\Lambda_{m,k}$ implies a host of useful facts that would be bothersome to prove directly. We state them as corollaries below.

Corollary 4.5. *Let m and k be positive integers. Then for all $p, q \in [0,1]$,*

$$q\Lambda_{m,k}(p) \leqslant \Lambda_{m,k}(pq).$$

Corollary 4.6. *Let m, k, ℓ be positive integers. Then for all $p, q \in [0,1]$,*

$$\Lambda_{m,k}(p)\,\Lambda_{m,\ell}\left(\frac{q}{m}\right) \leqslant \frac{\Lambda_{m,k+\ell}(pq)}{m}.$$

Corollary 4.7. *Let m and k be positive integers. Then for all $p \in [0,1]$,*

$$\Lambda_{m,k}(p) \leqslant \sqrt{2^k p}\cdot\Lambda_{m,k}(\sqrt{p}).$$

4.3 Contiguous Intervals

We have reached a focal point of this paper, where we analyze $T|_{\mathscr{E}}$ for arbitrary decision trees T and "canonical" elementary families $\mathscr{E}$. The families that we allow are those of the form

$$\mathscr{E} = \binom{I_1}{k_1} * \binom{I_2}{k_2} * \cdots * \binom{I_\ell}{k_\ell},$$

where $k_1, k_2, \ldots, k_\ell \in \{1,2\}$ and the integer intervals $I_1, I_2, \ldots, I_\ell$ form a partition of $\{1,2,\ldots,d\}$ with d being the depth of T. The proof proceeds by induction on ℓ. We will later generalize this result to arbitrary elementary families $\mathscr{E}$ and, from there, to all of $\mathscr{P}_{d,k}$ via the results of Section 3.

Theorem 4.8. *Let $T \in \mathscr{T}^*(n,d,p,0)$ be given, for some $0 \leqslant p \leqslant 1$ and integers $n, d \geqslant 1$. Let $\ell \geqslant 1$. Let $I_1, I_2, \ldots, I_\ell$ be pairwise disjoint integer intervals with $I_1 \cup I_2 \cup \cdots \cup I_\ell = \{1,2,\ldots,d\}$, and let $k_1, k_2, \ldots, k_\ell \in \{1,2\}$. Abbreviate $k = k_1 + k_2 + \cdots + k_\ell$. Then*

$$\left\| T|_{\binom{I_1}{k_1} * \binom{I_2}{k_2} * \cdots * \binom{I_\ell}{k_\ell}} \right\| \leqslant 2C^k\, 12^{\ell-1}\Lambda_{n^2,k}(p)\prod_{i=1}^{\ell}\binom{|I_i|}{k_i}^{1/2}, \tag{23}$$

where $C \geqslant 1$ is the absolute constant from Theorem 2.4.

Proof. The proof is by induction on ℓ. The base case $\ell = 1$ corresponds to $I_1 = \{1,2,\ldots,d\}$. Let $f\colon \{-1,1\}^n \to \{-1,0,1\}$ be the function computed by T. If $f \equiv 0$, we have $T|_{\binom{I_1}{k_1}} \equiv 0$ and the bound holds trivially. In the complementary case $f \not\equiv 0$, recall from Fact 2.3 that

$$\mathop{\mathbf{P}}_{x\in\{-1,1\}^n}[f(x) \neq 0] = p. \tag{24}$$

Then

$$\begin{aligned} \left\| T|_{\binom{I_1}{k_1}} \right\| &= \|L_{k_1} f\| \\ &\leqslant \binom{|I_1|}{k_1}^{1/2} C^{k_1} p \prod_{i=1}^{k_1}\sqrt{\ln\frac{en^{i-1}}{p}} \\ &\leqslant \binom{|I_1|}{k_1}^{1/2}\cdot 2C^{k_1} p\prod_{i=1}^{k_1}\sqrt{\ln\frac{en^{i-1}}{\sqrt{p}}} \\ &\leqslant \binom{|I_1|}{k_1}^{1/2}\cdot 2C^{k_1}\Lambda_{n^2,k_1}(p) \\ &= \binom{|I_1|}{k_1}^{1/2}\cdot 2C^{k}\Lambda_{n^2,k}(p), \end{aligned}$$

where the first step is valid by Lemma 4.2; the second step uses Theorem 2.4 along with (24) and $k_1 \leqslant 2$; and the fourth step applies Lemma 4.4. This settles the base case.

We now turn to the inductive step, $\ell \geqslant 2$. If $k_j > |I_j|$ for some j, then $T|_{\binom{I_1}{k_1}*\binom{I_2}{k_2}*\cdots*\binom{I_\ell}{k_\ell}} = T|_\varnothing = 0$, and the claimed bound holds trivially. We may therefore assume that $k_j \leqslant |I_j|$ for every $j = 1, 2, \ldots, \ell$. This means in particular that the intervals $I_1, I_2, \ldots, I_\ell$ are nonempty. Furthermore, by renumbering the intervals if necessary, we may assume that $I_1 < I_2 < \cdots < I_\ell$. Put $d' = \max I_{\ell-1}$, so that $I_\ell = \{d'+1, d'+2, \ldots, d\}$. Abbreviate

$$\mathscr{S}' = \binom{I_1}{k_1} * \binom{I_2}{k_2} * \cdots * \binom{I_{\ell-1}}{k_{\ell-1}},$$
$$\mathscr{S} = \mathscr{S}' * \binom{I_\ell}{k_\ell}.$$

For $j = 0, 1, 2, \ldots$, define a depth-d' decision tree T'_j by

$$T'_j(v) = \begin{cases} T(v) & \text{if } v \in \{-1,1\}^{\leqslant d'-1}, \\ T_v|_{\binom{\{1,2,\ldots,|I_\ell|\}}{k_\ell}} & \text{if } v \in \{-1,1\}^{d'} \text{ and } \frac{1}{3^{j+1}} < \operatorname{dns}(T_v) \leqslant \frac{1}{3^j}, \\ 0 & \text{otherwise.} \end{cases}$$

