Electronic Colloquium on Computational Complexity, Report No. 127 (2020)

k-Forrelation Optimally Separates Quantum and Classical Query Complexity

Nikhil Bansal[*] Makrand Sinha[†]

Abstract

Aaronson and Ambainis (SICOMP '18) showed that any partial function on N bits that can be computed with an advantage δ over a random guess by making q quantum queries, can also be computed classically with an advantage $\delta/2$ by a randomized decision tree making $O_q(N^{1-\frac{1}{2q}}\delta^{-2})$ queries. Moreover, they conjectured the k-Forrelation problem — a partial function that can be computed with $q = \lceil k/2 \rceil$ quantum queries — to be a suitable candidate for exhibiting such an extremal separation.

We prove their conjecture by showing a tight lower bound of $\widetilde{\Omega}_k(N^{1-1/k})$ for the randomized query complexity of k-Forrelation, where the advantage $\delta = 1/\text{polylog}^k(N)$ and $\widetilde{\Omega}_k$ hides $\text{polylog}^k(N)$ factors. Our proof relies on classical Gaussian tools, in particular, Gaussian interpolation and Gaussian integration by parts, and in fact, shows a more general statement, that to prove lower bounds for k-Forrelation against a family of functions, it suffices to bound the ℓ_1-weight of the Fourier coefficients at levels $k, 2k, 3k, \ldots, (k-1)k$ for functions in the family.

[*] CWI Amsterdam and TU Eindhoven, N.Bansal@cwi.nl. Supported by the NWO VICI grant 639.023.812.

[†] CWI Amsterdam, makrand@cwi.nl. Supported by the NWO VICI grant 639.023.812.

ISSN 1433-8092

1 Introduction

The last couple of decades have given us ample evidence to suggest that quantum computers can be exponentially more powerful in solving certain computational tasks than their classical counterparts. The *black-box* or *query* model offers a concrete setting to provably show such exponential speedups. In this model, a quantum algorithm has "black-box access" to the input and seeks to compute a function of the input while minimizing the number of queries. 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] or Shor's period-finding algorithm [Sho97], 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. In this case, the classical counterpart is also commonly known as a randomized decision tree. There are many connections between the settings of quantum and randomized query complexity and for more details, we refer the reader to the survey by Buhrman and de Wolf [BW02].

The above raises a natural question that was first asked by Buhrman, Fortnow, Newman and Röhrig [BFNR08]: what is the maximal possible separation between quantum and classical query complexities? Translating the results from slightly different query models to the setting where the queries return a bit of the input, Simon's problem [Sim97] and a work of Childs et al. [CCD+03] exhibited a separation of $O(\log^2 N)$ quantum queries vs $\widetilde{\Omega}(\sqrt{N})$ randomized queries for partial functions on N bits, while another work of de Beaudrap, Cleve and Watrous [BCW02] implied a 1 vs $\Omega(N^{1/4})$ separation. However, these works left open the possibility of a $O(1)$ vs $\Omega(N)$ separation, and towards answering this question, Aaronson and Ambainis [AA18] showed that for $q = O(1)$, any q-query quantum algorithm can be simulated by a randomized algorithm making $O(N^{1-\frac{1}{2q}})$ queries, thus ruling out the possibility of a $O(1)$ vs $\Omega(N)$ separation. In particular, they proved the following fundamental simulation result.

Theorem 1.1 ([AA18])**.** *Let $\mathcal{Q}$ be a quantum algorithm that makes q queries to an input $x \in \{\pm 1\}^N$. Then, with high probability, one can estimate $\mathbb{P}[\mathcal{Q}$ accepts $x]$ up to an additive δ factor by making $O(4^q N^{1-\frac{1}{2q}} \delta^{-2})$ classical randomized queries to x. Moreover, these queries are also non-adaptive.*

In the same paper, Aaronson and Ambainis showed that the (standard) Forrelation problem, exhibits a 1 vs $\widetilde{\Omega}(\sqrt{N})$ separation, improving upon a 1 vs $\Omega(N^{1/4})$ separation shown earlier by Aaronson [Aar10] where the standard Forrelation problem was introduced. Given the above theorem and ignoring polylog(N) factors, this is the maximal separation possible when $q = 1$.

[AA18] asked if Theorem 1.1 is also tight for any $q > 1$. If true, this would imply an $O(1)$ vs $\Omega(N^{1-\eta})$ separation where $\eta = O(1/q)$ could be made arbitrarily small. Towards this end, they suggested a natural generalization of the standard Forrelation problem, that they called k-Forrelation, which we introduce next in a slightly more general setting.

(δ, k)-Forrelation. Let $\mathsf{H} = \mathsf{H}_N$ denote the $N \times N$ Hadamard matrix where $N = 2^n$ for $n \in \mathbb{N}$ and H is normalized to have orthonormal columns, and hence operator norm 1. Let $k \geq 2$ be an integer and let $\underline{i} = (i_1, \cdots, i_k) \in [N]^k$, and $z := (z_1, \cdots, z_k) \in \{\pm 1\}^{kN}$. Define the function $\mathsf{forr}_k : \{\pm 1\}^{kN} \to \mathbb{R}$ as follows

$$\mathsf{forr}_k(z) = \frac{1}{N} \sum_{\underline{i} \in [N]^k} z_1(i_1) \cdot \mathsf{H}_{i_1, i_2} \cdot z_2(i_2) \cdot \mathsf{H}_{i_2, i_3} \cdots z_{k-1}(i_{k-1}) \cdot \mathsf{H}_{i_{k-1}, i_k} \cdot z_k(i_k). \tag{1.1}$$

Observe that this function can be written as the following quadratic form:

$$\mathsf{forr}_k(z) = \frac{1}{N} \cdot z_1^{\top} (\mathsf{H} \cdot \mathsf{Z}_2 \cdot \mathsf{H} \cdot \mathsf{Z}_3 \cdots \mathsf{H} \cdot \mathsf{Z}_{k-1} \cdot \mathsf{H}) z_k, \tag{1.2}$$

where $\mathsf{Z}_i = \mathsf{diag}(z_i)$ for $i \in \{2, \ldots, k-1\}$ is the diagonal matrix with $z_i \in \{\pm 1\}^N$ on the diagonal. From the above quadratic form description, it follows that $\mathsf{forr}_k(z) \in [-1, 1]$ always, since $z_1/\sqrt{N}$ and $z_k/\sqrt{N}$ are unit vectors, and the operator norm of the matrix appearing in the quadratic form is at most 1.

For a parameter $0 < \delta < 1$, the (δ, k)-*Forrelation function* is then defined in terms of forr_k as the following partial boolean function:

$$\mathsf{forr}_{\delta,k}(z) = \begin{cases} 1 & \text{if} \quad \mathsf{forr}_k(z) \geq \delta, \text{ and} \\ 0 & \text{if} \quad |\mathsf{forr}_k(z)| \leq \delta/2. \end{cases} \tag{1.3}$$

We overload the notation forr above to denote the real function forr_k, as well as the partial boolean function $\mathsf{forr}_{\delta,k}$, but the reader should not have any ambiguity as to what is meant. Note that the standard Forrelation promise problem of [AA18] is obtained by taking $\delta = 3/5$ and $k = 2$.

As already observed by [AA18], there is a simple and efficient quantum circuit that makes $\lceil k/2 \rceil$ queries and computes (δ, k)-Forrelation in the following manner.

Proposition 1.2 ([AA18])**.** *There exists a quantum circuit $\mathcal{Q}$ that makes $\lceil k/2 \rceil$ queries and uses $O(k \log N)$ gates, such that for any input $z \in \{\pm 1\}^{kN}$, it holds that $\mathbb{P}[\mathcal{Q}$ accepts $z] = \frac{1}{2}(1 + \mathsf{forr}_k(z))$.*

The above implies a $\delta/4$ gap between the acceptance probabilities on the 1-inputs and 0-inputs for (δ, k)-Forrelation. Standard tricks can then be used to show that with $\lceil k/2 \rceil$ quantum queries and a quantum circuit of $O(k \log N)$ size, one can compute (δ, k)-Forrelation with error at most $\frac{1}{2} - \delta/16$ on any input.

Combined with Theorem 1.1, this also shows that the (δ, k)-Forrelation function can be computed by making $O(2^k N^{1-1/k} \delta^{-2})$ classical randomized queries[1], even non-adaptively. For even values of k, this exactly matches the bound in Theorem 1.1 (upto $\mathrm{polylog}(kN)$ factors assuming $k = O(\log\log N)$) and Aaronson and Ambainis [AA18] proposed (δ, k)-Forrelation as a candidate for extremal separations between classical and quantum query complexities.

On the lower bound side, as mentioned before, Aaronson and Ambainis [AA18] showed that $\Omega(\sqrt{N}/\log N)$ classical queries are required for standard Forrelation. They also showed a slightly weaker lower bound of $\Omega(\sqrt{N}/\log^{7/2} N)$ for (δ, k)-Forrelation, for $\delta = 3/5$ and $k > 2$. One can improve this lower bound slightly by observing the following: in the quadratic form description (1.2) above, if we take $z_2, \cdots, z_{k-1}$ to be the all-one strings, and k is even, then (δ, k)-Forrelation reduces to the standard Forrelation as $\mathsf{H}^r = \mathsf{H}$ if r is an odd natural number. So the same $\Omega(\sqrt{N}/\log N)$ lower bound holds for $(\delta = 3/5, k)$-Forrelation as well, if k is even. Similarly, although not obvious, one can also design an input distribution achieving the same lower bound for odd k.

Thus, the current lower bounds for (δ, k)-Forrelation do not exhibit a better than $O(1)$ vs $\widetilde{\Omega}(\sqrt{N})$ separation, still leaving whether Theorem 1.1 is tight for $q > 1$ wide open.

Beyond $O(1)$ vs. $\widetilde{\Omega}(\sqrt{N})$ separation. Recently, motivated by this question, Tal [Tal19] considered a different variant of the (δ, k)-Forrelation problem, that he refers to as *Rorrelation*, to show a $\lceil k/2 \rceil$ vs $\widetilde{\Omega}(N^{2/3 - O(1/k)})$ separation. In particular, Tal shows that if one replaces the Hadamard matrix H in (1.1) and (1.3) by a random orthogonal matrix U, then to compute the resulting random partial function, one requires $\widetilde{\Omega}\left(N^{2(k-1)/(3k-1)}\right)$ classical queries with high probability for parameters $(\delta = 2^{-k}, k)$. Moreover, any such function can still be computed with $\lceil k/2 \rceil$ quantum queries, giving the $\lceil k/2 \rceil$ vs $\widetilde{\Omega}(N^{2/3 - O(1/k)})$ separation.

While this breaks the $\sqrt{N}$ barrier, the Rorrelation function is not explicit, and even though it is computable with a small number of quantum queries, the corresponding unitaries may not be efficiently implementable as a quantum circuit. This is in contrast to (δ, k)-Forrelation, where the resulting quantum query algorithms can also be efficiently implemented as a quantum circuit of polylogarithmic size. Tal's proof does not imply a better lower bound for (δ, k)-Forrelation than the $\widetilde{\Omega}(\sqrt{N})$ bound mentioned before, as it relies on various strong properties of random orthogonal matrices that the Hadamard matrix does not satisfy.

[1] For even k this follows from the statement of Theorem 1.1 as $\lceil k/2 \rceil = k/2$. The bound also holds for odd k as the proof of Theorem 1.1 in fact shows that any bounded *block-multilinear* polynomial of degree d can be approximated up to δ additive error with $O(2^d N^{1-1/d} \delta^{-2})$ randomized queries, and forr_k is a degree-k block-multilinear polynomial for all k. The connection with query complexity comes in as the acceptance probability of any q-query quantum algorithm can be written as a degree-$2q$ block-multilinear polynomial.

1.1 Our Results

In this work, we confirm the conjecture of Aaronson and Ambainis that (δ, k)-Forrelation does exhibit an extremal separation between classical and quantum query complexities by proving the following lower bound.

Theorem 1.3. *Let $k \geq 2$, $\epsilon = (64k^2\log(kN))^{-1}$, and set $\delta = \epsilon^k$. Then, any randomized decision tree that computes (δ, k)-Forrelation with error at most $\frac{1}{2} - \frac{\gamma}{2}$, must make at least the following number of queries,*

$$\Omega\left(\frac{N^{1-1/k}}{k^8\log(kN)} \cdot \frac{\gamma^2}{\log(1/\delta)}\right).$$

Note that for an even $k = O(1)$ and an advantage $\gamma = \delta/16$, the above lower bound is $\widetilde{\Omega}(N^{1-1/k})$ and it matches the upper bound for $(\delta = \epsilon^k, k)$-Forrelation implied by Theorem 1.1, up to a $\operatorname{polylog}(kN)$ factor. The bound is also tight for odd k, as mentioned before. We remark that the choice of $\epsilon = O(1/\log(kN))$ is due to technical reasons. One can try to take $\epsilon = \Omega(1)$ by a slight modification to our proof, but the computations get quite tedious. It will be interesting to see if this could be made to work.

The previous statement gives a lower bound for randomized algorithms that have a $\Theta(\delta)$ advantage, since we wish to compare it to the advantage of the $\lceil k/2 \rceil$-query quantum algorithm which has a success probability of $1/2 + \Theta(\delta)$. If one wants a success probability of at least $2/3$, then the quantum query complexity of (δ, k)-Forrelation becomes $O(k \cdot \delta^{-2}) = \operatorname{polylog}^k(kN)$ by using standard amplification tricks. This gives us that there exists an explicit partial boolean function on $M = kN$ bits that can be computed with error at most $1/3$ by quantum circuits of $\operatorname{polylog}^k(M)$ size, making $\operatorname{polylog}^k(M)$ queries, but requires $M^{1-\eta}$ randomized queries where $\eta = \Theta(1/k)$. If we set $k = O(\log\log N)$, then $\eta = o(1)$ while the number of quantum queries and the size of the quantum circuit is quasi-polylogarithmic in M. More precisely, we have the following.

Corollary 1.4. *Let $k \geq 2$, $\epsilon = (64k^2\log(kN))^{-1}$ and $\delta = \epsilon^k$. Then, there exists a quantum circuit with $\operatorname{polylog}^k(kN)$ gates, making $\operatorname{polylog}^k(kN)$ queries that computes (δ, k)-Forrelation with error at most $1/3$. On the other hand, any randomized decision tree that computes (δ, k)-Forrelation with error at most $1/3$, needs $\Omega\left(\frac{N^{1-1/k}}{k^8\log(kN)\log(1/\delta)}\right) = \Omega\left(\frac{N^{1-1/k}}{k^{10}\log^2(kN)}\right)$ queries.*

We remark that our proof also works even if one replaces the Hadamard matrix H in (1.1) and (1.3) by an arbitrary orthogonal matrix U where all entries are $\widetilde{O}(N^{-1/2})$ in magnitude. In particular, the lower bounds given above also hold for Rorrelation as all entries of a random orthogonal matrix are $O((N/\log N)^{-1/2})$ with high probability.

Separation for Total Boolean Functions. Our results also imply an improved separation for total boolean functions. Let $Q(\mathsf{f})$ (resp. $R(\mathsf{f})$) denote the minimum number of queries made by a quantum (resp. randomized) algorithm to compute a (partial or total) boolean function f with probability at least $2/3$.

Then, the results of Aaronson, Ben-David and Kothari [ABK16] imply that an $M^{o(1)}$ vs $M^{1-o(1)}$ separation between the quantum and randomized query complexity of a partial boolean function on M bits implies the existence of a *total* boolean function with cubic separation between the two measures. Combined with our results, this yields the following corollary.

Corollary 1.5. *There exists a total boolean function f for which $R(\mathsf{f}) \geq Q(\mathsf{f})^{3-o(1)}$.*

The recent work of Aaronson, Ben-David, Kothari and Tal [ABKT20] conjectures that for any total boolean function f, it always holds that $R(\mathsf{f}) = O(Q(\mathsf{f})^3)$, so if true, the above separation is optimal up to $o(1)$ factors in the exponent. The current best upper bound is a 4th power relationship which holds even for deterministic query algorithms: denoting by $D(\mathsf{f})$ the deterministic query complexity of f, [ABKT20] prove that $D(\mathsf{f}) = O(Q(\mathsf{f})^4)$. The above is tight for deterministic query algorithms due to an example of Ambainis et al. [ABB+17].

1.2 Overview and Techniques

Our proof of Theorem 1.3 is based on classical Gaussian tools, and builds on the stochastic calculus approach of Raz and Tal [RT19] for their breakthrough result on oracle separation between BQP and PH (see also the simplification of the results of [RT19] by Wu [Wu20]).

In fact, the input distribution that [RT19] use is a slight variant of the distribution used for standard Forrelation ($k = 2$) by [AA18]. However, as also noted by [Tal19], it is unclear how to use stochastic calculus already for $k = 3$, as the hard input distribution for randomized query algorithms has a non-linear structure involving the product of two Gaussians (we elaborate more on this later).

To get around this, our proof relies on using multilinearity of functions on the discrete cube and the properties of the underlying input distribution in a careful way, together with additional tools such as Gaussian interpolation and Gaussian integration by parts. In this overview, we focus on the special case of $k = 3$, which already suffices to illustrate the main difficulties in extending the previous approaches to prove lower bounds for (δ, k)-Forrelation.

The case of $k = 3$. In this case, for $\underline{i} = (i_1, i_2, i_3) \in [N]^3$ and $z = (z_1, z_2, z_3) \in \mathbb{R}^{3N}$, we have

$$\mathsf{forr}_3(z) = \frac{1}{N} \sum_{\underline{i} \in [N]^3} z_1(i_1) \cdot \mathsf{H}_{i_1,i_2} \cdot z_2(i_2) \cdot \mathsf{H}_{i_2,i_3} \cdot z_3(i_3).$$

It is not hard to see that the uniform distribution on $\{\pm 1\}^{3N}$ is mostly supported on 0-inputs for $\mathsf{forr}_3(z)$. We will give a distribution $p_1(Z)$ on $\{\pm 1\}^{3N}$ — a variant of the distribution considered in [RT19, Tal19] — that is mostly supported on 1-inputs.

Given an arbitrary randomized decision tree making d queries, let $f(z)$ be the acceptance probability of the decision tree on input z. To prove a lower bound it suffices to show that for any such f, the distinguishing advantage $|\mathbb{E}_{p_1}[f(Z)] - f(0)|$ is small, as $f(0)$ is exactly the average acceptance probability under the uniform distribution.