Observe that T'_j is a valid decision tree in that for every leaf $v \in \{-1,1\}^{d'}$, the label $T'_j(v) \in \mathbb{R}[x_1, x_2, \ldots, x_n]$ is a function that does not depend on any of the variables

$$x_{T(\varepsilon)}, x_{T(v_1)}, x_{T(v_1v_2)}, \ldots, x_{T(v_1v_2\ldots v_{d'-1})} \tag{25}$$

queried along the path from the root to v. Indeed, recall from Lemma 4.2 that $T_v|_{\binom{\{1,2,\ldots,|I_\ell|\}}{k_\ell}}$ is the k_ℓ-th homogeneous part of the function computed by the subtree T_v, which by definition does not use any of the variables (25). We also note that all but finitely many of the trees $T_0, T_1, T_2, \ldots$ are identically zero; however, working with the infinite sequence is more convenient from the point of view of notation and calculations.

The weighted densities of $T'_0, T'_1, T'_2, \ldots$ are given by

$$\begin{aligned} \sum_{j=0}^\infty 3^{-j} \operatorname{dns}(T'_j) &= \sum_{j=0}^\infty 3^{-j} \mathop{\mathbf{P}}_{v\in\{-1,1\}^{d'}}[T'_j(v) \neq 0] \\ &\leqslant \sum_{j=0}^\infty 3^{-j} \mathop{\mathbf{P}}_{v\in\{-1,1\}^{d'}}[3^{-j-1} < \operatorname{dns}(T_v) \leqslant 3^{-j}] \\ &\leqslant 3 \mathop{\mathbf{E}}_{v\in\{-1,1\}^{d'}} \operatorname{dns}(T_v) \\ &= 3 \operatorname{dns}(T) \\ &= 3p. \end{aligned} \tag{26}$$

The relevance of T'_j to our analysis of $T|_{\mathscr{S}}$ is clear from the following claims, whose proofs are available in the full version of this paper [23].

Claim 4.9. *$T|_{\mathscr{S}} = \sum_{j=0}^\infty T'_j|_{\mathscr{S}'}$.*

Claim 4.10. *Let $j = 0, 1, 2, \ldots$ be given. Then $\|T'_j|_{\mathscr{S}'}\|$ is at most*

$$8C^k\, 12^{\ell-2} \binom{|I_1|}{k_1}^{1/2} \cdots \binom{|I_\ell|}{k_\ell}^{1/2} \cdot \sqrt{3^{-j}}\, \Lambda_{n^2,k}(\sqrt{3^{-j}} \operatorname{dns}(T'_j)).$$

We now complete the proof of the theorem. Set $s = \sum_{i=0}^\infty \sqrt{3^{-i}} = 2.3660\ldots$. Then

$$\begin{aligned} \sum_{j=0}^\infty \sqrt{3^{-j}} \Lambda_{n^2,k}(\sqrt{3^{-j}} \operatorname{dns}(T'_j)) &= s \sum_{j=0}^\infty \frac{\sqrt{3^{-j}}}{s} \Lambda_{n^2,k}(\sqrt{3^{-j}} \operatorname{dns}(T'_j)) \\ &\leqslant s \Lambda_{n^2,k} \left(\sum_{j=0}^\infty \frac{\sqrt{3^{-j}}}{s} \cdot \sqrt{3^{-j}} \operatorname{dns}(T'_j) \right) \\ &\leqslant 3\Lambda_{n^2,k} \left(\frac{s}{3} \sum_{j=0}^\infty \frac{\sqrt{3^{-j}}}{s} \cdot \sqrt{3^{-j}} \operatorname{dns}(T'_j) \right) \\ &\leqslant 3\Lambda_{n^2,k}(p), \end{aligned} \tag{27}$$

where the second step is valid by Lemma 4.3 (iii); the third step uses Corollary 4.5 with $q = s/3$; and the final step is justified by (26) and Lemma 4.3 (ii). As a result,

$$\begin{aligned} \|T|_{\mathscr{S}}\| &\leqslant \sum_{j=0}^\infty \|T'_j|_{\mathscr{S}'}\| \\ &\leqslant 8C^k\, 12^{\ell-2} \binom{|I_1|}{k_1}^{1/2} \cdots \binom{|I_\ell|}{k_\ell}^{1/2} \sum_{j=0}^\infty \sqrt{3^{-j}} \Lambda_{n^2,k}(\sqrt{3^{-j}} \operatorname{dns}(T'_j)) \\ &\leqslant 2C^k\, 12^{\ell-1} \binom{|I_1|}{k_1}^{1/2} \cdots \binom{|I_\ell|}{k_\ell}^{1/2} \Lambda_{n^2,k}(p), \end{aligned}$$

where the first step is valid by Proposition 2.1 and Claim 4.9, bearing in mind once again that all but finitely many of the $T'_j|_{\mathscr{S}'}$ are identically zero; the second step is a substitution from Claim 4.10; and the final step uses (27). This completes the inductive step. □

4.4 Main Result

En route to our main result on the Fourier spectrum of decision trees, we now generalize Theorem 4.8 to arbitrary elementary families $\mathscr{E}$.

Theorem 4.11. *Let $T \in \mathscr{T}^*(n, d, p, 0)$ be given, for some $0 \leqslant p \leqslant 1$ and integers $n, d \geqslant 1$. Let k be an integer with $1 \leqslant k \leqslant d$. Then every elementary family $\mathscr{E} \subseteq \mathscr{P}_{d,k}$ satisfies*

$$\|T|_{\mathscr{E}}\| \leqslant (12C)^k \Lambda_{n^2,k}(p) \sqrt{|\mathscr{E}|}, \tag{28}$$

where $C \geqslant 1$ is the absolute constant from Theorem 2.4.