The distribution $p_1(Z)$. Consider the $2N \times 2N$ covariance matrix $\Sigma = \epsilon \begin{pmatrix} \mathsf{I}_N & \mathsf{H}_N \\ \mathsf{H}_N & \mathsf{I}_N \end{pmatrix}$ with $\epsilon = \Theta(1/\log N)$. A random Gaussian vector distributed as $\mathcal{N}(0, \Sigma)$ will typically lie inside the cube $[-1/2, 1/2]^{2N}$ as the variance of each coordinate is $O(1/\log N)$, and in this overview we assume that this is always the case, to avoid technicalities that can be dealt with truncating and bounding the error separately. Then, $p_1(Z)$ is the following distribution: Take two independent $2N$-dimensional Gaussian vectors $G = (U_1, V_1)$ and $B = (U_2, V_2)$ distributed as $\mathcal{N}(0, \Sigma)$ and obtain a vector $Z \in \{\pm 1\}^{3N}$ by rounding each coordinate independently to ± 1 with bias given by $(U_1, U_2 \odot V_1, V_2) \in [-1/2, 1/2]^{3N}$. Here $\odot$ denotes the Hadamard product[2] of two vectors. In other words, for $i \in [N]$,

$$\mathbb{E}_{p_1}[Z_1(i) \mid G, B] = U_1(i) \text{ and } \mathbb{E}_{p_1}[Z_2(i) \mid G, B] = U_2(i)V_1(i) \text{ and } \mathbb{E}_{p_1}[Z_3(i) \mid G, B] = V_2(i). \tag{1.4}$$

Therefore, we have

$$\mathbb{E}_{p_1}[\mathsf{forr}_3(Z)] = \frac{1}{N} \sum_{\underline{i} \in [N]^3} \mathbb{E}[U_1(i_1) \cdot \mathsf{H}_{i_1,i_2} \cdot V_1(i_2) G_2(i_2) \cdot \mathsf{H}_{i_2,i_3} \cdot V_2(i_3)] = \frac{\epsilon^2}{N} \sum_{\underline{i} \in [N]^3} \mathsf{H}^2_{i_1,i_2} \mathsf{H}^2_{i_2,i_3} = \Theta\left(\frac{1}{\log^2 N}\right),$$

as $\mathbb{E}[U_1(i)V_1(j)] = \mathbb{E}[U_2(i)V_2(j)] = \epsilon \cdot \mathsf{H}_{i,j}$, and since each entry of H is $\pm\frac{1}{\sqrt{N}}$ and $\epsilon = \Theta(1/\log N)$.

Extending f from $\{\pm 1\}^{3N}$ to a function from $\mathbb{R}^{3N}$ to $\mathbb{R}$, by identifying it with its Fourier expansion, and using the multilinearity of f and the equalities in (1.4), our task then reduces to showing that

$$\left|\mathbb{E}_{p_1}[f(Z)] - f(0)\right| = \left|\mathbb{E}[f(U_1, U_2 \odot V_1, V_2)] - f(0)\right| \ll 1/\log^2 N. \tag{1.5}$$

[2] For $u, v \in \mathbb{R}^m$, the Hadamard product is the vector $u \odot v \in \mathbb{R}^m$ defined as $u \odot v = (u(1) \cdot v(1), \cdots, u(m) \cdot v(m))$.

Previous approaches and their limitations. This is the starting point of all[3] previous approaches to bounding the above, which essentially proceed in the following two ways.

(a) Bounding all moments and Fourier weight of all levels. As $f(z) = \sum_{S\subseteq[3N]} \hat{f}(S)\chi_S(z)$ where $\{\chi_S(z)\}_{S\subseteq[3N]}$ are Fourier characters, one can bound

$$\left|\mathbb{E}[f(U_1, U_2 \odot V_1, V_2)] - f(0)\right| \leq \sum_{\ell=1}^{d} \mathsf{wt}_\ell(f) \cdot \max_{|S|=\ell} \left|\mathbb{E}[\chi_S(U_1, U_2 \odot V_1, V_2)]\right|,$$

writing $\mathsf{wt}_\ell(f) = \sum_{|S|=\ell} |\hat{f}(S)|$, as the ℓ_1-weight of the Fourier coefficients at level ℓ.

This approach needs a bound on the Fourier weight $\mathsf{wt}_\ell(f)$ for all levels $\ell \leq d$, as well as a bound on all the moments $|\mathbb{E}[\chi_S(U_1, U_2 \odot V_1, V_2)]|$, and consequently suffers from two drawbacks. First, the currently known bounds on wt_ℓ for decision trees degrade as ℓ gets large — [Tal19] shows that if f is computable by a randomized decision tree of depth d, then $\mathsf{wt}_\ell(f) \leq \widetilde{O}(d)^{\ell/2}$, which becomes weaker than the trivial bound of $\binom{d}{\ell}$ when $\ell \gg \sqrt{d}$. For this reason the bound of [Tal19] for Rorrelation does not go beyond $\widetilde{\Omega}(N^{2/3-O(1/k)})$.

Second, the moments can be very large for the Hadamard matrix (e.g. due to very large submatrices with all $1/\sqrt{N}$ entries). This is not an issue if a random orthogonal matrix is used instead (which allows [Tal19] to go beyond $N^{1/2}$ for Rorrelation). Another limitation is that using a worst case bound for the moment given by each Fourier character does not exploit the non-trivial cancellations that can occur for various terms in the sum. In fact, it is not even clear how to obtain the $\widetilde{\Omega}(N^{1/2})$ bound for $k = 2$ using this approach.

(b) Stochastic Calculus/Gaussian Interpolation. The second approach is based on utilizing the special properties of Gaussians and using tools from stochastic calculus [RT19, Wu20]. In this paper, we describe an alternate approach using the classical method of Gaussian interpolation which can also be recovered by stochastic (Itô) calculus. Gaussian interpolation is a way to continuously interpolate between jointly Gaussian random variables with different covariance structures. By choosing a suitable path to interpolate and controlling the derivatives along this path, one can compute functions of Gaussians with a more complicated covariance structure in terms of an easier one. Talagrand [Tal11] dubs this the *smart path method* to stress the important of choosing the right path.

In particular, let $G \in \mathbb{R}^m$ be a multivariate Gaussian and for an interpolation parameter $t \in (0, 1)$, define $\mathbf{G}(t) = \sqrt{t} \cdot G$. Then, the Gaussian interpolation formula (see Section 2.1) implies that for any *reasonable* function $h : \mathbb{R}^m \to \mathbb{R}$ one has

$$\mathbb{E}[h(G)] - h(0) = \int_0^1 \frac{d}{dt}\left(\mathbb{E}[h(\mathbf{G}(t))\right) dt = \frac{1}{2}\sum_{ij} \mathbb{E}[G_iG_j] \int_0^1 \mathbb{E}\left[\partial_{ij}h(\mathbf{G}(t))\right] dt. \tag{1.6}$$

in terms of the covariance of G and the second derivatives ∂_{ij} of h.

Note that if h is a multilinear polynomial, then $\partial_{ij}h(0) = \hat{h}(ij)$ if $i \neq j$ while $\partial_{ii}h$ is identically zero. The right-hand side above involves partial derivatives at arbitrary points $\mathbf{G}(t)$, but these can be reduced to derivatives at 0 (and hence level-two Fourier coefficients $\hat{h}(ij)$) by a clever random restriction. In particular, the derivative $\partial_{ij}h(\mu)$ at any $\mu \in [-1/2, 1/2]^m$ can be interpreted as a Fourier coefficient with respect to a biased product measure (details given later). Thus, this approach only requires a bound on the level-two weight $\mathsf{wt}_2(f)$, and works very nicely for $k = 2$, as in that case our function is a multilinear function of a Gaussian.

However for $k = 3$, as also noted by [Tal19], it is not immediately clear how to use the interpolation approach to bound the expression in (1.5), as it involves a product of Gaussians. In particular, the second block of coordinates consists of products of coordinates of Gaussians U_2 and V_1.

[3] We remark that the original approach of [AA18] does not fit in this framework and it is not clear how to generalize it either for $k > 2$.

1.2.1 Our Approach

Our main insight is that the advantage of f in (1.5) can essentially be bounded in terms of the third and sixth level Fourier weight of f (see (1.8) for the precise statement). More generally for any $k \geq 3$, the advantage of f can be bounded in terms of the Fourier weight of f at levels $k, 2k, \ldots, (k-1)k$.

To show this, we use Gaussian interpolation as in (1.6). In particular, for $k = 3$, given that our vector is of the form $(U_1, U_2 \odot V_1, V_2)$ and f is a multilinear polynomial, we can treat the function h in (1.6) as a function of the $4N$-dimensional Gaussian vector (U_1, U_2, V_1, V_2). Similarly, for an arbitrary k, using a suitable generalization of the distribution $p_1(Z)$, we get a function h of a $2(k-1)N$-dimensional Gaussian vector. The resulting expression in (1.6) is then a $k-1$ dimensional integral, which leads to partial derivatives of order $2k-2$ instead of ∂_{ij} in (1.6) above. However, due to the interactions between the variables of U_i and V_{i-1} (an issue which does not arise for $k = 2$), the partial deriatives with respect to U_i and V_{i-1} do not necessarily correspond to derivatives of f (with respect to its coordinates), and a key technical idea is to use Gaussian integration by parts to relate them. In particular, the order $2k-2$ derivatives of h can be related to order $k, 2k, \ldots, (k-1)k$ derivatives of f.

We remark that a recent work of Girish, Raz and Zhan [GRZ20] used a similar multi-dimensional stochastic walk to prove a lower bound for a different setting: they considered the partial function obtained by taking an XOR of multiple copies of the standard Forrelation problem, and their main focus was to prove a lower bound for quasipolynomially small advantage. The analysis for this setting is closer to the previously mentioned approaches of [RT19, Wu20] for the standard Forrelation problem. In particular, the technical challenges that arise while trying to prove a better than $\widetilde{\Omega}(\sqrt{N})$ lower bound for k-Forrelation for $k > 2$ do not arise in that case.

The case of $k = 3$. We explain the idea for $k = 3$ first, which is quite a bit simpler, and then sketch the additional ideas needed for higher k. We will crucially leverage the multilinearity of the function f and the specific structure of the random vector $(U_1, U_2 \odot V_1, V_2) \in \mathbb{R}^{3N}$. In particular, let $S = S_1 \sqcup S_2 \sqcup S_3$ where S_r for $r \in [3]$ is the projection of the subset on the r^{th} block of coordinates and $\sqcup$ denotes the disjoint union of the sets. Consider the monomial $\chi_S(z)$ in the multilinear representation of f. Using the multiplicativity of the characters, we have that

$$\chi_S(U_1, U_2 \odot V_1, V_2) = \chi_{S_1}(U_1)\chi_{S_2}(U_2) \cdot \chi_{S_2}(V_1)\chi_{S_3}(V_2).$$

Our starting point is that as $G = (U_1, V_1)$ and $B = (U_2, V_2)$ are independent, one can interpolate them separately, which leads to a two-dimensional integral in (1.6), and the integrand on the right side ranges over the following derivatives

$$\begin{aligned} &\mathbb{E}\left[\frac{\partial}{\partial u_1(i_1)\partial v_1(j_2)}\chi_{S_1}(\mathbf{U}_1(t_1))\chi_{S_2}(\mathbf{V}_1(t_1))\right]\mathbb{E}\left[\frac{\partial}{\partial u_2(i_2)\partial v_2(j_3)}\chi_{S_2}(\mathbf{U}_2(t_2))\chi_{S_3}(\mathbf{V}_2(t_2))\right] \\ =&\mathbb{E}\left[\chi_{S_1\backslash i_1}(\mathbf{U}_1(t_1))\chi_{S_2\backslash j_2}(\mathbf{V}_1(t_1))\right]\cdot\mathbb{E}\left[\chi_{S_2\backslash i_2}(\mathbf{U}_2(t_2))\chi_{S_3\backslash j_3}(\mathbf{V}_2(t_2))\right] \\ =&\mathbb{E}\left[\chi_{S_1\backslash i_1}(\mathbf{U}_1(t_1))\chi_{S_2\backslash j_2}(\mathbf{V}_1(t_1))\cdot\chi_{S_2\backslash i_2}(\mathbf{U}_2(t_2))\chi_{S_3\backslash j_3}(\mathbf{V}_2(t_2))\right], \end{aligned} \tag{1.7}$$

where $(i_1, i_2) \in S_1 \times S_2$, and $(j_2, j_3) \in S_2 \times S_3$, and $t_1, t_2 \in (0, 1)$ are interpolation parameters which we will drop from the notation henceforth.

The main difference now from the $k = 2$ case is that because of the presence of products $U_2 \odot V_1$, the above derivatives can not be interpreted in general as derivatives $\frac{\partial f}{\partial z_A}(z)$ evaluated at $(\mathbf{U}_1, \mathbf{U}_2 \odot \mathbf{V}_1, \mathbf{V}_2)$.

Let us consider this more closely. Suppose that $i_2 = j_2$. In this case, (1.7) becomes

$$\mathbb{E}\left[\chi_{S_1\backslash i_1}(\mathbf{U}_1)\cdot\chi_{S_2\backslash j_2}(\mathbf{U}_2 \odot \mathbf{V}_1)\cdot\chi_{S_3\backslash j_3}(\mathbf{V}_2)\right],$$

which corresponds to a third derivative of $\chi_S(z)$ evaluated at $z = (\mathbf{U}_1, \mathbf{U}_2 \odot \mathbf{V}_1, \mathbf{V}_2)$.

However, if $i_2 \neq j_2$, then the term in (1.7) does not correspond to a derivative of $f(z)$ with respect to z. To handle this, we note that $\chi_{S_2\setminus j_2}(\mathbf{V}_1)\cdot\chi_{S_{2\setminus i_2}}(\mathbf{U}_2)$ can be written as $\chi_{S_2\setminus\{i_2,j_2\}}(\mathbf{U}_2\odot\mathbf{V}_1)\cdot\mathbf{U}_2(j_2)\cdot\mathbf{V}_1(i_1)$, and hence (1.7) becomes

$$\mathbb{E}\left[\chi_{S_1\setminus i_1}(\mathbf{U}_1)\chi_{S_2\setminus\{i_2,j_2\}}(\mathbf{U}_2\odot\mathbf{V}_1)\chi_{S_3\setminus j_3}(\mathbf{V}_2)\cdot\mathbf{U}_2(j_2)\cdot\mathbf{V}_1(i_1)\right],$$

In particular, the term in the expectation corresponds to the derivative of $\chi_S(\mathbf{U}_1,\mathbf{U}_2\odot\mathbf{V}_1,\mathbf{V}_2)$ with respect to $J=\{i_1,i_2,j_2,j_3\}$ times the variables $\mathbf{U}_2(j_2)$ and $\mathbf{V}_1(i_2)$. However, this exactly fits the form required to use the Gaussian integration by parts formula (see Section 2.1). In particular, one can trade off the factors $\mathbf{U}_2(j_2)$ and $\mathbf{V}_1(i_2)$ for one additional derivative each, giving us the sixth order derivatives for χ_S. Both the cases above eventually allow us to bound the function in terms of Fourier weight of f at levels three and six.

To state the bound we obtain more formally, for $\mu\in[-1/2,1/2]^{3N}$, consider the product measure on $\{\pm1\}^{3N}$ where the i-th bit is 1 with probability $(1+\mu_i)/2$ and -1 with probability $(1-\mu_i)/2$, so that its bias is exactly μ_i. Define the level-ℓ Fourier weight with respect to bias μ as $\mathsf{wt}^{\mu}_{\ell}(f)=\sum_{|S|=\ell}|\hat{f}^{\mu}(S)|$, where $\hat{f}^{\mu}(S)$ is the Fourier coefficient with respect to the biased product measure above (see Section 2.2 for a formal definition). Then, we show the following key result towards bounding (1.5).

$$\left|\mathbb{E}[f(U_1,U_2\odot V_1,V_2)]-f(0)\right|\lesssim\sup_{\mu\in[-1/2,1/2]^{3N}}\frac{\epsilon}{N}\cdot\mathsf{wt}^{\mu}_{3}(f)+\frac{\epsilon^2}{N^2}\cdot\mathsf{wt}^{\mu}_{6}(f). \tag{1.8}$$

By a random-restriction argument similar to that in previous works, the level-ℓ Fourier weight for a decision tree with respect to biased measures is essentially the same as the Fourier weight with respect to the uniform measure (see Corollary 3.5 later) and hence at most $\widetilde{O}(d)^{\ell/2}$ by the bounds in [Tal19].

Plugging these bounds in (1.8) above, yields that for a depth-d randomized decision tree, the advantage is at most

$$\left|\mathbb{E}[f(U_1,U_2\odot V_1,V_2)]-f(0)\right|\leq\frac{\epsilon}{N}\cdot\widetilde{O}(d)^{3/2}+\left(\frac{\epsilon}{N}\cdot\widetilde{O}(d)^{3/2}\right)^2,$$

which is small for $d\ll N^{2/3}$. This gives the optimal bound for $(\delta,k=3)$-Forrelation, where $\delta=\Theta(1/\log^2 N)$.

Arbitrary k. For $k>3$, there is an additional complication that is not apparent in the case of $k=3$. In this case, a suitable generalization of the distribution $p_1(Z)$ involves $k-1$ independent $2N$-dimensional Gaussian vectors (U_κ,V_κ), for $\kappa\in[k-1]$ distributed as $\mathcal{N}(0,\Sigma)$. Moreover, there are $k-2$ blocks of the form $U_\kappa\odot V_{\kappa-1}$ for $\kappa\in\{2,\ldots,k-1\}$ (see Section 3 for the exact form). Due to this, when we apply Gaussian integration by parts to trade off the (unmatched) factors $U_\kappa(i)$ and $V_\kappa(j)$ with extra derivatives, this can lead to several more additional factors.

For example, suppose we apply Gaussian integration by parts to remove the factor $U_2(i)$, then since $U_2(i)$ is correlated with various $V_2(j)$ and each $V_2(j)$ appears together with a $U_3(j)$ in $V_2\odot U_3$, upon differentiating with respect to variables in V_2, this leads to multiple new terms with factors $U_3(j)$. Apriori, it is not obvious if applying Gaussian integration by parts leads to any progress. However, viewing this dynamics as a branching process and exploiting the multilinearity of the function f and the specific structure of the distribution $p_1(Z)$, we can show using a careful counting argument, that this process eventually terminates without giving too many higher order derivative terms.

In particular, even though the initial terms after the Gaussian interpolation step involve derivatives of order at most $2k-2$, we show that the final derivatives obtained after applying all the Gaussian integration by parts steps are of order $k,2k,\ldots,(k-1)k$. This allows us to show an overall bound on the advantage of f, in terms of the Fourier weight of f at levels $k,2k,3k,\ldots,(k-1)k$ where the relative contribution of the higher level weights gets progressively smaller. In the end, plugging in the bounds on the Fourier weight, we show that for an arbitrary k, the advantage a randomized depth-d decision tree has is at most

$$\left|\mathbb{E}_{p_1}[f(Z)]-f(0)\right|\;\leq\;\sum_{m=1}^{k-1}\left(\frac{\epsilon}{N}\right)^{m(k-1)/2}\cdot\widetilde{O}(d)^{mk/2}\;=\;\sum_{m=1}^{k-1}\left(\left(\frac{\epsilon}{N}\right)^{1-1/k}\cdot\widetilde{O}(d)\right)^{mk/2},$$

which is negligible if $d\ll N^{1-1/k}$. This gives the result for general k.