The proof of this result is available in the full version of our paper [23]. We now obtain our main result on the Fourier spectrum of decision trees by combining Theorem 4.11 with an efficient decomposition of $\mathscr{P}_{d,k}$ into elementary families (Theorem 3.3).

Theorem 4.12. *Let $f\colon \{-1,1\}^n \to \{-1,0,1\}$ be a function computable by a decision tree of depth d. Define $p = \mathbf{P}_{x\in\{-1,1\}^n}[f(x) \neq 0]$. Then*

$$\|L_k f\| \leqslant \binom{d}{k}^{1/2} (58Cc)^k\, \Lambda_{n^2,k}(p), \qquad k = 1, 2, \ldots, n,$$

where $C \geqslant 1$ and $c \geqslant 1$ are the absolute constants from Theorem 2.4 and Theorem 3.3, respectively.

Proof. Lemma 4.2 ensures that $L_k f = 0$ for $k > d$, so that the theorem holds vacuously in that case. We now examine the complementary possibility, $1 \leqslant k \leqslant d$. For some integer $N \geqslant 1$, Theorem 3.3 gives a partition $\mathscr{P}_{d,k} = \bigcup_{i=1}^N \mathscr{E}_i$ where $\mathscr{E}_1, \mathscr{E}_2, \ldots, \mathscr{E}_N$

are elementary families with

$$\sum_{i=1}^{N} |\mathscr{E}_i|^{1/2} \leqslant (2+2\sqrt{2})^k c^k \left(\frac{d}{k}\right)^{k/2}. \tag{29}$$

Fix a decision tree T of depth d that computes f. Then Fact 2.3 shows that $T \in \mathscr{T}^*(n,d,p,0)$. As a result,

$$\begin{aligned}
\|\!| L_k f |\!\| &= \|\!| T|_{\mathscr{P}_{d,k}} |\!\| \\
&= \left\|\!\left| \sum_{i=1}^{N} T|_{\mathscr{E}_i} \right|\!\right\| \\
&\leqslant \sum_{i=1}^{N} \|\!| T|_{\mathscr{E}_i} |\!\| \\
&\leqslant \sum_{i=1}^{N} (12C)^k \, \Lambda_{n^2,k}(p) \sqrt{|\mathscr{E}_i|} \\
&\leqslant \left(\frac{d}{k}\right)^{k/2} (58Cc)^k \, \Lambda_{n^2,k}(p),
\end{aligned}$$

where the first step is valid by Lemma 4.2; the second step uses Proposition 4.1; the third step uses Proposition 2.1; the fourth step applies Theorem 4.11; and the final step substitutes the upper bound from (29). In view of (14), the proof is complete. □

Maximizing over $0 \leqslant p \leqslant 1$, we establish the following clean bound conjectured by Tal [28].

Corollary 4.13. *Let $f\colon \{-1,1\}^n \to \{-1,0,1\}$ be a function computable by a decision tree of depth d. Then*

$$\|\!| L_k f |\!\| \leqslant C^k \sqrt{\binom{d}{k} (1+\ln n)^{k-1}}, \qquad k = 1,2,\ldots,n,$$

where $C \geqslant 1$ is an absolute constant.

Proof. Recall from Lemma 4.3 (ii) that $\Lambda_{n^2,k}(p) \leqslant \sqrt{(\ln en^2)^{k-1}}$ for all $0 \leqslant p \leqslant 1$. Now the claimed bound is immediate from Theorem 4.12 after a change of constant C. □

Corollary 4.13 settles Theorem 1.8 from the introduction. By convexity (Proposition 2.1), Corollary 4.13 holds more generally for any real function $f\colon \{-1,1\}^n \to [-1,1]$ computable by a decision tree of depth d.

5 QUANTUM VERSUS CLASSICAL QUERY COMPLEXITY

Using our newly derived bound for the Fourier spectrum of decision trees, we will now prove the main result of this paper on quantum versus randomized query complexity.

5.1 Quantum and Randomized Query Models

For a nonempty finite set X, a *partial Boolean function on X* is a mapping $X \to \{0,1,*\}$, where the output value $*$ is reserved for illegal inputs. Recall that a *randomized query algorithm of cost d* is a probability distribution on decision trees of depth at most d. For a (possibly partial) Boolean function f on the Boolean hypercube, we say that a randomized query algorithm *computes f with error ε* if, for every input $x \in f^{-1}(0) \cup f^{-1}(1)$, the algorithm outputs $f(x)$ with probability at least $1-\varepsilon$. Observe that in this formalism, the algorithm is allowed to exhibit arbitrary behavior on the illegal inputs, namely, those in $f^{-1}(*)$. The *randomized query complexity* $R_\varepsilon(f)$ is the minimum cost of a randomized query algorithm that computes f with error ε. The canonical setting of the error parameter is $\varepsilon = 1/3$. This choice is largely arbitrary because the error of a query algorithm can be reduced in an efficient manner by running the algorithm several times independently and outputting the majority answer. Quantitatively, the following relation follows from the Chernoff bound:

$$R_\varepsilon(f) \leqslant O\left(\frac{1}{\gamma^2} \log \frac{1}{\varepsilon}\right) \cdot R_{\frac{1}{2}-\gamma}(f) \tag{30}$$

for all $\varepsilon, \gamma \leqslant 1/2$.

These classical definitions carry over in the obvious way to the quantum model. Here, the cost is the worst-case number of quantum queries on any input, and a quantum algorithm is said to *compute f with error ε* if, for every input $x \in f^{-1}(0) \cup f^{-1}(1)$, the algorithm outputs $f(x)$ with probability at least $1-\varepsilon$. The *quantum query complexity* $Q_\varepsilon(f)$ is the minimum cost of a quantum query algorithm that computes f with error ε. For an excellent introduction to classical and quantum query complexity, we refer the reader to [8] and [29], respectively.

5.2 The Rorrelation Problem

We now formally state the problem of interest to us, Tal's *rorrelation* [28], which was briefly reviewed in the introduction. Let n and k be positive integers. For an orthogonal matrix $U \in \mathbb{R}^{n\times n}$, consider the multilinear polynomial $\phi_{n,k,U}\colon (\{-1,1\}^n)^k \to \mathbb{R}$ given by

$$\phi_{n,k,U}(x_1,x_2,\ldots,x_k) = \frac{1}{n} \mathbf{1}^\intercal D_{x_1} U D_{x_2} U D_{x_3} U \cdots U D_{x_k} \mathbf{1}, \tag{31}$$

where $\mathbf{1}$ denotes the all-ones vector and D_{x_i} denotes the diagonal matrix with vector x_i on the diagonal. In what follows, we treat the sets $(\{-1,1\}^n)^k$ and $\{-1,1\}^{n\times k}$ interchangeably, thereby interpreting the input to $\phi_{n,k,U}$ as an $n \times k$ sign matrix. Let $\|\cdot\|_2$ denote the Euclidean norm. Then for all $x_1, x_2, \ldots, x_k \in \{-1,1\}^n$, we have

$$\begin{aligned}
|\phi_{n,k,U}(x_1,x_2,\ldots,x_k)| &= \frac{1}{n} \langle \mathbf{1}, D_{x_1} U D_{x_2} U D_{x_3} U \cdots U D_{x_k} \mathbf{1} \rangle \\
&\leqslant \frac{1}{n} \|\mathbf{1}\|_2 \, \|D_{x_1} U D_{x_2} U D_{x_3} U \cdots U D_{x_k} \mathbf{1}\|_2 \\
&= \frac{1}{n} \|\mathbf{1}\|_2 \, \|\mathbf{1}\|_2 \\
&= 1,
\end{aligned} \tag{32}$$

where the second step applies the Cauchy–Schwarz inequality, and the third step is valid because each of the matrices involved preserves the Euclidean norm. In particular, the multivariate polynomial $\phi_{n,k,U}$ ranges in $[-1,1]$ for all inputs. Generalizing the forrelation problem of Aaronson and Ambainis [1], Tal [28] considered the partial Boolean function $f_{n,k,U}\colon \{-1,1\}^{n\times k} \to \{0,1,*\}$ given by

$$f_{n,k,U}(x) = \begin{cases} 1 & \text{if } \phi_{n,k,U}(x) \geqslant 2^{-k}, \\ 0 & \text{if } |\phi_{n,k,U}(x)| \leqslant 2^{-k-1}, \\ * & \text{otherwise.} \end{cases}$$

Aaronson and Ambainis [1] showed that there is a quantum algorithm with $\lceil k/2 \rceil$ queries whose acceptance probability on input

$x \in \{-1,1\}^{n\times k}$ is $(\phi_{n,k,H}(x)+1)/2$, where H is the Hadamard transform matrix. Their analysis generalizes to any orthogonal matrix in place of H, to the following effect.

Fact 5.1 (Tal [28, Claim 3.1]). *Let n and k be positive integers, where n is a power of 2. Let U be an arbitrary orthogonal matrix. Then there is a quantum query algorithm with $\lceil k/2\rceil$ queries whose acceptance probability on input $x \in \{-1,1\}^{n\times k}$ is $(\phi_{n,k,U}(x)+1)/2$.*

Corollary 5.2. *Let n and k be positive integers, where n is a power of 2. Let U be an arbitrary orthogonal matrix. Then*

$$Q_{\frac{1}{2}-\frac{1}{2^{k+4}}}(f_{n,k,U}) \leqslant \left\lceil \frac{k}{2} \right\rceil. \tag{33}$$

In particular,

$$Q_{1/3}(f_{n,k,U}) \leqslant O(k4^k). \tag{34}$$

Proof. On input x, the query algorithm for (33) is as follows: with probability p, run the algorithm of Fact 5.1 and output the resulting answer; with complementary probability $1-p$, output "no" regardless of x. By design, the proposed solution has query cost at most $\lceil k/2\rceil$ and accepts x with probability exactly

$$p \cdot \frac{\phi_{n,k,U}(x)+1}{2}.$$

We want this quantity to be at most $\frac{1}{2}-2^{-k-4}$ if $\phi_{n,k,U}(x) \leqslant 2^{-k-1}$, and at least $\frac{1}{2}+2^{-k-4}$ if $\phi_{n,k,U}(x) \geqslant 2^{-k}$. These requirements are both met for $p = (1+\frac{3}{2^{k+2}})^{-1}$. In summary, $f_{n,k,U}$ has a query algorithm with error at most $\frac{1}{2}-2^{-k-4}$ and query cost $\lceil k/2\rceil$. To reduce the error to $1/3$, run this algorithm independently $\Theta(4^k)$ times and output the majority answer; cf. (30). □

Corollary 5.2 shows that the rorrelation problem has small quantum query complexity. By contrast, we will show that its randomized complexity is essentially the maximum possible. Specifically, we will prove a near-linear lower bound on the randomized query complexity of rorrelation by combining Tal's work [28] with our near-optimal bounds for the Fourier spectrum of decision trees.

In what follows, let $\mathscr{U}_{n,k}$ denote the uniform probability distribution on $\{-1,1\}^{n\times k}$. Applying Parseval's identity to the multilinear polynomial $\phi_{n,k,U}$ gives:

Fact 5.3 (Tal [28, Claim 4.4]). $\mathbf{E}_{x\sim\mathscr{U}_{n,k}}[\phi_{n,k,U}(x)^2] = 1/n$.

The other result from [28] that we will need is as follows.