1.2.2 Organization

The rest of the paper is organized as follows. We introduce the notation and basic preliminaries in Section 2. Section 3 gives the input distribution, shows that the chosen input distribution has a large support on the 1 and 0 inputs of (δ, k)-Forrelation, and also gives a formal outline of the main proof. Section 4 shows that one can switch between the continuous and the discrete settings up to a small error. Section 5 contains the proof of the lower bound on randomized query algorithms.

2 Preliminaries

Notation. Throughout this paper, log denotes the natural logarithm unless the base is explicitly mentioned. We use $[k]$ to denote the set $\{1, 2, \ldots, k\}$. For a singleton set $\{x\}$, we sometimes write x for brevity. The set of natural numbers including zero is denoted by $\mathbb{N}_0$. Matrices are denoted by capital serif fonts (e.g. A).

For a random vector (or bit-string) z in $\mathbb{R}^n$, we will use z_i or $z(i)$ to denote the i-th coordinate of z, depending on whether we need to use the subscript for another index. If $z \in \mathbb{R}^{kn}$, then we will write $z = (z_1, \ldots, z_k)$ where $(z_\kappa)_{\kappa\in[k]}$ are vectors in $\mathbb{R}^n$ to denote the projections on the coordinates $\{(\kappa-1)n, \ldots, \kappa n\}$ — in this case, we will explicitly mention that $(z_\kappa)_{\kappa\in[k]}$ are vectors so that there is no ambiguity that z_κ refers to a coordinate of z. The operator and Frobenius norms of a matrix M are denoted by $\|\mathsf{M}\|_{\mathsf{op}}$ and $\|\mathsf{M}\|_F$.

Random variables are denoted by capital letters (e.g. A) and values they attain are denoted by lower-case letters possibly with subscripts and superscripts (e.g. a, a_1, a', etc.). Events in a probability space will be denoted by script letters (e.g. $\mathcal{B}$). We use $\mathbf{1}_{\mathcal{B}}$ or $\mathbf{1}[\mathcal{B}]$ to denote the indicator random variable for the event $\mathcal{B}$. Given a random variable X in a probability space p, we write $p(X)$ to denote the distribution of X in the probability space. For random variables X, Y, we write $p(X, Y)$ to denote the joint distribution and $p(X)$ to denote the marginal distribution. We write $p(\mathcal{B})$ to denote the probability of the event $\mathcal{B}$. For $\lambda \in [0, 1]$, we use $\lambda p(X) + (1 - \lambda)p'(X)$ to denote the convex combination of the two distributions, where the random variable X is sampled from $p(X)$ with probability λ, and from $p'(X)$ with probability $1 - \lambda$.

For a real valued function f, we write $\mathbb{E}_p[f(X)]$ to denote the expectation of the random variable $f(X)$ where X is in the probability space p. Similarly, $\mathbb{E}_p[f(X) \mid Y]$ denotes the conditional expectation of $f(X)$ with respect to Y. If the probability space p is clear from the context, we simply write $\mathbb{E}[f(X)]$ and $\mathbb{E}[f(X) \mid Y]$. We use $\mathcal{N}(0, \sigma^2)$ to denote a Gaussian random variable in $\mathbb{R}$ with mean zero and variance σ^2. For a positive semi-definite matrix $\mathsf{\Sigma} \in \mathbb{R}^{m\times m}$, we write $\mathcal{N}(0, \mathsf{\Sigma})$ to denote a centered (mean-zero) Gaussian random variable in $\mathbb{R}^m$ with covariance $\mathsf{\Sigma}$. We call an m-dimensional Gaussian standard, if $\mathsf{\Sigma}$ is the identity matrix I_m.

2.1 Gaussian Tools

Let us denote the density and cumulative distribution function for the standard Gaussian $\mathcal{N}(0, 1)$ by

$$\gamma(s) = \frac{1}{\sqrt{2\pi}} e^{-s^2/2} \quad \text{and} \quad \Phi(s) = \int_{-\infty}^{s} \gamma(t)dt.$$

The following estimate is standard.

Proposition 2.1 (Gaussian Concentration)**.** *For any $a > 0$, we have $1 - \Phi(a) \le e^{-a^2/2}$. In particular, if G is $\mathcal{N}(0, \sigma^2)$, then its density at s is $\sigma^{-1}\gamma(s/\sigma)$, and so for all $a > 0$,*

$$\mathbb{P}[|G| \ge a] = 2\int_a^\infty \sigma^{-1}\gamma(s/\sigma)ds = 2\int_{a/\sigma}^\infty \gamma(t)dt = 2(1 - \Phi(a/\sigma)) \le 2e^{-a^2/2\sigma^2}.$$

Proposition 2.2. *For any $a > 0$, it holds that $\int_a^\infty s^2\gamma(s)ds = a\gamma(a) + (1 - \Phi(a)) \le (a + 1)e^{-a^2/2}$.*

Proof. Since $\gamma'(s) = -s\gamma(s)$, integrating by parts, and using the proposition above

$$\int_a^\infty s^2\gamma(s)ds = \int_a^\infty -s\gamma'(s)ds = -s\gamma(s)\Big|_a^\infty + \int_a^\infty \gamma(s)ds = a\gamma(a) + (1-\Phi(a)) \leq (a+1)e^{-a^2/2}. \quad \blacksquare$$

Gaussian Interpolation and the Smart Path Method. We refer to Talagrand's book [Tal11] for a nice exposition, and in particular, §1.3 and Appendix A.4 there, for proofs of the lemmas given below.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be an infinitely differentiable function. We say that f is of moderate growth if all partial derivatives of f satisfy the following

$$\lim_{\|x\|\to\infty} \left|\partial_{\underline{i}} f(x)\right| \, e^{-a\|x\|^2} = 0 \text{ for every } \underline{i} = (i_1, \cdots, i_n) \in \mathbb{N}_0^n \text{ and } a \in \mathbb{R}_{>0}, \tag{2.1}$$

where $\partial_{\underline{i}}$ denotes the partial derivative $\dfrac{\partial}{\partial x_1^{i_1}} \cdots \dfrac{\partial}{\partial x_n^{i_n}}$ and $\|\cdot\|$ is the Euclidean norm. We remark that throughout this paper, we will only work with multilinear polynomials, which are always of moderate growth.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be of moderate growth and consider two centered jointly Gaussian random vectors G and B in $\mathbb{R}^n$. Let us define $\mathbf{G}(t) = (\mathbf{G}_i(t))_{i\leq n}$ where

$$\mathbf{G}_i(t) = \sqrt{t}\, G_i + \sqrt{1-t}\, B_i, \tag{2.2}$$

so that $G = \mathbf{G}(1)$ and $B = \mathbf{G}(0)$ and consider the function

$$\varphi(t) = \mathbb{E}[f(\mathbf{G}(t))]. \tag{2.3}$$

For clarity, we will use boldface font to refer to the interpolating Gaussian.

Lemma 2.3 (Gaussian Interpolation). *For $0 < t < 1$ we have*

$$\varphi'(t) = \frac{1}{2}\sum_{ij} \left(\mathbb{E}[G_iG_j] - \mathbb{E}[B_iB_j]\right) \, \mathbb{E}\left[\frac{\partial f}{\partial x_i \partial x_j}(\mathbf{G}(t))\right].$$

Choosing the covariance of B to be the all zero matrix, we have that $\mathbf{G}(t) = \sqrt{t}\, G$, and the following useful identity follows from the previous lemma by the fundamental theorem of calculus

$$\mathbb{E}[f(G)] - f(0) = \int_0^1 \varphi'(t)dt = \frac{1}{2}\sum_{ij} \mathbb{E}[G_iG_j] \int_0^1 \mathbb{E}\left[\frac{\partial f}{\partial x_i \partial x_j}(\mathbf{G}(t))\right] dt.$$

We remark that one can derive the same formula using Itô calculus.

Another important tool that we repeatedly use is the multivariate Gaussian integration by parts formula.

Lemma 2.4 (Gaussian Integration by Parts). *If $B, G_1, \ldots, G_n$ are real-valued random variables that are jointly Gaussian and $f : \mathbb{R}^n \to \mathbb{R}$ is of moderate growth, then*

$$\mathbb{E}[B \cdot f(G_1, \ldots, G_n)] = \sum_{i=1}^n \mathbb{E}[BG_i] \, \mathbb{E}\left[\frac{\partial f}{\partial x_i}(G_1, \ldots, G_n)\right].$$

Note that this formula replaces the expectation of the product of a Gaussian random variable with the function f, with a weighted sum of expectation of the derivatives of f.

The Gaussian integration by parts formula can be used to prove Lemma 2.3 and it turns out that it also uniquely characterizes the multivariate Gaussian distribution.

2.2 Fourier Analysis on the Discrete Cube

We give some facts from Fourier analysis on the discrete cube that we will need, and for more details we refer to the book [O'D14]. Every boolean function $f : \{\pm 1\}^m \to \mathbb{R}$ can be written uniquely as a sum of monomials $\chi_S(x) = \prod_{i \in S} x_i$,

$$f(x) = \sum_{S \subseteq [m]} \hat{f}(S)\chi_S(x), \tag{2.4}$$

where $\hat{f}(S) = \mathbb{E}_p[f(X)\chi_S(X)]$ is the Fourier coefficient with respect to the uniform measure p on $\{\pm 1\}^m$. The monomials $\chi_S(x) = \prod_{i \in S} x_i$ form an orthonormal basis for real-valued functions on $\{\pm 1\}^m$, called the *Fourier basis*.

Any function on $\{\pm 1\}^m$ can be extended to $\mathbb{R}^m$ by identifying it with the multilinear polynomial given by (2.4), which is also called the *harmonic extension* of f and is unique. We will denote the harmonic extension of f also by f and in general, we have the following identity by interpolating the values of f on the vertices of the discrete hypercube.

$$f(x) = \sum_{y \in \{\pm 1\}^m} w_x(y) f(y), \quad \text{where} \quad w_x(y) = \prod_{i=1}^{m} \frac{1 + x_i y_i}{2} \text{ for any } x \in \mathbb{R}^m. \tag{2.5}$$

The above implies that for a boolean function $f : \{\pm 1\}^m \to [-1, 1]$, the harmonic extension of f also satisfies $\max_{x \in [1,1]^m} |f(x)| \leq 1$.

The discrete derivative of a function on the hypercube $\{\pm 1\}^m$ is given by

$$\partial_i f(x) = \frac{1}{2}(f(x^{i \to 1}) - f(x^{i \to -1})),$$

where $x^{i \to b}$ is the same as x except that the i-th coordinate is set to b. It is easily checked that the harmonic extension of $\partial_i f(x)$ is the real partial derivative $\frac{\partial}{\partial x_i}$ of the harmonic extension of f and we will identify it as such. Furthermore, for a boolean function $f : \{\pm 1\}^m \to [-1, 1]$, the discrete derivative at any point $x \in \{\pm 1\}^m$ also satisfies $|\partial_i f(x)| \leq 1$ and hence (2.5) implies that $\max_{x \in [1,1]^m} |\partial_A f(x)| \leq 1$ for any $A \subseteq [m]$ identifying $\partial_A f$ as the harmonic extension of the real partial derivative of f. Moreover, from (2.4), it also follows that

$$\partial_A f(x) = \sum_{S : S \supseteq A} \hat{f}(S)\chi_{S \setminus A}(x) \tag{2.6}$$

for any subset $A \subseteq [m]$. The above also implies that $\partial_A f(0) = \hat{f}(A)$.

The level-ℓ Fourier weight of f is defined as $\mathsf{wt}_\ell(f) = \sum_{|S|=\ell} |\hat{f}(S)|$.

For a function $f(x_1, \ldots, x_m)$, a restriction $\rho \in \{-1, 1, \star\}^m$ gives a partial assignment to the variables $(x_i)_{i \leq m}$. We denote the set of coordinates of ρ whose value is $\star$ as $\mathsf{free}(\rho)$ while the set of coordinates that are fixed to ± 1 is denoted by $\mathsf{fix}(\rho)$. We use f_ρ to denote the function obtained from f by setting the variables in $\mathsf{fix}(\rho)$ to the values given by ρ.

Fourier basis for biased measures. For a proofs of the results below, see Chapter 8 in [O'D14]. Given any $\mu \in (-1, 1)^m$, let $p_\mu(X)$ be the biased product distribution over $\{\pm 1\}^m$ such that each coordinate of $X \in \{\pm 1\}^m$ is sampled independently so that $X_i = 1$ with probability $(1 + \mu_i)/2$ and $X_i = -1$ with probability $(1 - \mu_i)/2$. So the expectation and the variance of X_i are

$$\mathbb{E}_{p_\mu}[X_i] = \mu_i, \quad \text{and} \quad \mathbb{E}_{p_\mu}[(X_i - \mu_i)^2] = 1 - \mu_i^2.$$

Then, the Fourier basis with respect to the biased product measure p_μ is given by the following functions indexed by subsets $S \subseteq [n]$:

$$\phi_S^\mu(x) = \prod_{i \in S} \phi_i^\mu(x), \qquad \text{where} \qquad \phi_i^\mu(x) = \frac{x_i - \mu_i}{\sigma_i},$$

with $\sigma_i = (1-\mu_i^2)^{1/2}$ being the standard deviation of the biased random bit X_i. Note that

$$\mathbb{E}_{p_\mu}[\phi_S^\mu(X)^2] \;=\; \prod_{i\in S} \mathbb{E}_{p_\mu}[\phi_i^\mu(X)^2] \;=\; \prod_{i\in S} \frac{1}{\sigma_i^2}\cdot \mathbb{E}_{p_\mu}[(X_i-\mu_i)^2] \;=\; 1,$$

and that $\mathbb{E}_{p_\mu}[\phi_S^\mu(X)\phi_T^\mu(X)] = 0$ if $S \neq T$. So the functions $\phi_S^\mu(x)$ form an orthonormal basis for real-valued functions on $\{\pm 1\}^m$ with respect to the inner product obtained by taking expectation under p_μ. The Fourier expansion with respect to the biased product measure p_μ is given by

$$f(x) = \sum_{S\subseteq[n]} \hat{f}^\mu(S)\phi_S^\mu(x), \tag{2.7}$$

where $\hat{f}^\mu(S) = \mathbb{E}_{p_\mu(x)}[f(x)\phi_S^\mu(x)]$ are the Fourier coefficients with respect to p_μ.

The discrete derivative with respect to ϕ_i^μ is defined as

$$\partial_i^\mu f(x) := \frac{f(x^{i\to 1}) - f(x^{i\to -1})}{\phi_i^\mu(1) - \phi_i^\mu(-1)} = \sigma_i \cdot \frac{f(x^{i\to 1}) - f(x^{i\to -1})}{2} = \sigma_i \cdot \partial_i f(x), \tag{2.8}$$

where $\partial_i f(x)$ is the discrete derivative with respect to the standard Fourier basis (with respect to the uniform measure over $\{\pm 1\}^m$).

Since $\partial_i f$ can be viewed as the real partial derivative of the harmonic extension of f, using the chain rule for taking derivatives, $\dfrac{\partial f}{\partial \phi_i^\mu} = \sigma_i \cdot \partial_i f$, so one can identify $\partial_i^\mu f$ as the real partial derivative $\dfrac{\partial f}{\partial \phi_i^\mu}$ for the harmonic extension of f. Moreover, from (2.7), it also follows that $\partial_S^\mu f(\mu) = \hat{f}^\mu(S)$ for any subset $S \subseteq [n]$, so μ acts as the origin with respect to the biased measure.

The level-ℓ Fourier weight of f with respect to bias μ is defined as $\mathsf{wt}_\ell^\mu(f) = \sum_{|S|=\ell} |\hat{f}^\mu(S)|$.

3 Input Distribution and the Proof Outline

We now give a formal outline of the proof. We first give an input distribution for which (δ,k)-Forrelation is easy to compute using quantum queries, but hard for classical queries. Our distribution is a variant of those used in [RT19, Tal19].

To define the distribution we first introduce some notation. Let $\mathsf{trnc} : \mathbb{R} \to [-1/2, 1/2]$ denote the following function,

$$\mathsf{trnc}(s) = \begin{cases} \min\{1/2, s\} & \text{if } s \geq 0, \\ \max\{-1/2, s\} & \text{if } s \leq 0. \end{cases}$$

For notational convenience, we will write $\mathsf{trnc}(s_1,\ldots,s_m)$ to denote $(\mathsf{trnc}(s_1),\ldots,\mathsf{trnc}(s_m))$. Let us also introduce the following *block shifted Hadamard product* of two vectors: given vectors $x := (x_1,\cdots,x_{k-1}) \in \mathbb{R}^{(k-1)N}$ and $y := (y_1,\cdots,y_{k-1}) \in \mathbb{R}^{(k-1)N}$, we define $x \diamond y$ to be the following vector in $\mathbb{R}^{kN}$,

$$x \diamond y = (x_1,\cdots,x_{k-1},\mathbf{1}) \odot (\mathbf{1},y_1,\cdots,y_{k-1}) = (x_1, y_1\odot x_2, y_2 \odot x_3, \ldots, y_{k-2}\odot x_{k-1}, y_{k-1}), \tag{3.1}$$

where $\mathbf{1}$ is the all ones vector in $\mathbb{R}^N$ and $\odot$ is the Hadamard product of two vectors. The above product will allow a natural generalization of the input distribution described in Section 1.2 to the case of arbitrary k. To see some examples, for $k=2$ and vectors $x, y \in \mathbb{R}^n$, we have that $x \diamond y = (x,y)$; while for $k=3$, we have that $x \diamond y = (x_1, y_1\odot x_2, y_2) = (x_1, x_2 \odot y_1, y_2)$ reminiscent of the expression appearing in (1.5).

We can now describe the input distribution. Recall that $\epsilon = 1/(64k^2\log(kN))$ and $\delta = \epsilon^k$ and let $\Sigma = \epsilon\begin{pmatrix} \mathsf{I}_N & \mathsf{H}_N \\ \mathsf{H}_N & \mathsf{I}_N \end{pmatrix}$. Then, our input distribution $p(Z) = \frac{1}{2}p_0(Z) + \frac{1}{2}p_1(Z)$ where $p_0(Z)$ and $p_1(Z)$ are defined in Figure 1.

We now show that $p_b(Z)$ for $b \in \{0,1\}$ has a large support on b-inputs for (δ,k)-Forrelation.

Distribution $p_0(Z)$**:** Z is uniform over $\{\pm 1\}^{kN}$.

Distribution $p_1(Z)$**:** Let $(U_\kappa, V_\kappa)_{\kappa\in[k-1]}$ be independent random variables in $\mathbb{R}^{2N}$ that are distributed as $\mathcal{N}(0,\Sigma)$. Write $U = (U_\kappa)_{\kappa\in[k-1]}$ and $V = (V_\kappa)_{\kappa\in[k-1]}$ and define $W = \mathsf{trnc}(U) \diamond \mathsf{trnc}(V)$ where $W \in [-1/2, 1/2]^{kN}$. Let $Z = (Z_1, \ldots, Z_k) \in \{\pm 1\}^{kN}$ be obtained by rounding each coordinate of the vector W independently to ± 1 by interpreting them as means, i.e., for each coordinate $i \in [kN]$, we have $\mathbb{E}[Z(i) \mid U, V] = W(i)$.

Figure 1: Input Distributions $p_0(Z)$ and $p_1(Z)$

Theorem 3.1. *For the input distribution defined in Figure 1,*

$$p_0(\mathsf{forr}_{\delta,k} \textit{ outputs } 0) \geq 1 - \frac{4}{\delta^2 N} \quad \textit{and} \quad p_1(\mathsf{forr}_{\delta,k} \textit{ outputs } 1) \geq 6\delta.$$

Proof. We first consider p_0. Since $p_0(z)$ is uniform on $\{\pm 1\}^{kN}$ and $\mathsf{forr}_k(z)$ is a multilinear and homogeneous polynomial, clearly $\mathbb{E}_{p_0(z)}[\mathsf{forr}_k(z)] = 0$. Next, we claim that $\mathbb{E}_{p_0}[\mathsf{forr}_k(Z)^2] \leq 1/N$. To see this, we use the quadratic form description (1.2). Fix any values $z_2, \ldots, z_{k-1}$, and let $\mathsf{A} = \mathsf{H}\cdot\mathsf{diag}(z_2)\cdot\cdots\cdot\mathsf{H}\cdot\mathsf{diag}(z_{k-1})\cdot\mathsf{H}$ be the matrix appearing in the quadratic form which satisfies $\|\mathsf{A}\|_{\mathsf{op}} \leq 1$. Then, we have

$$\begin{aligned}\mathbb{E}_{p_0}[\mathsf{forr}_k(Z)^2] &= \frac{1}{N^2}\mathbb{E}_{p_0}[(Z_1^\top \mathsf{A} Z_k)^2] = \frac{1}{N^2}\sum_{ij,rs}\mathbb{E}_{p_0}[\mathsf{A}_{ij}\mathsf{A}_{rs}\cdot Z_1(i)Z_k(j)Z_1(r)Z_k(s)] \\ &= \frac{1}{N^2}\sum_{ij}\mathsf{A}_{ij}^2 = \frac{\|\mathsf{A}\|_F^2}{N^2} \leq \frac{N\|\mathsf{A}\|_{\mathsf{op}}^2}{N^2} \leq \frac{1}{N}.\end{aligned}$$

By Chebshev's inequality, it follows that $p_0(\mathsf{forr}_{\delta,k}(Z) \text{ outputs } 1) \leq p_0(|\mathsf{forr}_k(Z)| \geq \delta/2) \leq (4/\delta^2 N)$.

We now consider p_1. As $\mathsf{forr}_k(z)$ is a multilinear polynomial, from the description of $p_1(Z)$, we have that $\mathbb{E}_{p_1}[\mathsf{forr}_k(z) \mid U, V] = \mathsf{forr}_k(\mathsf{trnc}(U)\diamond\mathsf{trnc}(V))$. Using Lemma 4.1, which shows that truncating the Gaussians has negligible effect in expectation, i.e. $|\mathbb{E}_{p_1}[f(\mathsf{trnc}(U)\diamond\mathsf{trnc}(V))] - \mathbb{E}_{p_1}[f(U,V)]| \leq N^{-3k^2}$, we have that

$$\begin{aligned}&\mathbb{E}_{p_1}[\mathsf{forr}_k(Z)] + N^{-3k^2} \geq \mathbb{E}_{p_1}[\mathsf{forr}_k(U\diamond V)] \\ &= \frac{1}{N}\sum_{\underline{i}}\mathbb{E}[U_1(i_1)\cdot\mathsf{H}_{i_1,i_2}\cdot V_1(i_2)U_2(i_2)\cdot\mathsf{H}_{i_2,i_3}\cdots\mathsf{H}_{i_{k-2},i_{k-1}}\cdot V_{k-2}(i_{k-1})U_{k-1}(i_k)\cdot\mathsf{H}_{i_{k-1},i_k}\cdot V_{k-1}(i_k)] \\ &= \frac{1}{N}\sum_{\underline{i}}\mathbb{E}[U_1(i_1)\cdot\mathsf{H}_{i_1,i_2}\cdot V_1(i_2)]\cdot\mathbb{E}[U_2(i_2)\cdot\mathsf{H}_{i_2,i_3}\cdot V_2(i_3)]\cdots\mathbb{E}[U_{k-1}(i_k)\cdot\mathsf{H}_{i_{k-1},i_k}\cdot V_{k-1}(i_k)] \\ &= \frac{1}{N}\sum_{\underline{i}}\epsilon^{k-1}\cdot\mathsf{H}_{i_1,i_2}^2\cdots\mathsf{H}_{i_{k-1},i_k}^2 = \frac{1}{N}\sum_{\underline{i}}\epsilon^{k-1}\cdot\frac{1}{N^{k-1}} = \epsilon^{k-1},\end{aligned}$$

where the second equality used that (U_κ, V_κ) are independent Gaussians for different values of κ, the third equality follows since $\mathbb{E}[U_\kappa(i)V_\kappa(j)] = \epsilon\cdot\mathsf{H}_{ij}$, and the fourth equality follows since each entry of H is $\pm\frac{1}{\sqrt{N}}$ and the sum if over N^k indices. It thus follows that $\mathbb{E}_{p_1}[\mathsf{forr}_k(Z)] \geq \epsilon^{k-1} - N^{-3k^2} \geq 10\epsilon^k = 10\delta$.

Let $\alpha = p_1(\mathsf{forr}_k(Z) \geq \delta)$. Recalling (1.2), we have that $|\mathsf{forr}_k(z)| \leq 1$ for $z \in [-1,1]^{kN}$. So, the above gives that $\alpha + (1-\alpha)\delta \geq 10\delta$ and hence in particular that $\alpha \geq 6\delta$, as $\delta \ll 1$. ■

To prove a lower bound for classical query algorithms (decision trees), we show that the advantage of any bounded real-valued function on $\{\pm 1\}^{kN}$ can be computed in terms of the low-level Fourier weight of the function f with respect to biased measures, as mentioned in Section 1.2. In particular, for $\mu \in [-1/2, 1/2]^{kN}$, consider the product measure p_μ induced on $Z \in \{\pm 1\}^{kN}$ by sampling each bit independently so that $\mathbb{E}_{p_\mu}[Z_i] = \mu_i$. Then, we prove the following which is the main contribution of this work.

Theorem 3.2. *Let $f : \{\pm 1\}^{kN} \to [0,1]$. Then,*

$$\left|\mathbb{E}_{p_1}[f(Z)] - \mathbb{E}_{p_0}[f(Z)]\right| \le \sup_{\mu \in [-\frac{1}{2},\frac{1}{2}]^{kN}} \sum_{m=1}^{k-1} \left(\frac{\epsilon}{\sqrt{N}}\right)^{m(k-1)} \cdot (8k)^{4mk} \cdot \mathsf{wt}^{\mu}_{mk}(f) + N^{-k}.$$

Note that in the previous work of [RT19] for the standard Forrelation problem ($k = 2$), one only gets an upper bound in terms of the level-2 weight of the function f, but here we have an upper bound in terms of level ℓ weights where $\ell \in \{mk \mid m \in [k-1]\}$. We stress that the weight of the higher level mk can be much larger than the level-k weight, but the extra $\epsilon/\sqrt{N}$ factors in the above theorem takes care of it.

To bound the level-ℓ Fourier weight with respect to biased measures, we use the following bound proven in [Tal19] for Fourier weights under the uniform measure.

Theorem 3.3 ([Tal19])**.** *Let $f : \{\pm 1\}^m \to [0,1]$ be the acceptance probability function of a randomized depth-d decision tree. Then, for any $\ell \le d$, the following holds for a universal constant c,*

$$\mathsf{wt}_\ell(f) \le \left((cd)^\ell \log^{\ell-1} m\right)^{1/2},$$

where the Fourier weight $\mathsf{wt}_\ell(f)$ is with respect to the uniform measure on $\{\pm 1\}^m$.

We prove the following general statement showing that if a function and all its restrictions have a small Fourier weight on level-ℓ with respect to the uniform measure, then the Fourier weight with respect to an arbitrary bias $\mu \in [-1/2, 1/2]^m$ is also small.

Theorem 3.4. *Let $f : \{\pm 1\}^m \to \mathbb{R}$ and $\ell \in [m]$. Let w be such that for any restriction $\rho \in \{-1, 1, \star\}^m$, we have $\mathsf{wt}_\ell(f_\rho) \le w$ where the Fourier weight is with respect to the uniform measure. Then, for any $\mu \in [-1/2, 1/2]^m$, we have $\mathsf{wt}^{\mu}_\ell(f) \le 4^\ell w$.*

Since depth-d decision trees are closed under restrictions, combining Theorem 3.4 with Theorem 3.3 gives us that the level-ℓ weight of depth-d decision trees with an arbitrary bias $\mu \in [-1/2, 1/2]^m$ is also bounded by $\left((cd)^\ell \log^{\ell-1}(m)\right)^{1/2}$.

Corollary 3.5. *Let $f : \{\pm 1\}^m \to [0,1]$ be the acceptance probability function of a randomized depth-d decision tree. Then, for any $\mu \in [-1/2, 1/2]^m$ and $\ell \le d$, we have $\mathsf{wt}^{\mu}_\ell(f) \le \left((cd)^\ell \log^{\ell-1} m\right)^{1/2}$ for a universal constant c.*

Combined with Theorem 3.2, the above implies that if the depth d of the decision tree satisfies $d \ll N^{1-1/k}$, then the advantage of f would be much smaller than δ.

4 Bounding the Truncation Error

To prove our results, it will be much more convenient to ignore the truncation and work with Gaussians. We show in this section that this can be done with a small error. In particular, let $f : \{\pm 1\}^{kN} \to \mathbb{R}$ be a function on the hypercube, then identifying f with its harmonic extension, the definition of the distribution $p_1(Z)$ implies that

$$\mathbb{E}_{p_1}[f(Z)] = \mathbb{E}_{p_1}[f(\mathsf{trnc}(U) \diamond \mathsf{trnc}(V))], \tag{4.1}$$

where $(U_\kappa, V_\kappa)_{\kappa \in [k-1]}$ are independent multivariate Gaussians. The equality follows since f is multilinear and each bit of Z is independently chosen so that $\mathbb{E}[Z \mid U, V] = \mathsf{trnc}(U) \diamond \mathsf{trnc}(V)$.

The next lemma shows that up to a small error, one can replace the truncated Gaussian with the Gaussian.

Lemma 4.1. *Let $f : \{\pm 1\}^{kN} \to [-1,1]$. Then, identifying f with its harmonic extension,*

$$\mathbb{E}_{p_1}[f(\mathsf{trnc}(U) \diamond \mathsf{trnc}(V))] = \mathbb{E}_{p_1}[f(U \diamond V)] + \mathsf{err},$$

where $|\mathsf{err}| \leq N^{-3k^2}$.

Proof. Let $\mathcal{U}_\kappa$ and $\mathcal{V}_\kappa$ respectively be the events that $X_\kappa \neq U_\kappa$ and $Y_\kappa \neq V_\kappa$. Let us also define bad events $\mathcal{B}_\kappa = (\mathcal{U}_\kappa \cup \mathcal{V}_\kappa)$ and $\mathcal{B} = \cup_\kappa \mathcal{B}_\kappa$. Then, using (4.1), it suffices to bound

$$\mathsf{err} := \mathbb{E}[\mathbf{1}_{\mathcal{B}} f(\mathsf{trnc}(U) \diamond \mathsf{trnc}(V))] - \mathbb{E}[\mathbf{1}_{\mathcal{B}} f(U \diamond V)], \tag{4.2}$$

since otherwise the values taken by f are the same.

To bound the above, we first claim that since $|f|$ is bounded by 1 inside the hypercube $[-1,1]^{kN}$, one can bound the value of $|f|$ at any point by the following claim.

Claim 4.2. *For any $z \in \mathbb{R}^{kN}$, we have $|f(z)| \leq \prod_{j \in [k]} h(z_j)$, where $h : \mathbb{R}^N \to [0, \infty)$ is the function defined as $h(x) = \prod_{i=1}^{N} \max\{1, |x(i)|\}$.*

Given the above claim, which we shall prove later, we can proceed as follows. For $a, b \in \mathbb{R}$, it holds that $\max\{1, |ab|\} \leq \max\{1, |a|\} \cdot \max\{1, |b|\}$, so using Claim 4.2,

$$|\mathsf{err}| \quad \leq \quad \mathbb{E}[\mathbf{1}_{\mathcal{B}}] + \mathbb{E}\left[\mathbf{1}_{\mathcal{B}} \cdot \prod_{\kappa \in [k-1]} h(U_\kappa) \cdot h(V_\kappa)\right]. \tag{4.3}$$

Note that the marginal distribution on each coordinate of U_κ is $\mathcal{N}(0, \epsilon)$, so from a union bound and Gaussian concentration (Proposition 2.1), it follows that $\mathbb{E}[\mathbf{1}_{\mathcal{B}}] \leq 4kNe^{-1/(8\epsilon)} \leq N^{-4k^2}$. To bound the second term, we shall prove that

Claim 4.3. *For any $\kappa \in [k-1]$, we have $\mathbb{E}[\mathbf{1}_{\mathcal{B}_\kappa} \cdot h(U_\kappa) \cdot h(V_\kappa)] \leq N^{-6k^2}$.*

We first finish the proof of Lemma 4.1 assuming the above. First, notice that (U_κ, V_κ) are independent for different values of κ, and when $\mathcal{B}_k$ does not occur, then $h(U_\kappa) \cdot h(V_\kappa) \leq 1$. Therefore, decomposing the event $\mathcal{B}$ into further sub-events and using Claim 4.3, we can bound

$$\mathbb{E}\left[\mathbf{1}_{\mathcal{B}} \cdot \prod_{\kappa \in [k-1]} h(U_\kappa) \cdot h(V_\kappa)\right] \;\leq\; \sum_{\substack{T \subseteq [k-1] \\ T \neq \emptyset}} \prod_{\kappa \in T} \mathbb{E}\left[\mathbf{1}_{\mathcal{B}_\kappa} \cdot h(U_\kappa) \cdot h(V_\kappa)\right] \;\leq\; 2^k N^{-6k^2}.$$

Altogether, we get that $|\mathsf{err}| \leq N^{-3k^2}$ for large enough N. This finishes the proof of Lemma 4.1 assuming Claim 4.2 and Claim 4.3 which we prove next. ■

Proof of Claim 4.2. The harmonic extension of f is explicitly given by (2.5) using interpolation on the vertices of the hypercube $[-1,1]^{kN}$. Thus, for any $z = (z_1, \ldots, z_\kappa) \in \mathbb{R}^{kN}$, we have

$$\begin{aligned} |f(z)| &\leq \sum_{y \in \{\pm 1\}^{kN}} |f(y)| \cdot \left| \prod_{i=1}^{kN} \frac{1 + z(i)y(i)}{2} \right| \\ &\leq \sum_{y \in \{\pm 1\}^{kN}} \prod_{i=1}^{kN} \frac{|1 + z(i)y(i)|}{2} = \prod_{i=1}^{kN} \left(\frac{|1 + z(i)|}{2} + \frac{|1 - z(i)|}{2} \right) = \prod_{i=1}^{kN} \max\{1, |z(i)|\} = \prod_{j \in [k]} h(z_j) \end{aligned}$$ ■

Proof of Claim 4.3. Recall that $\mathcal{B}_\kappa = \mathcal{U}_\kappa \cup \mathcal{V}_\kappa$ and when $\mathcal{U}_\kappa$ (resp. $\mathcal{V}_\kappa$) does not occur then $h(U_\kappa) \leq 1$ (resp. $h(V_\kappa) \leq 1$). Therefore,

$$\begin{aligned} \mathbb{E}[\mathbf{1}_{\mathcal{B}_\kappa} \cdot h(U_\kappa) \cdot h(V_\kappa)] &\leq \mathbb{E}[\mathbf{1}_{\mathcal{U}_\kappa} \cdot h(U_\kappa)] + \mathbb{E}[\mathbf{1}_{\mathcal{V}_\kappa} \cdot h(V_\kappa)] + \mathbb{E}[\mathbf{1}_{\mathcal{U}_\kappa} h(U_\kappa) \cdot \mathbf{1}_{\mathcal{V}_\kappa} h(V_\kappa)] \\ &\leq \mathbb{E}[\mathbf{1}_{\mathcal{U}_\kappa} \cdot h^2(U_\kappa)] + \mathbb{E}[\mathbf{1}_{\mathcal{V}_\kappa} \cdot h^2(V_\kappa)] + \sqrt{\mathbb{E}[\mathbf{1}_{\mathcal{U}_\kappa} \cdot h^2(U_\kappa)] \cdot \mathbb{E}[\mathbf{1}_{\mathcal{V}_\kappa} \cdot h^2(V_\kappa)]}, \end{aligned} \tag{4.4}$$

where the second inequality follows as $|h(x)| \leq h^2(x)$ for all $x \in \mathbb{R}^N$, and applying the Cauchy-Schwarz inequality.

As U_κ and V_κ are marginally distributed as the N-dimensional Gaussian $\mathcal{N}(0, \epsilon \cdot I_N)$, it follows that the right hand side in (4.4) is $3 \cdot \mathbb{E}[\mathbf{1}_{\mathcal{U}_\kappa} \cdot h^2(U_\kappa)]$. Therefore, decomposing the event $\mathcal{B}_k$ into further sub-events that correspond to a single coordinate taking value larger than $1/2$, we have

$$\mathbb{E}[\mathbf{1}_{\mathcal{B}_\kappa} \cdot h(U_\kappa) \cdot h(V_\kappa)] \leq 3 \cdot \sum_{t=1}^{N} \binom{N}{t} \cdot \alpha^t = 3\left((1+\alpha)^N - 1\right), \tag{4.5}$$

where $\alpha = \mathbb{E}\left[\mathbf{1}_{|G| \geq 1/2} \cdot \max\{1, G^2\}\right]$ for $G \in \mathbb{R}$ drawn from $\mathcal{N}(0, \epsilon)$. A direct computation using Proposition 2.2 shows that

$$\begin{aligned} \alpha \quad &= \quad \frac{2}{\sqrt{\epsilon}} \int_{1/2}^{1} \gamma\left(\frac{s}{\sqrt{\epsilon}}\right) ds \quad + \quad \frac{2}{\sqrt{\epsilon}} \int_{1}^{\infty} s^2 \gamma\left(\frac{s}{\sqrt{\epsilon}}\right) ds \\ &\leq \quad \frac{2}{\sqrt{\epsilon}} \cdot e^{-\frac{1}{8\epsilon}} \quad + \quad 2\epsilon \int_{1/\sqrt{\epsilon}}^{\infty} s^2 \gamma(s) ds \quad \leq \quad \frac{4}{\sqrt{\epsilon}} \cdot e^{-\frac{1}{8\epsilon}} \quad \leq \quad N^{-6k^2}. \end{aligned}$$

Plugging it back in (4.5), we get that

$$\mathbb{E}[\mathbf{1}_{\mathcal{B}_\kappa} \cdot h(U_\kappa) \cdot h(V_\kappa)] \quad \leq \quad 3\left((1+\alpha)^N - 1\right) \quad \leq \quad 6\alpha N \quad \leq \quad N^{-4k^2},$$

where we used that $(1+r)^N \leq 1 + rN + (rN)^2 \leq 1 + 2rN$ holds for $r \leq 1/N^2$, by Taylor series expansion. ■

5 Lower Bound for Decision Trees

We first prove Theorem 3.2 that bounds the advantage of the randomized decision tree in terms of biased Fourier weights. Following that, we show how to bound the Fourier weight of a function under a biased measure (Theorem 3.4) by using a random restriction argument. We assemble all the pieces together to prove Theorem 1.3 and Corollary 1.4 following that.