Fact 5.4 (Tal [28, Lemmas 5.6, 5.7, and Claim 4.1]). *Let n and k be positive integers. Let $U \in \mathbb{R}^{n\times n}$ be a uniformly random orthogonal matrix. Then with probability $1-o(1)$, there exists a probability distribution $\mathscr{D}_{n,k,U}$ on $\{-1,1\}^{n\times k}$ such that:*

$$\mathop{\mathbf{E}}_{x\sim\mathscr{D}_{n,k,U}} \phi_{n,k,U}(x) \geqslant \left(\frac{2}{\pi}\right)^{k-1}, \tag{35}$$

$$\mathop{\mathbf{E}}_{x\sim\mathscr{D}_{n,k,U}} \prod_{(i,j)\in S} x_{i,j} = 0, \qquad |S| = 1,2,\ldots,k-1, \tag{36}$$

$$\left|\mathop{\mathbf{E}}_{x\sim\mathscr{D}_{n,k,U}} \prod_{(i,j)\in S} x_{i,j}\right| \leqslant \left(\frac{c|S|\log n}{n}\right)^{\frac{|S|}{2}\cdot\frac{k-1}{k}}, \qquad |S| = k,k+1,\ldots,nk, \tag{37}$$

where $c \geqslant 1$ is an absolute constant independent of n,k,U.

5.3 The Quantum-Classical Separation

In this section, we derive our lower bound on the randomized query complexity of the rorrelation problem by combining Tal's Facts 5.3 and 5.4 with our main result on decision trees (Corollary 4.13). The technical centerpiece of this derivation is the following "indistinguishability" lemma, which is a polynomial improvement on the analogous calculation by Tal [28, Theorem 5.8] that used weaker Fourier bounds for decision trees.

Lemma 5.5. *Let n and k be positive integers. Let $U \in \mathbb{R}^{n\times n}$ be a uniformly random orthogonal matrix. Then with probability $1-o(1)$, every function $g\colon \{-1,1\}^{n\times k} \to \{0,1\}$ obeys*

$$\left|\mathop{\mathbf{E}}_{\mathscr{U}_{n,k}} g - \mathop{\mathbf{E}}_{\mathscr{D}_{n,k,U}} g\right| \leqslant \left(cd \cdot \frac{\log^{2-\frac{1}{k}}(n+k)}{n^{1-\frac{1}{k}}}\right)^{k/2}, \tag{38}$$

where $\mathscr{D}_{n,k,U}$ is as defined in Fact 5.4; d is the minimum depth of a decision tree that computes g; and $c \geqslant 1$ is an absolute constant independent of n,k,U,g.

Proof. Fact 5.4 guarantees that with probability $1-o(1)$, there is a probability distribution $\mathscr{D}_{n,k,U}$ on $\{-1,1\}^{n\times k}$ that obeys (35)–(37). Conditioned on this event, we will prove (38). To start with, fix g and write out the Fourier expansion

$$\begin{aligned} g(x) &= \sum_{S\subseteq\{1,2,\ldots,n\}\times\{1,2,\ldots,k\}} \hat{g}(S) \prod_{(i,j)\in S} x_{i,j} \\ &= \sum_{\ell=0}^{nk} \sum_{|S|=\ell} \hat{g}(S) \prod_{(i,j)\in S} x_{i,j}. \end{aligned}$$

Then

$$\begin{aligned} &\left|\mathop{\mathbf{E}}_{\mathscr{U}_{n,k}} g - \mathop{\mathbf{E}}_{\mathscr{D}_{n,k,U}} g\right| \\ &\quad\leqslant \sum_{\ell=0}^{nk} \sum_{|S|=\ell} |\hat{g}(S)| \left|\mathop{\mathbf{E}}_{\mathscr{U}_{n,k}} \prod_{(i,j)\in S} x_{i,j} - \mathop{\mathbf{E}}_{\mathscr{D}_{n,k,U}} \prod_{(i,j)\in S} x_{i,j}\right| \\ &\quad\leqslant \sum_{\ell=1}^{nk} \sum_{|S|=\ell} |\hat{g}(S)| \left|\mathop{\mathbf{E}}_{\mathscr{U}_{n,k}} \prod_{(i,j)\in S} x_{i,j} - \mathop{\mathbf{E}}_{\mathscr{D}_{n,k,U}} \prod_{(i,j)\in S} x_{i,j}\right| \\ &\quad\leqslant \sum_{\ell=k}^{nk} \sum_{|S|=\ell} |\hat{g}(S)| \left|\mathop{\mathbf{E}}_{\mathscr{D}_{n,k,U}} \prod_{(i,j)\in S} x_{i,j}\right|, \end{aligned}$$

where the first step uses the triangle inequality; the second step is justified by $\mathbf{E}_{\mathscr{U}_{n,k}} 1 = \mathbf{E}_{\mathscr{D}_{n,k,U}} 1 = 1$; and the third step is valid due to (36) and the identity $\mathbf{E}_{\mathscr{U}_{n,k}} \prod_{(i,j)\in S} x_{i,j} = 0$ for nonempty S. Let d be the minimum depth of a decision tree that computes g. Applying (37) then Corollary 4.13, we conclude that

$$\left|\mathop{\mathbf{E}}_{\mathscr{U}_{n,k}} g - \mathop{\mathbf{E}}_{\mathscr{D}_{n,k,U}} g\right| \leqslant \sum_{\ell=k}^{nk} c_1^\ell \sqrt{\binom{d}{\ell}(1+\ln nk)^{\ell-1}} \left(\frac{c_2 \ell \log n}{n}\right)^{\frac{\ell}{2}\cdot\frac{k-1}{k}},$$

where $c_1 \geqslant 1$ and $c_2 \geqslant 1$ are the absolute constants in Corollary 4.13 and Fact 5.4. In view of (14), this gives