5.1 Advantage in terms of Fourier weight: Proof of Theorem 3.2

Using Lemma 4.1 we have that

$$\mathbb{E}_{p_1}[f(Z)] - f(0) = \mathbb{E}_{p_1}[f(U \diamond V)] - f(0) + \mathsf{err}, \tag{5.1}$$

where $|\mathsf{err}| \leq N^{-3k^2}$.

To evaluate the first term on the right hand side, we will use Gaussian interpolation. Recall that $(U_\kappa, V_\kappa)_{\kappa \in [k-1]}$ are independent multivariate Gaussians. We will interpolate them separately. In particular, for each $\kappa \in [k]$ and $t_\kappa \in (0,1)$, we define

$$(\mathbf{U}_\kappa(t_\kappa), \mathbf{V}_\kappa(t_\kappa)) = \sqrt{t_\kappa} \cdot (U_\kappa, V_\kappa).$$

We will refer to the interpolation parameter $t = (t_1, \cdots, t_{k-1})$ as time and we will drop the time index and just write $\mathbf{U}$ and so on, if there is no ambiguity. We remind the reader of our convention that bold fonts will always refer to the interpolated Gaussian.

To use Gaussian interpolation, we consider the function $\varphi : (0,1)^{k-1} \to \mathbb{R}$ defined as

$$\varphi(t) = \mathbb{E}[f(\mathbf{U}(t) \diamond \mathbf{V}(t))].$$

For any fixed values of $t_1, \cdots, t_{k-2}$, by the fundamental theorem of calculus we have that

$$\begin{aligned} \mathbb{E}[f(\mathbf{U}(t_1, \cdots, t_{k-2}, 1) \diamond \mathbf{V}(t_1, \cdots, t_{k-2}, 1))] - \mathbb{E}[f(\mathbf{U}(t_1, \cdots, t_{k-2}, 0) \diamond \mathbf{V}(t_1, \cdots, t_{k-2}, 0))] \\ = \int_0^1 \frac{\partial \varphi}{\partial t_{k-1}}(t) dt_{k-1}. \end{aligned}$$

Repeating the above and fixing each index of the time parameter one by one, we obtain

$$\begin{aligned}\mathbb{E}[f(U \diamond V)] - f(0) &= \mathbb{E}[f(\mathbf{U}(\mathbf{1}) \diamond \mathbf{V}(\mathbf{1})] - \mathbb{E}[f(\mathbf{U}(0) \diamond \mathbf{V}(0)] \\ &= \underset{[0,1]^{k-1}}{\int \cdots \int} \frac{\partial \varphi}{\partial t_1 \cdots \partial t_{k-1}}(t) dt_{k-1} \cdots dt_1, \end{aligned} \tag{5.2}$$

where $\mathbf{1}$ is the all ones vector in $\mathbb{R}^{k-1}$.

To bound the value of the above partial derivative (taken with respect to the time parameters) at any point, we will use Lemma 2.3. Since $f(z)$ is a multilinear polynomial, it suffices to compute the derivative of a character and towards this end, we show the following key lemma in terms of derivatives $\partial_J f = \frac{\partial f}{\partial z_J}$ where the order of the derivative $|J|$ is always between k and $k(k-1)$.

Lemma 5.1. *Let $t \in (0,1)^{k-1}$ and $S \subseteq [kN]$. Defining $\varphi_S(t) = \mathbb{E}[\chi_S(\mathbf{U}(t) \diamond \mathbf{V}(t))]$, the following holds*

$$\frac{\partial \varphi_S}{\partial t_1 \cdots \partial t_{k-1}}(t) = \frac{1}{2^{k-1}} \sum_{m=1}^{k-1} \left(\frac{\epsilon}{\sqrt{N}}\right)^{m(k-1)} \cdot \sum_{\substack{J \subseteq S \\ |J| = mk}} \theta_J(t) \cdot \mathbb{E}[\chi_{S \setminus J}(\mathbf{U}(t) \diamond \mathbf{V}(t))],$$

where $\theta_J(t) = \theta_J(t_1, \cdots, t_{k-1})$ is a polynomial that only depends on J (and not on S) and the sum of the absolute value of all the coefficients of θ_J is bounded by $(4k)^{4|J|}$.

We first finish the proof of Theorem 3.2 and then prove the above lemma. Given Lemma 5.1, since $\varphi(t) = \sum_{S \subseteq [kN]} \hat{f}(S) \varphi_S(t)$, by linearity of expectation and exchanging the order of summation, it follows that for a given time t,

$$\begin{aligned}\frac{\partial \varphi}{\partial t_1 \cdots \partial t_{k-1}}(t) &= \frac{1}{2^{k-1}} \sum_{m=1}^{k-1} \left(\frac{\epsilon}{\sqrt{N}}\right)^{m(k-1)} \cdot \mathbb{E}\left[\sum_{\substack{J \subseteq [kN] \\ |J| = mk}} \theta_J(t) \cdot \sum_{S : S \supseteq J} \hat{f}(S) \chi_{S \setminus J}(\mathbf{U} \diamond \mathbf{V})\right] \\ &= \frac{1}{2^{k-1}} \sum_{m=1}^{k-1} \left(\frac{\epsilon}{\sqrt{N}}\right)^{m(k-1)} \cdot \mathbb{E}\left[\sum_{\substack{J \subseteq [kN] \\ |J| = mk}} \theta_J(t) \cdot \partial_J f(\mathbf{U} \diamond \mathbf{V})\right],\end{aligned}$$

where the second equality uses (2.6).

To proceed, we want to relate the derivatives above to the Fourier coefficients under a biased measure but the coordinates of $\mathbf{U} \diamond \mathbf{V}$ can be large with a small probability. This is however not a problem, since we can use Lemma 4.1 once again to go back to the truncated Gaussians with a small error. In particular, since for each subset $J \subseteq [kN]$, the derivative $\partial_J(f)$ is bounded by one on the hypercube $\{\pm 1\}^{kN}$, we can apply Lemma 4.1 (the statement still holds for the interpolated Gaussians by following the same proof), to get that

$$\frac{\partial \varphi}{\partial t_1 \cdots \partial t_{k-1}}(t) = \frac{1}{2^{k-1}} \sum_{m=1}^{k-1} \left(\frac{\epsilon}{\sqrt{N}}\right)^{m(k-1)} \cdot \mathbb{E}\left[\sum_{\substack{J \subseteq [kN] \\ |J| = mk}} \theta_J(t) \cdot \partial_J f(\mathsf{trnc}(\mathbf{U}) \diamond \mathsf{trnc}(\mathbf{V}))\right] + \mathsf{err}, \tag{5.3}$$

where we replaced $\mathbf{U} \diamond \mathbf{V}$ with $\mathsf{trnc}(\mathbf{U}) \diamond \mathsf{trnc}(\mathbf{V})$ at the expense of a real number err satisfying

$$|\mathsf{err}| \le \frac{1}{2^{k-1}} \sum_{m=1}^{k-1} \left(\frac{\epsilon}{\sqrt{N}}\right)^{m(k-1)} \cdot (4k)^{4mk} \cdot (kN)^{mk} \cdot N^{-3k^2} \le N^{-k},$$

for large enough N. The $(kN)^{mk}$ term above comes from the number of possible choices of sets J of size mk, and for $t \in (0,1)^{k-1}$, we used that $|\theta_J(t)| \le (4k)^{4mk}$ upper bounding it by the sum of the absolute value of all the coefficients.

Next, we express the derivatives in (5.3) as biased Fourier coefficients. For any fixed value $\mu \in [-1/2, 1/2]^{kN}$ taken by $\mathsf{trnc}(\mathbf{U}(t)) \diamond \mathsf{trnc}(\mathbf{V}(t))$ and recalling the identity (2.8), we see that $\partial_J f(z) = \sigma_J^{-1} \hat{f}^{\mu}(J)$ where $\sigma_J = \prod_{i \in J} \sigma_i$ with $\sigma_i = \sqrt{1-\mu_i^2} \geq 1/2$. Furthermore, as $|\theta_J(t)| \leq (4k)^{4mk}$, equation (5.3) gives us that the following holds for any $t \in (0,1)^{k-1}$,

$$\begin{aligned}\left|\frac{\partial \varphi}{\partial t_1 \cdots \partial t_{k-1}}(t)\right| &\leq \sup_{\mu \in [-1/2,1/2]^{kN}} \sum_{m=1}^{k-1} \left(\frac{\epsilon}{\sqrt{N}}\right)^{m(k-1)} \sum_{\substack{J \subseteq [kN] \\ |J| = mk}} |\theta_J(t)| \cdot \sigma_J^{-1} \cdot |\hat{f}^{\mu}(J)| + N^{-k}. \\ &\leq \sup_{\mu \in [-1/2,1/2]^{kN}} \sum_{m=1}^{k-1} \left(\frac{\epsilon}{\sqrt{N}}\right)^{m(k-1)} \cdot (8k)^{4mk} \cdot \mathsf{wt}^{\mu}_{mk}(f) + N^{-k}.\end{aligned}$$

Finally, using (5.2) and (5.1), the above implies that

$$|\mathbb{E}_{p_1}[f(Z)] - f(0)| \leq \sup_{\mu \in [-1/2,1/2]^{kN}} \sum_{m=1}^{k-1} \left(\frac{\epsilon}{\sqrt{N}}\right)^{m(k-1)} \cdot (8k)^{4mk} \cdot \mathsf{wt}^{\mu}_{mk}(f) + N^{-k},$$

completing the proof of Theorem 3.2 given Lemma 5.1, which we prove next.

5.2 Proof of Lemma 5.1

For ease of exposition we first give the proof for the simpler case of $k = 3$. The application of Gaussian integration by parts is much easier here, as it does not recursively lead to other terms. For larger values of k, we need more technical care and additional ideas in the form of a careful counting argument.

Proof for the $k = 3$ Case

In this case, we shall prove that

$$4 \cdot \frac{\partial \varphi}{\partial t_1 \partial t_2}(t) = \sum_{\substack{J \subseteq S \\ |J| = 3}} \left(\frac{\epsilon}{\sqrt{N}}\right)^2 \cdot \theta_J(t) \cdot \mathbb{E}[\chi_{S \setminus J}(\mathbf{U} \diamond \mathbf{V})] + \sum_{\substack{J \subseteq S \\ |J| = 6}} \left(\frac{\epsilon}{\sqrt{N}}\right)^4 \cdot \theta_J(t) \cdot \mathbb{E}[\chi_{S \setminus J}(\mathbf{U} \diamond \mathbf{V})], \qquad (5.4)$$

where $\theta_J(t) = \theta_J(t_1, t_2)$ is a polynomial for which the sum of absolute value of all the coefficients is at most $12^{4|J|}$ and θ_J only depends on J and not on S.

Let $S = S_1 \sqcup S_2 \sqcup S_3$ where $S_1 \subseteq [N], S_2 \subseteq \{N+1, \ldots, 2N\}$ and $S_3 \subseteq \{2N+1, \ldots, 3N\}$. We first observe that because of the multiplicativity of the characters χ_S and the definition of block-shifted Hadamard product, we have that for any $u, v \in \mathbb{R}^{2N}$,

$$\chi_S(u \diamond v) = \chi_{S_1}(u_1)\chi_{S_2}(v_1) \cdot \chi_{S_2}(u_2)\chi_{S_3}(v_2). \qquad (5.5)$$

We will treat $\chi_{S_1}(u_1)\chi_{S_2}(v_1)$ and $\chi_{S_2}(u_2)\chi_{S_3}(v_2)$ as function in the variables $u_1 = (u_1(i))_{i \in S_1}, v_1 = (v_1(j))_{j \in S_2}$ and $u_2 = (u_2(i))_{i \in S_2}, v_2 = (v_2(j))_{j \in S_3}$ respectively and write $\frac{\partial}{\partial u_1(i)}, \frac{\partial}{\partial v_1(j)}$ to denote the corresponding partial derivatives. To prevent any confusion, we clarify that $\partial_i = \frac{\partial}{\partial z_i}$ will always denote the derivative with respect to z.

Now, since (U_1, V_1) and (U_2, V_2) are independent Gaussians and they are being interpolated separately, we can apply the Gaussian interpolation formula separately to the functions of (U_1, V_1) and (U_2, V_2) appearing in (5.5). Since these functions are multilinear in the variables (u_1, v_1) and (u_2, v_2), any second order derivative with respect to the same variable is zero, so to apply Lemma 2.3, we only need to worry about the covariance between pairs of coordinates which are different. Moreover because of the covariance structure,

$\mathbb{E}[U_\kappa(i)U_\kappa(j)] = \mathbb{E}[V_\kappa(i)V_\kappa(j)] = 0$ for $i \neq j$, while $\mathbb{E}[U_\kappa(i)V_\kappa(j)] = \epsilon\mathsf{H}_{ij}$ for $i, j \in [N]$ and $\kappa \in [2]$. Therefore, applying Lemma 2.3 and using linearity of expectation, we have

$$\begin{aligned}
&4 \cdot \frac{\partial\varphi}{\partial t_1 \cdots \partial t_2}(t) \\
&= \sum_{\underline{i},\underline{j}} (\epsilon\mathsf{H}_{i_1,j_2}) \cdot (\epsilon\mathsf{H}_{i_2,j_3}) \cdot \mathbb{E}\left[\frac{\partial}{\partial u_1(i_1)\partial v_1(j_2)}\chi_{S_1}(\mathbf{U}_1)\chi_{S_2}(\mathbf{V}_1)\right] \cdot \mathbb{E}\left[\frac{\partial}{\partial u_2(i_2)\partial v_2(j_3)}\chi_{S_2}(\mathbf{U}_2)\chi_{S_3}(\mathbf{V}_2)\right] \\
&= \sum_{\underline{i},\underline{j}} (\epsilon\mathsf{H}_{i_1,j_2}) \cdot (\epsilon\mathsf{H}_{i_2,j_3}) \cdot \mathbb{E}[\chi_{S_1\setminus i_1}(\mathbf{U_1})\chi_{S_2\setminus j_2}(\mathbf{V_1})] \cdot \mathbb{E}[\chi_{S_2\setminus i_2}(\mathbf{U_2})\chi_{S_3\setminus j_3}(\mathbf{V_2})], \qquad (5.6)
\end{aligned}$$

writing $\underline{i} = (i_1, i_2) \in S_1 \times S_2$ and $\underline{j} = (j_2, j_3) \in S_2 \times S_3$. Note that the indices are shifted for $\underline{j}$ to clarify that they lie in the corresponding set S_r and we will keep using this indexing convention.

We can classify the terms in (5.6) into two types: terms where $i_2 = j_2$ and where $i_2 \neq j_2$. These behave very differently, and we bound their contributions separately.

(a) Terms where $i_2 = j_2$: In this case, defining $i_3 = j_3$, extending the tuple $\underline{i} = (i_1, i_2, i_3)$, the corresponding terms in (5.6) are given by

$$\begin{aligned}
&(\epsilon\mathsf{H}_{i_1,i_2})\cdot(\epsilon\mathsf{H}_{i_2,i_3}) \cdot \mathbb{E}[\chi_{S_1\setminus i_1}(\mathbf{U_1})\chi_{S_2\setminus i_2}(\mathbf{V_1} \odot \mathbf{U_2})\chi_{S_3\setminus i_3}(\mathbf{V_2})] \\
&\qquad = (\epsilon\mathsf{H}_{i_1,i_2}) \cdot (\epsilon\mathsf{H}_{i_2,i_3}) \cdot \mathbb{E}[\chi_{S\setminus\{i_1,i_2,i_3\}}(\mathbf{U} \diamond \mathbf{V})] = \theta_{\underline{i}} \cdot \left(\frac{\epsilon}{\sqrt{N}}\right)^2 \cdot \mathbb{E}[\chi_{S\setminus\{i_1,i_2,i_3\}}(\mathbf{U} \diamond \mathbf{V})],
\end{aligned}$$

where $\theta_{\underline{i}} = \mathsf{sign}(\mathsf{H}_{i_1,i_2} \cdot \mathsf{H}_{i_2,i_3})$. Viewing the tuple $\underline{i}$ as a set $J \subseteq S$ of size 3, this gives us that the sum of all the terms in (5.6) where $i_2 = j_2$ is exactly

$$\sum_{\substack{J\subseteq S\\|J|=3}} \theta_J \cdot \left(\frac{\epsilon}{\sqrt{N}}\right)^2 \cdot \mathbb{E}[\chi_{S\setminus J}(\mathbf{U} \diamond \mathbf{V})], \qquad (5.7)$$

where θ_J is a constant satisfying $|\theta_J| \leq 1$.

(b) Terms where $i_2 \neq j_2$: To bound these terms, we use Gaussian integration by parts to reduce them to sixth order derivatives. Consider a fixed term where $i_2 \neq j_2$. Then, the corresponding expectation term in (5.6) is

$$\mathbb{E}[\chi_{S_1\setminus i_1}(\mathbf{U_1})\chi_{S_2\setminus j_2}(\mathbf{V_1})] \cdot \mathbb{E}[\chi_{S_2\setminus i_2}(\mathbf{U_2})\chi_{S_3\setminus j_3}(\mathbf{V_2})]. \qquad (5.8)$$

As $i_2 \neq j_2$, the first expectation involving $(\mathbf{U}_1, \mathbf{V}_1)$ still depends on the random variable $\mathbf{V}_1(i_2)$ (while $\chi_{S_2\setminus i_2}(\mathbf{U_2})$ does not). Since eventually we need a function of $\mathbf{U}_2 \odot \mathbf{V}_1$, we pull out $\mathbf{V}_1(i_2)$ and write,

$$\begin{aligned}
\mathbb{E}[\chi_{S_{1\setminus i_1}}(\mathbf{U_1})\chi_{S_2\setminus j_2}(\mathbf{V_1})] &= \mathbb{E}\left[\mathbf{V}_1(i_2) \cdot \chi_{S_1\setminus i_1}(\mathbf{U_1})\chi_{S_2\setminus\{i_2,j_2\}}(\mathbf{V_1})\right] \\
&= \sum_{\substack{q_1\in S_1,\\ q_1\neq i_1}} (\epsilon\mathsf{H}_{q_1,i_2}t_1) \cdot \mathbb{E}\left[\frac{\partial}{\partial u_1(q_1)}\chi_{S_1\setminus i_1}(\mathbf{U_1})\chi_{S_2\setminus\{i_2,j_2\}}(\mathbf{V_1})\right] \\
&= \sum_{\substack{q_1\in S_1,\\ q_1\neq i_1}} (\epsilon\mathsf{H}_{q_1,i_2}t_1) \cdot \mathbb{E}[\chi_{S_1\setminus\{i_1,q_1\}}(\mathbf{U}_1)\chi_{S_2\setminus\{j_2,i_2\}}(\mathbf{V}_1)] \qquad (5.9)
\end{aligned}$$

where the second equality follows from Lemma 2.4 since $\mathbb{E}[\mathbf{U_1}(i)\mathbf{V_1}(j)] = \epsilon\mathsf{H}_{i,j}t_1$.

Analogously, the second expectation involving $\mathbf{U}_2, \mathbf{V}_2$ in (5.8) still depends on the random variable $\mathbf{U}_2(j_2)$ (while $\chi_{S_2\setminus j_2}(\mathbf{V_1})$ does not), and applying Gaussian integration by parts, one gets

$$\mathbb{E}[\chi_{S_{2\setminus i_2}}(\mathbf{U_2})\chi_{S_3\setminus j_3}(\mathbf{V_2})] \;=\; \sum_{\substack{q_3\in S_3,\\ q_3\neq j_3}} (\epsilon\mathsf{H}_{j_2,q_3}t_2) \cdot \mathbb{E}[\chi_{S_2\setminus\{j_2,i_2\}}(\mathbf{U}_2)\chi_{S_3\setminus\{j_3,q_3\}}(\mathbf{V}_2)] \qquad (5.10)$$

Combining (5.9) and (5.10), we get the sum of all the terms in (5.6) where $i_2 \neq j_2$. In particular, defining new tuples $\underline{\alpha} = (i_1, j_2, q_3) \in S_1 \times S_2 \times S_3$ and $\underline{\beta} = (q_1, i_2, j_3) \in S_1 \times S_2 \times S_3$, we get that the sum of the terms in (5.6) where $i_2 \neq j_2$ equals

$$
\begin{aligned}
&\sum_{\underline{\alpha},\underline{\beta}} \left(\frac{\epsilon}{\sqrt{N}}\right)^4 \theta_{\underline{\alpha},\underline{\beta}} \cdot t_1 t_2 \cdot \mathbb{E}[\chi_{S_1 \setminus \{i_1,q_1\}}(\mathbf{U}_1) \cdot \chi_{S_2 \setminus \{j_2,i_2\}}(\mathbf{V}_1) \cdot \chi_{S_2 \setminus \{j_2,i_2\}}(\mathbf{U}_2) \cdot \chi_{S_3 \setminus \{j_3,q_3\}}(\mathbf{V}_2)] \\
&= \sum_{\underline{\alpha},\underline{\beta}} \left(\frac{\epsilon}{\sqrt{N}}\right)^4 \theta_{\underline{\alpha},\underline{\beta}} \cdot t_1 t_2 \cdot \mathbb{E}[\chi_{S_1 \setminus \{i_1,q_1\}}(\mathbf{U}_1) \cdot \chi_{S_2 \setminus \{j_2,i_2\}}(\mathbf{V}_1 \odot \mathbf{U}_2) \cdot \chi_{S_3 \setminus \{j_3,q_3\}}(\mathbf{V}_2)] \\
&= \sum_{\underline{\alpha},\underline{\beta}} \left(\frac{\epsilon}{\sqrt{N}}\right)^4 \theta_{\underline{\alpha},\underline{\beta}} \cdot t_1 t_2 \cdot \mathbb{E}[\chi_{S \setminus \{i_1,q_1,i_2,j_2,j_3,q_3\}}(\mathbf{U} \diamond \mathbf{V})], \qquad (5.11)
\end{aligned}
$$

where $\theta_{\underline{\alpha},\underline{\beta}} = \mathsf{sign}(\mathsf{H}_{i_1,j_2}\mathsf{H}_{j_2,q_3}\mathsf{H}_{q_1,i_2}\mathsf{H}_{i_2,j_3})$ and the sum ranges over all possible tuples $\underline{\alpha}, \underline{\beta}$ where $\underline{\alpha}(r) \neq \underline{\beta}(r)$ for $r \in [3]$. Note that there are exactly 8 possible tuples $\underline{\alpha}, \underline{\beta}$ that give rise to the set $J = \{i_1, q_1, i_2, j_2, j_3, q_3\}$. It follows that the sum in (5.11) is exactly

$$
\sum_{\substack{J \subseteq S \\ |J|=6}} \left(\frac{\epsilon}{\sqrt{N}}\right)^4 \cdot \theta_J(t_1, t_2) \cdot \mathbb{E}[\chi_{S \setminus J}(\mathbf{U} \diamond \mathbf{V})], \qquad (5.12)
$$

where $\theta_J(t)$ is a polynomial for which the sum of the absolute values of all the coefficients is at most 8 and it only depends on J.

Then, plugging in the bounds from (5.7) and (5.11) for the two cases in (5.6), we get (5.4).

Proof for Arbitrary k

Let $S = S_1 \sqcup S_2 \sqcup \cdots \sqcup S_k$ where $S_j \subseteq \{(j-1)N+1, \ldots, jN\}$ for $r \in [k]$. As before, we first observe that because of the multiplicativity of the characters χ_S and the definition of block-shifted Hadamard product, we have that for any $u, v \in \mathbb{R}^{(k-1)N}$,

$$
\chi_S(u \diamond v) = \prod_{\kappa \in [k-1]} \chi_{S_\kappa}(u_\kappa)\chi_{S_{\kappa+1}}(v_\kappa),
$$

so it decomposes as a product of the $k-1$ functions from $\mathbb{R}^{2N}$ to $\mathbb{R}$ where the κ^{th} function is evaluated at (u_κ, v_κ). Let us treat them as functions in the variables $u_\kappa = (u_\kappa(i))_{i \in S_\kappa}$ and $v_\kappa = (v_\kappa(j))_{j \in S_{\kappa+1}}$ and write the derivatives as $\frac{\partial}{\partial u_\kappa(i)}, \frac{\partial}{\partial v_\kappa(j)}$.

Now, since (U_κ, V_κ) are independent Gaussians for different values of κ and they are being interpolated separately, we can apply the Gaussian interpolation formula separately to the functions of (U_κ, V_κ) appearing in (5.5). Since these functions are multilinear in the variables (u_κ, v_κ), any second order derivative with respect to the same variable is zero, so to apply Lemma 2.3, we only need to worry about the covariance between pairs of coordinates which are different. Moreover because of the covariance structure, $\mathbb{E}[U_\kappa(i)U_\kappa(j)] = \mathbb{E}[V_\kappa(i)V_\kappa(j)] = 0$ for $i \neq j$, while $\mathbb{E}[U_\kappa(i)V_\kappa(j)] = \epsilon\mathsf{H}_{ij}$ for $i, j \in [N]$ and $\kappa \in [k-1]$. Therefore, applying Lemma 2.3 and using linearity of expectation, we have

$$
\begin{aligned}
2^{k-1} \cdot \frac{\partial \varphi}{\partial t_1 \cdots \partial t_{k-1}}(t) &= \sum_{\underline{i},\underline{j}} \prod_{\kappa \in [k-1]} \epsilon\mathsf{H}_{i_\kappa, j_{\kappa+1}} \cdot \mathbb{E}\left[\frac{\partial}{\partial u_\kappa(i_\kappa)\partial v_\kappa(j_{\kappa+1})} \left(\chi_{S_\kappa}(\mathbf{U}_\kappa)\chi_{S_{\kappa+1}}(\mathbf{V}_\kappa)\right)\right], \\
&= \sum_{\underline{i},\underline{j}} \prod_{\kappa \in [k-1]} \epsilon\mathsf{H}_{i_\kappa, j_{\kappa+1}} \cdot \mathbb{E}[\chi_{S_\kappa \setminus i_\kappa}(\mathbf{U}_\kappa)\chi_{S_{\kappa+1} \setminus j_{\kappa+1}}(\mathbf{V}_\kappa)], \qquad (5.13)
\end{aligned}
$$

where $\underline{i} = (i_1, \cdots, i_{k-1}) \in S_1 \times \cdots \times S_{k-1}$ and $\underline{j} = (j_2, \cdots, j_k) \in S_2 \times \cdots \times S_k$ are tuples. As before, the indices are shifted for $\underline{j}$ to clarify that they lie in the corresponding set S_r.

Unlike the case of $k = 3$, there are many types of terms in the above summation. Some of them correspond to derivatives ∂_J of $\chi_S(z)$ where ∂_J is of order k, and others which do not correspond to any such derivative.

Terms that correspond to derivatives with respect to z. Observe that when tuples $\underline{i}$ and $\underline{j}$ satisfy $i_\ell = j_\ell$ for $2 \le \ell \le k-2$, then defining $i_k := j_k$ (this extends the $(k-1)$-tuple $\underline{i}$ to a k-tuple), the corresponding term in (5.13) is

$$
\begin{aligned}
&\prod\nolimits_{\kappa\in[k-1]} \epsilon \mathsf{H}_{i_\kappa,i_{\kappa+1}} \mathbb{E}[\chi_{S_\kappa\setminus i_\kappa}(\mathbf{U}_\kappa)\chi_{S_{\kappa+1}\setminus i_{\kappa+1}}(\mathbf{V}_\kappa)] \\
&= \Big(\prod\nolimits_{\kappa} \epsilon \mathsf{H}_{i_\kappa,i_{\kappa+1}}\Big) \cdot \mathbb{E}[\chi_{S_1\setminus i_1}(\mathbf{U}_1)\chi_{S_2\setminus i_2}(\mathbf{V}_1)\cdot\chi_{S_2\setminus i_2}(\mathbf{U}_2)\chi_{S_3\setminus i_3}(\mathbf{V}_2)\cdots\chi_{S_{k-1}\setminus i_{k-1}}(\mathbf{U}_{k-1})\chi_{S_k\setminus i_k}(\mathbf{V}_{k-1})] \\
&= \Big(\prod\nolimits_{\kappa} \epsilon \mathsf{H}_{i_\kappa,i_{\kappa+1}}\Big) \cdot \mathbb{E}[\chi_{S_1\setminus i_1}(\mathbf{U}_1)\chi_{S_2\setminus i_2}(\mathbf{V}_1\odot\mathbf{U}_2)\cdot\chi_{S_3\setminus i_3}(\mathbf{V}_2\odot\mathbf{U}_3)\cdots\chi_{S_k\setminus i_k}(\mathbf{V}_{k-1})] \\
&= \Big(\prod\nolimits_{\kappa} \epsilon \mathsf{H}_{i_\kappa,i_{\kappa+1}}\Big) \cdot \mathbb{E}[\chi_{S\setminus\{i_1,\ldots,i_k\}}(\mathbf{U}\diamond\mathbf{V}),] \qquad (5.14)
\end{aligned}
$$

by the definition of the $\diamond$ product. This corresponds to taking the partial derivative $\partial_J\chi_S(\mathbf{U}\diamond\mathbf{V}) = \chi_{S\setminus J}(\mathbf{U}\diamond\mathbf{V})$ for $J = \{i_1,\cdots,i_k\}$. However, the other terms in (5.13) can not be written as such a partial derivative. We will give a process that reduces such terms to a higher order derivative by repeated application of the Gaussian integration by parts identity.

Setup to apply Gaussian integration by parts. To describe the process, we will need some additional notation. Let us consider the terms appearing in (5.13) and drop the $\prod_{\kappa\in[k-1]}\epsilon\mathsf{H}_{i_\kappa,j_{\kappa+1}}$ scaling factor for notational convenience. By the independence of $(\mathbf{U}_r,\mathbf{V}_r)$ for different r, any term

$$
\prod_{r\in[k-1]} \mathbb{E}[\chi_{S_r\setminus i_r}(\mathbf{U}_r)\cdot\chi_{S_{r+1}\setminus j_{r+1}}(\mathbf{V}_r)] = \mathbb{E}\Big[\prod_{r\in[k-1]}\chi_{S_r\setminus i_r}(\mathbf{U}_r)\cdot\chi_{S_{r+1}\setminus j_{r+1}}(\mathbf{V}_r)\Big].
$$

Define the sets $A_1,\ldots,A_k$ and $B_1,\ldots,B_k$, where $A_r, B_r \subseteq S_r$ as follows:

$$
A_r = \{i_r\}, B_{r+1} = \{j_{r+1}\} \text{ for } r\in[k-1] \text{ and } A_k = \emptyset, B_1 = \emptyset. \qquad (5.15)
$$

Let us also define $\mathbf{V}_0 = \mathbf{U}_k = \mathbf{1}$ where $\mathbf{1}$ is the all ones vector in $\mathbb{R}^N$. Then the above can be written as

$$
\begin{aligned}
(5.16)\qquad &\mathbb{E}\Big[\prod_{r\in[k-1]}\chi_{S_r\setminus A_r}(\mathbf{U}_r)\cdot\chi_{S_{r+1}\setminus B_{r+1}}(\mathbf{V}_r)\Big] \\
&= \mathbb{E}\Big[\chi_{S_1\setminus A_1}(\mathbf{U}_1\odot\mathbf{V}_0)\Big(\prod_{r=2}^{k-1}\chi_{S_r\setminus A_r}(\mathbf{U}_r)\cdot\chi_{S_r\setminus B_r}(\mathbf{V}_{r-1})\Big)\cdot\chi_{S_k\setminus B_k}(\mathbf{U}_k\odot\mathbf{V}_{k-1})\Big].
\end{aligned}
$$

Let us write the middle term in the above expectation in a different way as follows,

$$
\Big(\prod_{r=2}^{k-1}\chi_{S_r\setminus A_r}(\mathbf{U}_r)\cdot\chi_{S_r\setminus B_r}(\mathbf{V}_{r-1})\Big) = \prod_{r=2}^{k-1}\chi_{S_r\setminus(A_r\cup B_r)}(\mathbf{U}_r\odot\mathbf{V}_{r-1})\cdot\chi_{B_r\setminus A_r}(\mathbf{U}_r)\cdot\chi_{A_r\setminus B_r}(\mathbf{V}_{r-1}).
$$

Writing $A = (A_1,\ldots,A_k)$ and $B = (B_1,\ldots,B_k)$, let us call (A,B) a *configuration*. We will always require a configuration to have $A_k = \emptyset$ and $B_1 = \emptyset$ and we call such configurations *valid*. For notational convenience, let us write $A\cup B$ to denote $(A_1\cup B_1)\cup\ldots\cup(A_k\cup B_k)$ for a configuration (A,B). Then, as $A_k = \emptyset, B_1 = \emptyset$, the expression in (5.16) can be written as

$$
\Gamma(S,A,B) := \mathbb{E}\Big[\chi_{S\setminus(A\cup B)}(\mathbf{U}\diamond\mathbf{V})\cdot\prod_{r=2}^{k-1}\Big(\chi_{B_r\setminus A_r}(\mathbf{U}_r)\cdot\chi_{A_r\setminus B_r}(\mathbf{V}_{r-1})\Big)\Big].
$$

Note that for $r \in \{2, \ldots, k-1\}$, the sets $B_r \setminus A_r$ (resp. $A_r \setminus B_r$) keep track of the excess $\mathbf{U}_r$ (resp. $\mathbf{V}_{r-1}$) variables that can not be absorbed in $\chi_{S\setminus(A\cup B)}(\mathbf{U} \diamond \mathbf{V})$.

For terminology, let us call a configuration *active* if there is some $r \in \{2, \ldots, k-1\}$ such that either $B_r \setminus A_r \neq \emptyset$ or $A_r \setminus B_r \neq \emptyset$. Any configuration that is not active, is referred to as being *inactive*. Note that the Γ value of any inactive configuration has the form $\mathbb{E}[\chi_{S\setminus J}(\mathbf{U} \diamond \mathbf{V})]$ for some $J \subseteq S$. In particular, inactive configurations correspond to derivatives $\partial_J \chi_S(z)$ evaluated at $\mathbf{U} \diamond \mathbf{V}$, for instance, the case of (5.14) corresponds to the inactive configuration with $J = \{i_1, i_2, \ldots, i_k\}$.

A Branching Process from Gaussian integration by parts. Given an initial active configuration (A, B), to compute $\Gamma(S, A, B)$, we will apply Gaussian integration by parts. Doing so will lead to several other terms of the same type with different configurations (A', B'), and we recursively continue this way until all resulting configurations are inactive. This can be viewed as a branching process, where starting from the configuration (A, B), we get a tree, where the leaves correspond to inactive configurations, and $\Gamma(S, A, B)$ is a weighted sum of the Γ values of the leaf configurations.

Below, we first describe the branching process and how the Γ values of the child nodes produced by one step of the process are related to the Γ value of the parent. After that, we will describe the properties of the leaf configurations generated by the process and relate it to the left hand side of (5.13).

(a) The branching process. Fix the set S, and consider an active configuration (A, B). Suppose that $B_q \setminus A_q \neq \emptyset$ for some $q \in \{2, \ldots, k-1\}$. Consider some arbitrary $i_q \in B_q \setminus A_q$. Then, we have the following key lemma.

Lemma 5.2. *For any $q \in \{2, \ldots, k-1\}$ and for any $i_q \in B_q \setminus A_q$, we have that*

$$\Gamma(S, A, B) = \sum_{j_{q+1} \in S_{q+1} \setminus B_{q+1}} \epsilon \mathsf{H}_{i_q, j_{q+1}} t_q \cdot \Gamma(S, A \cup \{i_q\}, B \cup \{j_{q+1}\})$$

where $A \cup \{i_q\}$ is obtained from A by setting $A_q = A_q \cup \{i_q\}$, and $B \cup \{j_{q+1}\}$ from B by setting $B_{q+1} = B_{q+1} \cup \{j_{q+1}\}$.

Before giving a proof, we remark that it may be useful to interpret the above lemma in the following way: any configuration $(A', B') = (A \cup \{i_q\}, B \cup \{j_{q+1}\})$ that appears on the right hand side above, absorbs excess variables $\mathbf{U}_q(i_q)$ and $\mathbf{V}_q(j_{q+1})$ into the $\chi_{S\setminus(A'\cup B')}(\mathbf{U} \diamond \mathbf{V})$ term, but might add an excess variable $\mathbf{U}_{q+1}(j_{q+1})$ in the $\chi_{B_{q+1}\setminus A_{q+1}}(\mathbf{U}_{q+1})$ term depending on whether $j_{q+1} \in A_{q+1}$ or not. If an excess variable is not added (when $j_{q+1} \in A_{q+1}$), we call it a type I transition, otherwise (when $j_{q+1} \notin A_{q+1}$) we call it a type II transition.