$$\begin{aligned}\left|\mathop{\mathbf{E}}_{\mathscr{U}_{n,k}} g - \mathop{\mathbf{E}}_{\mathscr{D}_{n,k,U}} g\right| &\leqslant \sum_{\ell=k}^{\infty}\left(\frac{c_1^2 ed}{\ell}\cdot(1+\ln nk)^{\frac{\ell-1}{\ell}}\left(\frac{c_2\ell\log n}{n}\right)^{\frac{k-1}{k}}\right)^{\frac{\ell}{2}}\\ &\leqslant \sum_{\ell=k}^{\infty}\left(c_1^2\cdot ed\cdot(1+\ln nk)\cdot\left(\frac{c_2\log n}{n}\right)^{\frac{k-1}{k}}\right)^{\frac{\ell}{2}}\\ &\leqslant \sum_{\ell=k}^{\infty}\left(\frac{cd}{4}\cdot\frac{\log^{2-\frac{1}{k}}(n+k)}{n^{1-\frac{1}{k}}}\right)^{\frac{\ell}{2}},\end{aligned}$$

where $c \geqslant 1$ in the last step is a sufficiently large absolute constant. This settles (38) in the case when $cd\log^{(2k-1)/k}(n+k) \leqslant n^{(k-1)/k}$. In the complementary case, (38) follows from the trivial bound $|\mathbf{E}_{\mathscr{U}_{n,k}} g - \mathbf{E}_{\mathscr{D}_{n,k,U}} g| \leqslant 1$. □

We have reached the main result of this section, an essentially tight lower bound on the randomized query complexity of the k-fold rorrelation problem.

Theorem 5.6. *Let n and k be positive integers, with $k \leqslant \frac{1}{3}\log n - 1$. Let $U \in \mathbb{R}^{n\times n}$ be a uniformly random orthogonal matrix. Then with probability $1 - o(1)$,*

$$R_{1/2^{k+1}}(f_{n,k,U}) = \Omega\left(\frac{n^{1-\frac{1}{k}}}{(\log n)^{2-\frac{1}{k}}}\right) \tag{39}$$

and in particular

$$R_{\frac{1}{2}-\gamma}(f_{n,k,U}) = \Omega\left(\frac{\gamma^2}{k}\cdot\frac{n^{1-\frac{1}{k}}}{(\log n)^{2-\frac{1}{k}}}\right), \qquad 0 \leqslant \gamma \leqslant \frac{1}{2}. \tag{40}$$

Proof. We will prove the lower bound for every U that satisfies (35) and (38), which happens with probability $1 - o(1)$ by Fact 5.4 and Lemma 5.5. To begin with,

$$\begin{aligned}\mathop{\mathbf{P}}_{\mathscr{U}_{n,k}}[f_{n,k,U}(x) \neq 0] &= \mathop{\mathbf{P}}_{\mathscr{U}_{n,k}}[|\phi_{n,k,U}(x)| > 2^{-k-1}]\\ &\leqslant 4^{k+1}\mathop{\mathbf{E}}_{\mathscr{U}_{n,k}}[\phi_{n,k,U}(x)^2]\\ &\leqslant \frac{4^{k+1}}{n}\\ &\leqslant \frac{1}{2^{k+1}},\end{aligned} \tag{41}$$

where the last three steps use Markov's inequality, Fact 5.3, and $k \leqslant \frac{1}{3}\log n - 1$, respectively. Also,

$$\begin{aligned}\left(\frac{2}{\pi}\right)^{k-1} &\leqslant \mathop{\mathbf{E}}_{\mathscr{D}_{n,k,U}}\phi_{n,k,U}(x)\\ &\leqslant 2^{-k}\mathop{\mathbf{P}}_{\mathscr{D}_{n,k,U}}[\phi_{n,k,U}(x) < 2^{-k}] + \mathop{\mathbf{P}}_{\mathscr{D}_{n,k,U}}[\phi_{n,k,U}(x) \geqslant 2^{-k}]\\ &= 2^{-k}(1 - \mathop{\mathbf{P}}_{\mathscr{D}_{n,k,U}}[f_{n,k,U}(x) = 1]) + \mathop{\mathbf{P}}_{\mathscr{D}_{n,k,U}}[f_{n,k,U}(x) = 1]\\ &= 2^{-k} + (1-2^{-k})\mathop{\mathbf{P}}_{\mathscr{D}_{n,k,U}}[f_{n,k,U}(x) = 1],\end{aligned}$$

where the first and second steps are justified by (35) and (32), respectively. The last equation shows that

$$\begin{aligned}\mathop{\mathbf{P}}_{\mathscr{D}_{n,k,U}}[f_{n,k,U}(x) = 1] &\geqslant \left(\frac{2}{\pi}\right)^{k-1} - 2^{-k}\\ &\geqslant 2^{-k}.\end{aligned} \tag{42}$$

Now fix arbitrary parameters $d \geqslant 1$ and $0 \leqslant \varepsilon \leqslant 1/2$, and consider a randomized query algorithm of cost d that computes $f_{n,k,U}$ with error at most ε. Then the algorithm's acceptance probability on given input x is $\mathbf{E}_r\, g_r(x)$, where r denotes a random string and each $g_r\colon \{-1,1\}^{n\times k} \to \{0,1\}$ is computable by a decision tree of depth at most d. Since the error is at most ε, we have

$$\mathop{\mathbf{P}}_r[f_{n,k,U}(x) = 0,\ g_r(x) = 1] + \mathop{\mathbf{P}}_r[f_{n,k,U}(x) = 1,\ g_r(x) = 0] \leqslant \varepsilon \tag{43}$$

for every $x \in \{-1,1\}^{n\times k}$. We thus obtain the two inequalities

$$\mathop{\mathbf{E}}_r \mathop{\mathbf{P}}_{\mathscr{U}_{n,k}}[f_{n,k,U}(x) = 0,\ g_r(x) = 1] \leqslant \varepsilon, \tag{44}$$

$$\mathop{\mathbf{E}}_r \mathop{\mathbf{P}}_{\mathscr{D}_{n,k,U}}[f_{n,k,U}(x) = 1,\ g_r(x) = 0] \leqslant \varepsilon, \tag{45}$$

by passing to expectations in (43) with respect to $x \sim \mathscr{U}_{n,k}$ and $x \sim \mathscr{D}_{n,k,U}$, respectively. On the other hand, (38) and $k = O(\log n)$ imply

$$\mathop{\mathbf{E}}_r\left|\mathop{\mathbf{E}}_{\mathscr{D}_{n,k,U}} g_r - \mathop{\mathbf{E}}_{\mathscr{U}_{n,k}} g_r\right| \leqslant \left(c'd\cdot\frac{(\log n)^{2-\frac{1}{k}}}{n^{1-\frac{1}{k}}}\right)^{\frac{k}{2}} \tag{46}$$

for some absolute constant $c' \geqslant 1$.

We now have all the ingredients to complete the proof. For each r, we have

$$\begin{aligned}\mathop{\mathbf{E}}_{\mathscr{D}_{n,k,U}} g_r &= \mathop{\mathbf{P}}_{\mathscr{D}_{n,k,U}}[g_r(x) = 1]\\ &\geqslant \mathop{\mathbf{P}}_{\mathscr{D}_{n,k,U}}[f_{n,k,U}(x) = 1]\\ &\qquad - \mathop{\mathbf{P}}_{\mathscr{D}_{n,k,U}}[f_{n,k,U}(x) = 1,\ g_r(x) = 0]\\ &\geqslant 2^{-k} - \mathop{\mathbf{P}}_{\mathscr{D}_{n,k,U}}[f_{n,k,U}(x) = 1,\ g_r(x) = 0],\end{aligned} \tag{47}$$

where the last step uses (42). Similarly,

$$\begin{aligned}\mathop{\mathbf{E}}_{\mathscr{U}_{n,k}} g_r &= \mathop{\mathbf{P}}_{\mathscr{U}_{n,k}}[g_r(x) = 1]\\ &\leqslant \mathop{\mathbf{P}}_{\mathscr{U}_{n,k}}[f_{n,k,U}(x) \neq 0] + \mathop{\mathbf{P}}_{\mathscr{U}_{n,k}}[f_{n,k,U}(x) = 0,\ g_r(x) = 1]\\ &\leqslant 2^{-k-1} + \mathop{\mathbf{P}}_{\mathscr{U}_{n,k}}[f_{n,k,U}(x) = 0,\ g_r(x) = 1],\end{aligned} \tag{48}$$

where the last step uses (41). Passing to expectations in (47) and (48) with respect to r gives

$$\begin{aligned}&\mathop{\mathbf{E}}_r\left[\mathop{\mathbf{E}}_{\mathscr{D}_{n,k,U}} g_r - \mathop{\mathbf{E}}_{\mathscr{U}_{n,k}} g_r\right]\\ &\quad\geqslant 2^{-k-1} - \mathop{\mathbf{E}}_r\mathop{\mathbf{P}}_{\mathscr{D}_{n,k,U}}[f_{n,k,U}(x) = 1,\ g_r(x) = 0]\\ &\qquad - \mathop{\mathbf{E}}_r\mathop{\mathbf{P}}_{\mathscr{U}_{n,k}}[f_{n,k,U}(x) = 0,\ g_r(x) = 1],\end{aligned}$$

which in view of (44) and (45) simplifies to

$$\mathop{\mathbf{E}}_r\left[\mathop{\mathbf{E}}_{\mathscr{D}_{n,k,U}} g_r - \mathop{\mathbf{E}}_{\mathscr{U}_{n,k}} g_r\right] \geqslant 2^{-k-1} - 2\varepsilon.$$

Comparing this lower bound with (46), we arrive at

$$\left(c'd \cdot \frac{(\log n)^{2-\frac{1}{k}}}{n^{1-\frac{1}{k}}}\right)^{\frac{k}{2}} \geqslant 2^{-k-1} - 2\varepsilon.$$

Taking $\varepsilon = 2^{-k-3}$ and solving for d, we find that

$$R_{2^{-k-3}}(f_{n,k,U}) = \Omega\left(\frac{n^{1-\frac{1}{k}}}{(\log n)^{2-\frac{1}{k}}}\right).$$

By the error reduction formula (30), this settles (39) and (40). □

Theorem 5.6 settles Theorem 1.1 from the introduction. Corollary 1.2 now follows from (34) and Theorem 1.1 by taking $k = \lceil 1/\varepsilon \rceil + 1$ and $\gamma = 1/6$. Similarly, Corollary 1.3 follows from (34) and Theorem 1.1 by setting $\gamma = 1/6$ and taking $k = k(n)$ to be a sufficiently slow-growing function.

ACKNOWLEDGMENTS

This work was supported in part by NSF grant CCF-1814947. The authors are thankful to Nikhil Bansal, Dmitry Gavinsky, Makrand Sinha, Avishay Tal, and Ronald de Wolf for valuable comments on an earlier version of this paper. We are additionally thankful to Nikhil and Makrand for reminding us that quantum-classical query separations automatically imply analogous separations in communication complexity.

REFERENCES

[1] Scott Aaronson and Andris Ambainis. 2018. Forrelation: A Problem That Optimally Separates Quantum from Classical Computing. *SIAM J. Comput.* 47, 3 (2018), 982–1038. https://doi.org/10.1137/15M1050902

[2] Scott Aaronson, Shalev Ben-David, and Robin Kothari. 2016. Separations in query complexity using cheat sheets. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing* (STOC). 863–876. https://doi.org/10.1145/2897518.2897644

[3] Scott Aaronson, Shalev Ben-David, Robin Kothari, and Avishay Tal. 2020. Quantum Implications of Huang's Sensitivity Theorem. Available at https://arxiv.org/abs/2004.13231.

[4] Nikhil Bansal and Makrand Sinha. 2020. k-Forrelation Optimally Separates Quantum and Classical Query Complexity. Available at https://arxiv.org/abs/2008.07003.