Proof of Lemma 5.2. Writing $\chi_{B_q\setminus A_q}(\mathbf{U}_q) = \mathbf{U}_q(i_q) \cdot \chi_{B_q\setminus(A_q\cup\{i_q\})}(\mathbf{U}_q)$ and applying Gaussian integration by parts (Lemma 2.4) gives that

$$\begin{aligned} (5.17) \quad \Gamma(S, A, B) &= \mathbb{E}\Big[\chi_{S\setminus(A\cup B)}(\mathbf{U} \diamond \mathbf{V}) \cdot \prod_{r=2}^{k-1} \Big(\chi_{B_r\setminus A_r}(\mathbf{U}_r) \cdot \chi_{A_r\setminus B_r}(\mathbf{V}_{r-1})\Big)\Big] \\ &= \sum_{j_{q+1} \in S_{q+1}\setminus B_{q+1}} \mathbb{E}[\mathbf{U}_q(i_q)\mathbf{V}_q(j_{q+1})] \cdot \mathbb{E}\Big[\frac{\partial}{\partial v_q(j_{q+1})}\Big(\chi_{S\setminus(A'\cup B)}(\mathbf{U} \diamond \mathbf{V}) \cdot \prod_{r=2}^{k-1}\Big(\chi_{B_r\setminus A'_r}(\mathbf{U}_r) \cdot \chi_{A'_r\setminus B_r}(\mathbf{V}_{r-1})\Big)\Big)\Big] \end{aligned}$$

where we denote $A' = A \cup \{i_q\}$. In the summation above, we only need to consider $j_{q+1} \in S_{q+1}$ as $\mathbf{U}_q(i_q)$ has non-zero correlation only with coordinates of $\mathbf{V}_q$ and with $\mathbf{U}_q(i_q)$. However, the $\mathbb{E}[\mathbf{U}_q(i_q)^2]$ term does not contribute to the above sum as the partial derivative with respect to $u_q(i_q)$ is identically zero as $u_q(i_q)$ does not appear in

$$\chi_{S\setminus(A'\cup B)}(u \diamond v) = \prod_{r=1}^{k} \chi_{S_r\setminus(A'_r\cup B_r)}(u_r \odot v_{r-1})$$

or in $\chi_{B_q\setminus A'_q}(\mathbf{U}_q)$ as $i_q \in A'_q$. Furthermore, we can further restrict $j_{q+1} \in S_{q+1} \setminus B_{q+1}$ in the summation (5.17) above, since for $j_{q+1} \in B_{q+1}$, the derivative with respect to $v_q(j_{q+1})$ is identically zero as $v_q(j_{q+1})$ does not appear in either $\chi_{A'_{q+1}\setminus B_{q+1}}(v_q)$ or $\chi_{S\setminus(A'\cup B)}(u \diamond v)$.

Simplifying further, we have $\mathbb{E}[\mathbf{U}_q(i_q)\mathbf{V}_q(j_{q+1})] = \epsilon \mathsf{H}_{i_q,j_{q+1}} t_q$. Next, the expectation containing the derivative simplifies as follows.

Claim 5.3. *For $j_{q+1} \in S_{q+1} \setminus B_{q+1}$, we have that*

$$\mathbb{E}\Bigg[\frac{\partial}{\partial v_q(j_{q+1})}\Bigg(\chi_{S\setminus(A'\cup B)}(\mathbf{U}\diamond\mathbf{V}) \cdot \prod_{r=2}^{k-1}\Big(\chi_{B_r\setminus A'_r}(\mathbf{U}_r)\cdot\chi_{A'_r\setminus B_r}(\mathbf{V}_{r-1})\Big)\Bigg)\Bigg] = \Gamma(S, A', B \cup \{j_{q+1}\}).$$

The statement of Lemma 5.2 follows from the claim above and (5.17). We now prove Claim 5.3.

Proof of Claim 5.3. We have two cases, depending on whether $v_q(j_{q+1})$ appears in $\chi_{A'_{q+1}\setminus B_{q+1}}(v_q)$, which happens if $j_{q+1} \in A'_{q+1}$, or whether $v_q(j_{q+1})$ appears in $\chi_{S\setminus(A'\cup B)}(u\diamond v)$, which happens if $j_{q+1} \notin A'_{q+1}$.

Suppose first that $j_{q+1} \in A'_{q+1}$. As $v_q(j_{q+1})$ appears in $\chi_{A'_{a+1}\setminus B_{q+1}}(v_q)$, upon taking the derivative this term becomes $\chi_{A'_{q+1}\setminus(B_{q+1}\cup\{j_{q+1}\})}(v_q)$. Upon setting $B' = B \cup \{j_{q+1}\}$, note that the other terms $\chi_{S\setminus(A'\cup B')}(u\diamond v)$ and $\chi_{B'_{q+1}\setminus A'_{q+1}}(u_q)$ remain unchanged as $j_{q+1} \in A'_{q+1}$. It follows that

$$\begin{aligned}(5.18)\qquad &\mathbb{E}\left[\frac{\partial}{\partial v_q(j_{q+1})}\left(\chi_{S\setminus(A'\cup B)}(\mathbf{U}\diamond\mathbf{V}) \cdot \prod_{r=2}^{k-1}\Big(\chi_{B_r\setminus A'_r}(\mathbf{U}_r)\cdot\chi_{A'_r\setminus B_r}(\mathbf{V}_{r-1})\Big)\right)\right]\\ = \quad &\mathbb{E}\left[\chi_{S\setminus(A'\cup B')}(\mathbf{U}\diamond\mathbf{V}) \cdot \prod_{r=2}^{k-1}\Big(\chi_{B'_r\setminus A'_r}(\mathbf{U}_r)\cdot\chi_{A'_r\setminus B'_r}(\mathbf{V}_{r-1})\Big)\right] = \Gamma(S,A',B') = \Gamma(S,A',B\cup\{j_{q+1}\}).\end{aligned}$$

Now we consider the more interesting case where $v_q(j_{q+1})$ appears in the term

$$\chi_{S\setminus(A'\cup B)}(u\diamond v) = \chi_{S_{q+1}\setminus(A'_{q+1}\cup B_{q+1})}(u_{q+1}\odot v_q)\cdot \prod_{\substack{r\in[k]\\ r\neq q+1}} \chi_{S_r\setminus(A'_r\cup B_r)}(u_r\odot v_{r-1}).$$

Consider the term $\chi_{S_{q+1}\setminus(A'_{q+1}\cup B_{q+1})}(u_{q+1}\odot v_q)$ appearing in the above expression. Upon taking the derivative $\frac{\partial}{\partial v_q(j_{q+1})}$ this becomes

$$u_{q+1}(j_{q+1})\cdot\chi_{S_{q+1}\setminus(A'_{q+1}\cup B_{q+1}\cup\{j_{q+1}\})}(u_{q+1}\odot v_q),$$

which has the extra factor $u_{q+1}(j_{q+1})$. However, setting $B' = B\cup\{j_{q+1}\}$, this $u_{q+1}(j_{q+1})$ factor is absorbed in $\chi_{B'_{q+1}\setminus A'_{q+1}}(u_{q+1})$ as $j_{q+1} \notin A'_{q+1}$ by our assumption. Finally, note that this does not affect $\chi_{A'_{q+1}\setminus B'_{q+1}}(v_q)$. It follows that (5.18) holds in this case as well. ■

This completes the proof of Lemma 5.2. ■

A completely analogous lemma holds if $i_q \in A_q \setminus B_q$, and Gaussian integration by parts is applied with respect to the variable $V_{q-1}(i_q)$. In particular we have the following.

Lemma 5.4. *For any $q \in \{2,\ldots,k-1\}$ and for any $i_q \in A_q \setminus B_q$, we have that*

$$\Gamma(S,A,B) = \sum_{j_{q-1}\in S_{q-1}\setminus A_{q-1}} \epsilon\mathsf{H}_{i_q,j_{q-1}} t_{q-1}\cdot\Gamma(S, A\cup\{j_{q-1}\}, B\cup\{i_q\})$$

where $A\cup\{j_{q-1}\}$ is obtained from A by setting $A_{q-1} = A_{q-1}\cup\{j_{q-1}\}$, and $B\cup\{i_q\}$ from B by setting $B_q = B_q\cup\{i_q\}$.

Similar to Lemma 5.2, it may be useful to interpret the above lemma as follows: any configuration $(A', B') = (A \cup \{j_{q-1}\}, B \cup \{i_q\})$ that appears on the right hand side above, absorbs excess variables $\mathbf{U}_{q-1}(j_{q-1})$ and $\mathbf{V}_{q-1}(i_q)$ into the $\chi_{S\setminus(A'\cup B')}(\mathbf{U} \diamond \mathbf{V})$ term, but might add an excess variable $\mathbf{V}_{q-1}(j_{q-1})$ in the $\chi_{A_{q-1}\setminus B_{q-1}}(\mathbf{V}_{q-1})$ term depending on whether $j_{q-1} \in B_{q-1}$ or not. If an excess variable is not added (when $j_{q-1} \in B_{q-1}$), we call it a type I transition, otherwise (when $j_{q-1} \notin B_{q-1}$) we call it a type II transition.

Also, we remark that in both Lemma 5.2 and Lemma 5.4 above, the resulting configurations A', B' still satisfy $A'_k = B'_1 = \emptyset$ and hence are valid. In particular, as $q \in \{2, \ldots, k-1\}$, neither A_k or B_1 are ever updated in either of the lemmas.

Finally, note that Lemma 5.2 and Lemma 5.4 take different actions — Lemma 5.2 chooses an $i_q \in B_q \setminus A_q$ and uses an application of Gaussian integration by parts using the variable $U_q(i_q)$; on the other hand, Lemma 5.4 chooses an $i_q \in A_q \setminus B_q$ and applies Gaussian integration by parts using the variable $V_{q-1}(i_q)$. However, both Lemma 5.2 and Lemma 5.4 allow us to express the value $\Gamma(S, A, B)$ for a configuration (A, B) as a weighted sum of Γ values of other configurations. Hence, given a starting configuration (A, B) and applying Lemma 5.2 and Lemma 5.4 alternately gives the claimed branching process. Note that a branch of the process terminates at a leaf configuration which is inactive. We remark that the same configuration may appear multiple times as different nodes of the branching tree, but we will treat each node of the branching tree as a separate configuration.

(b) Properties of the leaf configurations. We next show some properties of the configurations that arise in the branching process where each initial configuration is given by (5.15). Note that an initial configuration (A, B) corresponds uniquely to a tuple $\underline{i}, \underline{j}$ appearing in (5.13). We define the weight of an initial configuration (A, B) that corresponds to the tuple $\underline{i}, \underline{j}$ as

$$\mathsf{wt}(S, A, B) := \prod_{\kappa \in [k-1]} \epsilon \mathsf{H}_{i_\kappa, j_{\kappa+1}}.$$

Note that left hand side of (5.13) equals the weighted sum of Γ values of all the initial configurations where the weights are given by the wt values. For each initial configuration, we will start a separate branching process in parallel, and we will always maintain the invariant that the left hand side of (5.13) always equals the weighted sum of the Γ values of all the configurations generated at any intermediate step. To do this we define the weight of each node in one such branching tree in the following way: if a configuration $(A \cup \{i_q\}, B \cup \{j_{q+1}\})$ is a child of (A, B) in the branching tree, then the weight of $(A \cup \{i_q\}, B \cup \{j_{q+1}\})$ is defined as

$$\mathsf{wt}(S, A \cup \{i_q\}, B \cup \{j_{q+1}\}) = \epsilon \mathsf{H}_{i_q, j_{q+1}} t_q \cdot \mathsf{wt}(S, A, B),$$

where $\epsilon \mathsf{H}_{i_q, j_{q+1}} t_q$ is the factor appearing in front of $\Gamma(S, A \cup \{i_q\}, B \cup \{j_{q+1}\})$ in Lemma 5.2 or Lemma 5.4. Note that the weight of a node in the branching tree depends on the path from the initial configuration to that node in the tree.

The following proposition is an immediate consequence of the definition of weight and of Lemma 5.2 and Lemma 5.4.

Proposition 5.5. *Let $\mathcal{L}(S)$ denote the collection of all the leaf configurations generated by all parallel branching processes (viewed as a multiset). Then the left hand side of* (5.13) *equals*

$$\sum_{(A,B) \in \mathcal{L}(S)} \mathsf{wt}(S, A, B) \cdot \Gamma(S, A, B).$$

Recall that for any final inactive leaf configuration (A, B), we have that $\Gamma(S, A, B) = \mathbb{E}[\chi_{S\setminus J}(\mathbf{U} \diamond \mathbf{V})]$ where $J = A \cup B$. This is exactly the derivative $\partial_J \chi_S(z)$ evaluated at $\mathbf{U} \diamond \mathbf{V}$. Next, towards bounding (5.13), we compute the contribution of each leaf configuration (A, B) in terms of these derivatives. Note that the same J may correspond to multiple leaf configurations (A, B) appearing with potentially different weights.

Contribution of a Leaf Configuration. Given a fixed S, let $(A^{(0)}, B^{(0)})$ denote some initial configuration, and consider some path in the branching tree starting from $(A^{(0)}, B^{(0)})$ and ending in $(A^{(T)}, B^{(T)})$. Consider a step on this path where the configuration changes from $(A^{(\tau)}, B^{(\tau)})$ to $(A^{(\tau+1)}, B^{(\tau+1)})$ and also recall that in either application of Lemma 5.2 or Lemma 5.4 at each step, there are two types of transitions — either of type I or type II. We note that if $(A^{(\tau+1)}, B^{(\tau+1)})$ is derived from a type I transition, then $|A^{(\tau+1)} \cup B^{(\tau+1)}| = |A^{(\tau)} \cup B^{(\tau)}|$ while if it is derived from a type II transition, then $|A^{(\tau+1)} \cup B^{(\tau+1)}| = |A^{(\tau)} \cup B^{(\tau)}| + 1$.

The following lemma shows that each leaf configuration corresponds to a derivative of order mk where $m \in [k-1]$ and also gives a bound on the contribution of each leaf configuration towards (5.13).

Lemma 5.6. *Consider the branching tree started at the initial configuration $(A^{(0)}, B^{(0)})$. For any inactive leaf configuration $(A^{(T)}, B^{(T)})$ in this branching tree, we have that $|A^{(T)} \cup B^{(T)}| = mk$ where $m \in [k-1]$. Moreover, in this case $T = (m-1)(k-1)$ and there exists a sign $\theta \in \{\pm 1\}$ and a monomial $\zeta(t) = \zeta(t_1, \ldots, t_{k-1}) = \prod_{\kappa \in [k-1]} t_\kappa^{\alpha_\kappa}$ of total degree T such that*

$$\mathsf{wt}(S, A^{(T)}, B^{(T)}) = \theta \cdot \zeta(t_1, \ldots, t_{k-1}) \cdot \left(\frac{\epsilon}{\sqrt{N}}\right)^{m(k-1)},$$

where θ and ζ depend only on the path from $(A^{(0)}, B^{(0)})$ to $(A^{(T)}, B^{(T)})$ in the branching tree.

We remark that the monomial $\zeta(t_1, \ldots, t_{k-1})$ can be taken to be $\prod_{\kappa \in [k-1]} t_\kappa^{m-1}$ by a more careful analysis. Since, it is not needed, we do not prove the statement with the exact form of ζ.

Proof of Lemma 5.6. In the initial configuration $(A^{(0)}, B^{(0)})$, recall that $|A_r^{(0)}| = |B_{r+1}^{(0)}| = 1$ for $r \in [k-1]$ and $|A_k^{(0)}| = |B_1^{(0)}| = 0$. Let β denote the number of blocks $r \in \{2, \ldots, k-1\}$ for which $|A_r^{(0)} \setminus B_r^{(0)}| = 0$ (or equivalently $|A_r^{(0)} \cup B_r^{(0)}| = 1$.

For a configuration (A, B), consider the following potential which nicely captures many properties of the dynamics of the configurations generated by the branching process.

$$\Phi(A, B) = |A \cup B| + \sum_{r=2}^{k-1} (k-r)|B_r \setminus A_r| + \sum_{r=2}^{k-1} (r-1)|A_r \setminus B_r|.$$

Claim 5.7. *For an initial configuration, $\Phi(A^{(0)}, B^{(0)}) = k(k-1-\beta)$.*

Proof. We compute each of the terms in the potential. First

$$|A^{(0)} \cup B^{(0)}| = 1 + 2(k-2-\beta) + \beta + 1 = 2k - 2 - \beta$$

as $|A_1^{(0)} \cup B_1^{(0)}| = |A_k^{(0)} \cup B_k^{(0)}| = 1$, and for $r \in \{2, \ldots, k-1\}$, we have $|A_r^{(0)} \cup B_r^{(0)}| = 1$ for β indices and 2 for the remaining $k-2-\beta$ indices.

We now consider the terms $(k-r)|B_r \setminus A_r| + (r-1)|A_r \setminus B_r|$ for $r \in \{2, \ldots, k-1\}$. This contributes exactly $(k-r) + (r-1) = k-1$ whenever $B_r^{(0)} \neq A_r^{(0)}$, which happens for $k-2-\beta$ indices, and contributes 0 for all other indices. This gives

$$\Phi(A^{(0)}, B^{(0)}) = 2k - 2 - \beta + (k-2-\beta)(k-1) = k(k-1-\beta). \qquad \blacksquare$$

Next, we show how Φ evolves along any edge of the branching tree.

Claim 5.8. *Consider any transition $(A^{(\tau)}, B^{(\tau)})$ to $(A^{(\tau+1)}, B^{(\tau+1)})$. If this is type I transition then the potential decreases by exactly k, otherwise for a type II transition the potential remains unchanged.*

Proof. We first consider the type I transition, and consider the setting of Lemma 5.2, where $i_q \in B_q \setminus A_q$. Then, $|A^{(\tau)} \cup B^{(\tau)}|$ does not change and $|B_q \setminus A_q|$ and $|A_{q+1} \setminus B_{q+1}|$ both decrease by exactly 1 (as i_q is added to A_q and j_{q+1} to B_{q+1} where $j_{q+1} \in A_{q+1}$). Thus the potential change is

$$\Phi(A^{(\tau+1)}, B^{(\tau+1)}) - \Phi(A^{(\tau)}, B^{(\tau)}) = -(k-q) - (q+1-1) = -k$$

An exactly analogous argument works for type I transition corresponding to the setting of Lemma 5.4.

For type II transition, again consider the setting of Lemma 5.2. Then $|A^{(\tau+1)} \cup B^{(\tau+1)}| = |A^{(\tau)} \cup B^{(\tau)}| + 1$. As j_{q+1} is added to $B^{(\tau)}_{q+1}$ (and j_{q+1} is not in $A^{(\tau)}_{q+1}$), the quantity $|B^{(\tau)}_{q+1} \setminus A^{(\tau)}_{q+1}|$ increases by 1, and as i_q is added to $A^{(\tau)}_q$ (and i_q is in $B^{(\tau)}_q \setminus A^{(\tau)}_q$ before it is added to $A^{(\tau)}_q$) the quantity $|B^{(\tau)}_{q+1} \setminus A^{(\tau)}_{q+1}|$ decreases by 1. As the coefficient of these terms in the potential is $k-(q+1)$ and $k-q$ respectively, overall we have that

$$\Phi(A^{(\tau+1)}, B^{(\tau+1)}) - \Phi(A^{(\tau)}, B^{(\tau)}) = 1 + (k-(q+1)) - (k-q) = 0$$

The setting of Lemma 5.4 is exactly analogous. ■

Finally, for an inactive configuration as $|A_r \setminus B_r| = |B_r \setminus A_r| = 0$ for $r \in \{2, \dots, k-1\}$, and hence, the value of the potential is exactly $|A \cup B|$.

We can now prove the first result in the statement of the lemma. Consider some path from $(A^{(0)}, B^{(0)})$ to $(A^{(T)}, B^{(T)})$, and let λ_1 denote the number of type I transitions. Then by the properties above,

$$|A^{(T)} \cup B^{(T)}| = \Phi(A^{(T)}, B^{(T)}) = \Phi(A^{(0)}, B^{(0)}) - \lambda_1 k = k(k-1-\beta) - \lambda_1 k = k(k-1-\beta-\lambda_1) = mk$$

where $m = k-1-\beta-\lambda_1$ is a positive integer as the potential Φ is always non-negative.

Next we express T in terms of other parameters and show that $T = (m-1)(k-1)$. For this, we note that $|A \cup B|$ rises by 1 exactly for $T - \lambda_1$ steps (at type II transitions). As $|A^{(0)} \cup B^{(0)}| = 2k-2-\beta$ initially, we have that $T - \lambda_1 = mk - (2k-2-\beta)$. Therefore, since $m = k-1-\beta-\lambda_1$ from the first part of the lemma, we can express

$$T = k(k-1-\beta-\lambda_1) - (2k-2-\beta) + \lambda_1 = (k-1)(k-2-\beta-\lambda_1) = (k-1)(m-1).$$

Finally, we bound the weight of the leaf configuration. Note the weight of the initial configuration $(A^{(0)}, B^{(0)})$ consists of a product of $k-1$ terms of the form $\epsilon H_{i_\kappa, j_{\kappa+1}}$ where $\kappa \in [k-1]$. By the definition of weight of a child node, at each step of the branching process, we gain exactly one $\epsilon H_{i_q, j_{q+1}} t_q$ factor in the weight. It follows that the weight of the leaf configuration $(A^{(T)}, B^{(T)})$ equals

$$\mathsf{wt}(A^{(T)}, B^{(T)}) = \theta \cdot \zeta(t_1, \dots, t_{k-1}) \cdot \left(\frac{\epsilon}{\sqrt{N}}\right)^{T+(k-1)}, \tag{5.19}$$

where θ is the sign of corresponding products of Hadamard entries and $\zeta(t) = \zeta(t_1, \dots, t_{k-1}) = \prod_{\kappa \in [k-1]} t_\kappa^{\alpha_\kappa}$ is a monomial of total degree T.

Plugging in the value of T in (5.19) gives us the statement of the lemma regarding the weight. ■

Bounding the Total Contribution. For any leaf configuration (A, B), we have that $\Gamma(S, A, B) = \mathbb{E}[\chi_{S \setminus J}(\mathbf{U} \diamond \mathbf{V})]$ where $J = A \cup B$. Therefore, using Proposition 5.5 and Lemma 5.6, we get that the left hand side in (5.13) equals

$$\sum_{m=1}^{k-1} \left(\frac{\epsilon}{\sqrt{N}}\right)^{m(k-1)} \cdot \sum_{\substack{J \subseteq S \\ |J| = mk}} \theta_{S,J}(t) \cdot \mathbb{E}[\chi_{S \setminus J}(\mathbf{U}(t) \diamond \mathbf{V}(t))], \tag{5.20}$$

where $\theta_{S,J}(t) = \theta_{S,J}(t_1, \cdots, t_{k-1})$ is a polynomial which is determined by the collection of paths in all the branching trees that lead to a leaf configuration $(A, B) \in \mathcal{L}(S)$ satisfying $A \cup B = J$. Moreover, the sum of the absolute value of all the coefficients of $\theta_{S,J}(t)$ is bounded by the number of such paths.

To finish the proof, we argue that $\theta_{S,J}$ only depends on J and we also bound how many leaf configurations correspond to the set J via an encoding argument.

Lemma 5.9. *$\theta_{S,J}(t)$ depends only on J and not on S. Moreover, there are at most $(4k)^{4|J|}$ leaf configurations (A,B) in the multiset $\mathcal{L}(S)$ for which $A \cup B = J$ and hence, the sum of the absolute value of all the coefficients of $\theta_J(t) := \theta_{S,J}(t)$ is at most $(4k)^{4|J|}$.*

Proof. Let us write $J = J_1 \sqcup J_2 \sqcup \ldots \sqcup J_k$ where $J_r \subseteq S_r$ for each $r \in [k]$. First, we note that the branching process only adds elements to the configuration and never removes them. Therefore, all the paths that lead to leaf configurations (A,B) satisfying $A \cup B = J$ can only contain intermediate configurations (A', B') where the sets $A'_r, B'_r \subseteq J_r$. Moreover, since there is a branching tree for all initial configurations $(A^{(0)}, B^{(0)})$ satisfying $|A_r^{(0)}| = |B_{r+1}^{(0)}| = 1$ for $r \in [k-1]$ and $|A_k^{(0)}| = |B_1^{(0)}| = 0$ where $A_r, B_r \in S_r$, it follows that to determine the collection of paths that lead to a leaf configuration $(A,B) \in \mathcal{L}(S)$ satisfying $A \cup B = J$, we can assume without any loss of generality that $S_r = J_r$ for every $r \in [k]$. This proves that $\theta_{S,J}(t)$ only depends on J and not on S as it is determined by this collection of paths.

Next, we bound the number of paths in this collection. We will describe an encoding that stores at most $\log_2((4k)^{4|J|})$ bits and uniquely determines the entire path of the branching process from the initial configuration $(A^{(0)}, B^{(0)})$ to the final leaf configuration $(A^{(T)}, B^{(T)})$ for which $A^{(T)} \cup B^{(T)} = J$. From this, it follows that the number of leaf configurations (A,B) for which $A \cup B = J$ is at most $(4k)^{4|J|}$.

For the encoding, we initialize a bit-string of length $2|J|$ to the indicator vectors of the initial configuration $(A^{(0)}, B^{(0)})$. We only need $2|J| = 2\sum_{r=1}^{k} |J_r|$ bits as we only need to store subsets of each J_r. This bit-string will be updated at every step of the branching process, along with some auxiliary information.

At time $\tau \in [T]$, the configuration is updated from $(A^{(\tau-1)}, B^{(\tau-1)})$ to $(A^{(\tau)}, B^{(\tau)})$ using either Lemma 5.2 or Lemma 5.4. In each case, for exactly one $r \in \{2, \cdots, k-1\}$, A_r is updated to $A_r \cup \{i_r\}$ and B_{r+1} is updated to $B_{r+1} \cup \{j_{r+1}\}$. To reconstruct this information, we store the following:

- We update the two $|J|$-length bit-strings to store the indicator vectors of the configuration $A^{(\tau)}, B^{(\tau)}$ at time τ. This requires changing a zero bit to a one bit in each of the two bit-strings as we only ever add elements to the sets.

- We also store the indices of the two locations where the above bit-strings were updated. This requires exactly $2\lceil \log_2 |J| \rceil$ bits. Note that which set A_r or $B_{r'}$ was updated is also determined by the indices. Moreover, we record the $2\lceil \log_2 |J| \rceil$ indices in order, so the exact time τ when the bits were written is also determined by this information.

Overall, given the above information, one can uniquely determine the exact path from $(A^{(0)}, B^{(0)})$ to $(A^{(T)}, B^{(T)})$. The total number of bits of information is at most $2|J| + T(2 + 2\lceil \log_2 J \rceil)$.

Now, from Lemma 5.6, it follows that $|J| = mk$ where $m \in [k-1]$ and also $T = (m-1)(k-1) \leq |J|$, so the total number of bits information is at most

$$2|J|\log_2 |J| + 6|J| \leq 4|J|\log_2 k + 6|J| = \log_2(2^{6|J|} k^{4|J|}) \leq \log_2((4k)^{4|J|}). \qquad \blacksquare$$

Using the above lemma in conjunction with (5.20) completes the proof of Lemma 5.1 for an arbitrary k.

5.3 Fourier Weight of Decision Trees under Biased Measures

We use a random restriction argument to prove Theorem 3.4. Recall the basic notation about random restrictions introduced in Section 2.2.

Theorem 3.4. *Let $f : \{\pm 1\}^m \to \mathbb{R}$ and $\ell \in [m]$. Let w be such that for any restriction $\rho \in \{-1, 1, \star\}^m$, we have $\mathsf{wt}_\ell(f_\rho) \leq w$ where the Fourier weight is with respect to the uniform measure. Then, for any $\mu \in [-1/2, 1/2]^m$, we have $\mathsf{wt}_\ell^\mu(f) \leq 4^\ell w$.*

Proof. Define the following product distribution over restrictions $\rho \in \{-1, 1, \star\}^m$,

$$\rho_i = \begin{cases} \star & \text{with probability } \ (1-\mu_i^2)/2 = \sigma_i^2/2, \\ 1 & \text{with probability } \ (1+\mu_i)^2/4, \\ -1 & \text{with probability } \ (1-\mu_i)^2/4. \end{cases}$$

For notational convenience, let us abbreviate $\sigma_W = \prod_{i \in W} \sigma_i$ and $\mu_W = \prod_{i \in W} \mu_i$ for $W \subseteq [m]$. Then, the Fourier coefficient of f_ρ under the uniform measure and of f under the bias μ are related by the following claim.

Claim 5.10. *Let $S \subseteq [m]$. Then, we have $\mathbb{E}[\hat{f}_\rho(S)] = 2^{-|S|}\sigma_S \cdot \hat{f}^\mu(S)$ where the expectation is taken over ρ.*

Given the above claim, we can finish the proof of Corollary 3.5 as follows. Let us define $\mathsf{wt}_\ell^\mu(f, \theta) := \sum_{|S|=\ell} \theta_S \hat{f}^\mu(S)$ for any sequence of signs $\theta := (\theta_S)_{|S|=\ell}$. Then, using Claim 5.10 and taking expectation over ρ, we obtain

$$\begin{aligned}\mathsf{wt}_\ell^\mu(f, \epsilon) = \sum_{|S|=\ell} \theta_S \cdot 2^\ell \sigma_S^{-1} \cdot \mathbb{E}[\hat{f}_\rho(S)] &= \mathbb{E}\Bigg[\sum_{|S|=\ell} \theta_S \cdot 2^\ell \sigma_S^{-1} \cdot \hat{f}_\rho(S)\Bigg] \\ &\le \mathbb{E}\Bigg[\sum_{|S|=\ell} 2^\ell \sigma_S^{-1} \cdot |\hat{f}_\rho(S)|\Bigg] \le 4^\ell \cdot \mathbb{E}[\mathsf{wt}_\ell(f_\rho)] \le 4^\ell w.\end{aligned}$$

The second last inequality above follows since $\sigma_i = \sqrt{1-\mu_i^2} \ge 1/2$ as $\mu \in [-1/2, 1/2]^m$, and the last inequality uses our assumption on the Fourier weight of the restricted function f_ρ under the uniform measure. Since the above is true for an arbitrary sequence of signs θ, it follows that

$$\mathsf{wt}_\ell^\mu(f) \le 4^\ell w.$$

This finishes the proof assuming Claim 5.10 which we prove next. ■

Proof of Claim 5.10. We note that for any subset $S \subseteq [m]$,

$$\hat{f}_\rho(S) = \sum_{T : T \supseteq S} \hat{f}(T) \cdot \mathbf{1}[S \subseteq \mathsf{free}(\rho) \text{ and } T \setminus S \subseteq \mathsf{fix}(\rho)] \cdot \chi_{T \setminus S}(\rho).$$

Taking expectation over the random restriction ρ, we get that

$$\mathbb{E}[\hat{f}_\rho(S)] = \sum_{T : T \supseteq S} \hat{f}(T) \cdot \prod_{i \in S} \frac{\sigma_i^2}{2} \cdot \prod_{i \in T \setminus S} \left(\frac{(1+\mu_i)^2}{4} - \frac{(1-\mu_i)^2}{4}\right) = \sum_{T : T \supseteq S} \hat{f}(T) \cdot 2^{-|S|}\sigma_S^2 \cdot \mu_{T \setminus S}. \tag{5.21}$$

Next, recalling that the Fourier basis with respect to bias μ is given by $\phi_S(x) = \prod_{i \in S} \frac{(x_i - \mu_i)}{\sigma_i}$, we have

$$\hat{f}^\mu(S) = \mathbb{E}_{p_\mu}[f(X)\phi_S(X)] = \sum_{T \subseteq [m]} \hat{f}(T) \cdot \sigma_S^{-1} \cdot \mathbb{E}_{p_\mu}\Big[\chi_T(X) \cdot \prod_{i \in S}(X_i - \mu_i)\Big].$$

Since $\mathbb{E}_{p_\mu}[X_i] = \mu_i$, it follows that all the terms above where $S \setminus T$ is not the empty set are zero. Moreover, $\mathbb{E}_{p_\mu}[X_i(X_i - \mu_i)] = 1 - \mu_i^2 = \sigma_i^2$. Therefore,

$$\hat{f}^\mu(S) = \sum_{T : T \supseteq S} \hat{f}(T) \cdot \sigma_S \cdot \mu_{T \setminus S}. \tag{5.22}$$

Comparing (5.21) and (5.22) gives us the claim. ■

5.4 Proof of Main Lower Bound: Theorem 1.3 and Corollary 1.4

Given Corollary 3.5 and Theorem 3.2, the proof is straightforward.

Proof of Theorem 1.3. Given a randomized decision tree of depth d that has advantage γ, we first amplify the success probability of the decision tree to $1 - \delta$, by making $\tau = \Theta(\gamma^{-2} \log(1/\delta))$ repetitions and taking the majority vote. Since the error of this randomized decision of $\Theta(d\tau)$ depth is at most δ on each valid input, we have that for large enough N,

$$\left|\mathbb{E}_{p_1}[f(Z)] - f(0)\right| \geq 6\delta - \frac{\delta^2}{4N} - 2\delta \geq \delta, \tag{5.23}$$

because of Theorem 3.1.

Next, we will show a contradiction to the above statement if the depth d was too small. In particular, applying Theorem 3.2 and Corollary 3.5 to the decision tree of depth $d_1 = \Theta(d\tau)$, we obtain

$$\begin{aligned}\left|\mathbb{E}_{p_1}[f(Z)] - f(0)\right| - N^{-k} &\leq \sup_{\mu\in[-1/2,1/2]^{kN}} \sum_{m=1}^{k-1} \left(\frac{\epsilon}{\sqrt{N}}\right)^{m(k-1)} \cdot (8k)^{4mk} \cdot \mathsf{wt}^{\mu}_{mk}(f) \\ &\leq \sum_{m=1}^{k-1} \left(\frac{\epsilon}{\sqrt{N}}\right)^{m(k-1)} \cdot (8k)^{4mk} \cdot \sqrt{d_1^{mk} \log^{km-1}(kN)} \\ &\leq \sum_{m=1}^{k-1} \left(cd_1k^8 \cdot \left(\frac{\epsilon^2 \log(kN)}{N}\right)^{1-1/k}\right)^{mk/2},\end{aligned}$$

for a universal constant c. Thus, if

$$d_1 < c_1 \cdot \frac{\epsilon^{1/k}}{k^8} \cdot \left(\frac{N}{\log(kN)}\right)^{1-1/k} < C \cdot \frac{N^{1-1/k}}{k^8 \log(kN)},$$

for suitable constants c_1 and C, where the second inequality holds by our choice of $\epsilon = \frac{1}{64k^2 \log(kN)}$, then the advantage of the decision tree on the input distribution $p(Z)$ is at most $\epsilon^k/4 = \delta/4$ which contradicts (5.23). This gives us that d must be at least $\Omega\left(\frac{N^{1-1/k}}{\tau k^8 \log(kN)}\right)$ giving us the bound in the statement of the theorem after substituting the value of τ. ■

Corollary 1.4 can be obtained analogous to the above.

Acknowledgements

We thank Ronald de Wolf for discussions throughout the course of this work and for providing helpful feedback on the writing. We also thank Avishay Tal for very useful comments.

References

[AA18] Scott Aaronson and Andris Ambainis. Forrelation: A problem that optimally separates quantum from classical computing. *SIAM J. Comput.*, 47(3):982–1038, 2018.

[Aar10] Scott Aaronson. BQP and the Polynomial Hierarchy. STOC '10, page 141–150, New York, NY, USA, 2010.

[ABB$^+$17] Andris Ambainis, Kaspars Balodis, Aleksandrs Belovs, Troy Lee, Miklos Santha, and Juris Smotrovs. Separations in query complexity based on pointer functions. *J. ACM*, 64(5), September 2017.

[ABK16] Scott Aaronson, Shalev Ben-David, and Robin Kothari. Separations in query complexity using cheat sheets. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*, page 863–876, 2016.

[ABKT20] Scott Aaronson, Shalev Ben-David, Robin Kothari, and Avishay Tal. Quantum implications of huang's sensitivity theorem. *Electronic Colloquium on Computational Complexity (ECCC)*, 27:66, 2020.

[BCW02] J. Niel de Beaudrap, Richard Cleve, and John Watrous. Sharp Quantum versus Classical Query Complexity Separations. *Algorithmica*, 34(4):449–461, 2002.

[BFNR08] Harry Buhrman, Lance Fortnow, Ilan Newman, and Hein Röhrig. Quantum property testing. *SIAM Journal on Computing*, 37(5):1387–1400, 2008.

[BV97] Ethan Bernstein and Umesh Vazirani. Quantum complexity theory. *SIAM Journal on Computing*, 26(5):1411–1473, 1997.

[BW02] Harry Buhrman and Ronald de Wolf. Complexity measures and decision tree complexity: a survey. *Theoretical Computer Science*, 288(1):21 – 43, 2002. Complexity and Logic.

[CCD$^+$03] Andrew M. Childs, Richard Cleve, Enrico Deotto, Edward Farhi, Sam Gutmann, and Daniel A. Spielman. Exponential algorithmic speedup by a quantum walk. STOC '03, page 59–68, 2003.

[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.

[Gro96] Lov K. Grover. A fast quantum mechanical algorithm for database search. STOC '96, page 212–219, 1996.

[GRZ20] Uma Girish, Ran Raz, and Wei Zhan. Lower Bounds for XOR of Forrelations. *CoRR*, abs/2007.03631, 2020.

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.

[RT19] Ran Raz and Avishay Tal. Oracle separation of BQP and PH. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019*, pages 13–23. ACM, 2019.

[Sho97] Peter W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.*, 26(5):1484–1509, October 1997.

[Sim97] Daniel R. Simon. On the power of quantum computation. *SIAM J. Comput.*, 26(5):1474–1483, October 1997.

[Tal11] Michel Talagrand. *Mean Field Models for Spin Glasses, Volume I: Basic Examples*. Springer-Verlag Berlin Heidelberg, 2011.

[Tal19] Avishay Tal. Towards optimal separations between quantum and randomized query complexities. *CoRR*, abs/1912.12561, 2019. To appear in FOCS '20.

[Wu20] Xinyu Wu. A stochastic calculus approach to the oracle separation of BQP and PH. *CoRR*, abs/2007.02431, 2020.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il