[5] Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald de Wolf. 2001. Quantum lower bounds by polynomials. *J. ACM* 48, 4 (2001), 778–797. https://doi.org/10.1145/502090.502097

[6] Ethan Bernstein and Umesh V. Vazirani. 1997. Quantum Complexity Theory. *SIAM J. Comput.* 26, 5 (1997), 1411–1473. https://doi.org/10.1137/S0097539796300921

[7] Harry Buhrman, Richard Cleve, and Avi Wigderson. 1998. Quantum vs. Classical Communication and Computation. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing* (STOC). 63–68. https://doi.org/10.1145/276698.276713

[8] Harry Buhrman and Ronald de Wolf. 2002. Complexity measures and decision tree complexity: A survey. *Theor. Comput. Sci.* 288, 1 (2002), 21–43. https://doi.org/10.1016/S0304-3975(01)00144-X

[9] Harry Buhrman, Lance Fortnow, Ilan Newman, and Hein Röhrig. 2008. Quantum Property Testing. *SIAM J. Comput.* 37, 5 (2008), 1387–1400. https://doi.org/10.1137/S0097539704442416

[10] Arkadev Chattopadhyay, Yuval Filmus, Sajin Koroth, Or Meir, and Toniann Pitassi. 2019. Query-To-Communication Lifting for BPP Using Inner Product. In *Proceedings of the Forty-Sixth International Colloquium on Automata, Languages and Programming* (ICALP) *(LIPIcs, Vol. 132)*. 35:1–35:15. https://doi.org/10.4230/LIPIcs.ICALP.2019.35

[11] Eshan Chattopadhyay, Pooya Hatami, Kaave Hosseini, and Shachar Lovett. 2018. Pseudorandom Generators from Polarizing Random Walks. In *Proceedings of the Thirty-Third Annual IEEE Conference on Computational Complexity* (CCC), Vol. 102. 1:1–1:21. https://doi.org/10.4230/LIPIcs.CCC.2018.1

[12] Eshan Chattopadhyay, Pooya Hatami, Omer Reingold, and Avishay Tal. 2018. Improved pseudorandomness for unordered branching programs through local monotonicity. In *Proceedings of the Fiftieth Annual ACM Symposium on Theory of Computing* (STOC). 363–375. https://doi.org/10.1145/3188745.3188800

[13] David Deutsch and Richard Jozsa. 1992. Rapid solution of problems by quantum computation. *Proc. R. Soc. Lond. A* 439 (1992), 553–558. https://doi.org/10.1098/rspa.1992.0167

[14] Dmitry Gavinsky. 2020. Entangled Simultaneity Versus Classical Interactivity in Communication Complexity. *IEEE Trans. Inf. Theory* 66, 7 (2020), 4641–4651. https://doi.org/10.1109/TIT.2020.2976074

[15] Parikshit Gopalan, Rocco A. Servedio, and Avi Wigderson. 2016. Degree and Sensitivity: Tails of Two Distributions. In *Proceedings of the Thirty-First Annual IEEE Conference on Computational Complexity* (CCC), Vol. 50. 13:1–13:23. https://doi.org/10.4230/LIPIcs.CCC.2016.13

[16] Lov K. Grover. 1996. A Fast Quantum Mechanical Algorithm for Database Search. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing* (STOC). 212–219. https://doi.org/10.1145/237814.237866

[17] Hao Huang. 2019. Induced subgraphs of hypercubes and a proof of the Sensitivity Conjecture. *Annals of Mathematics* 190, 3 (2019), 949–955. https://doi.org/10.4007/annals.2019.190.3.6

[18] Stasys Jukna. 2011. *Extremal Combinatorics with Applications in Computer Science* (2nd ed.). Springer-Verlag Berlin Heidelberg. https://doi.org/10.1007/978-3-642-17364-6

[19] Ryan O'Donnell. 2014. *Analysis of Boolean Functions.* Cambridge University Press.

[20] Ryan O'Donnell and Rocco A. Servedio. 2007. Learning Monotone Decision Trees in Polynomial Time. *SIAM J. Comput.* 37, 3 (2007), 827–844. https://doi.org/10.1137/060669309

[21] Ran Raz. 1999. Exponential Separation of Quantum and Classical Communication Complexity. In *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing* (STOC). 358–367. https://doi.org/10.1145/301250.301343

[22] Oded Regev and Bo'az Klartag. 2011. Quantum one-way communication can be exponentially stronger than classical communication. In *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing* (STOC). 31–40. https://doi.org/10.1145/1993636.1993642

[23] Alexander A. Sherstov, Andrey Storozhenko, and Pei Wu. 2020. An Optimal Separation of Randomized and Quantum Query Complexity. In *Electronic Colloquium on Computational Complexity (ECCC).* Report TR20-128.

[24] Peter W. Shor. 1997. Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. *SIAM J. Comput.* 26, 5 (1997), 1484–1509. https://doi.org/10.1137/S0097539795293172

[25] Daniel R. Simon. 1997. On the Power of Quantum Computation. *SIAM J. Comput.* 26, 5 (1997), 1474–1483. https://doi.org/10.1137/S0097539796298637

[26] Thomas Steinke, Salil P. Vadhan, and Andrew Wan. 2017. Pseudorandomness and Fourier-Growth Bounds for Width-3 Branching Programs. *Theory Comput.* 13, 1 (2017), 1–50. https://doi.org/10.4086/toc.2017.v013a012

[27] Avishay Tal. 2017. Tight Bounds on the Fourier Spectrum of AC0. In *Proceedings of the Thirty-Second Annual IEEE Conference on Computational Complexity* (CCC), Vol. 79. 15:1–15:31. https://doi.org/10.4230/LIPIcs.CCC.2017.15

[28] Avishay Tal. 2020. Towards Optimal Separations between Quantum and Randomized Query Complexities. In *Proceedings of the Sixty-First Annual IEEE Symposium on Foundations of Computer Science* (FOCS). 228–239. https://doi.org/10.1109/FOCS46700.2020.00030

[29] Ronald de Wolf. 2001. *Quantum Computing and Communication Complexity.* Ph.D. Dissertation. University of Amsterdam.