



Lower Bounds for XOR of Forrelations

Uma Girish  

Department of Computer Science, Princeton University, NJ, USA

Ran Raz  

Department of Computer Science, Princeton University, NJ, USA

Wei Zhan  

Department of Computer Science, Princeton University, NJ, USA

Abstract

The Forrelation problem, first introduced by Aaronson [1] and Aaronson and Ambainis [2], is a well studied computational problem in the context of separating quantum and classical computational models. Variants of this problem were used to give tight separations between quantum and classical query complexity [2]; the first separation between poly-logarithmic quantum query complexity and bounded-depth circuits of super-polynomial size, a result that also implied an oracle separation of the classes BQP and PH [15]; and improved separations between quantum and classical communication complexity [12]. In all these separations, the lower bound for the classical model only holds when the advantage of the protocol (over a random guess) is more than $\approx 1/\sqrt{N}$, that is, the success probability is larger than $\approx 1/2 + 1/\sqrt{N}$. This is unavoidable as $\approx 1/\sqrt{N}$ is the correlation between two coordinates of an input that is sampled from the Forrelation distribution, and hence there are simple classical protocols that achieve advantage $\approx 1/\sqrt{N}$, in all these models.

To achieve separations when the classical protocol has smaller advantage, we study in this work the XOR of k independent copies of (a variant of) the Forrelation function (where $k \ll N$). We prove a very general result that shows that any family of Boolean functions that is closed under restrictions, whose Fourier mass at level $2k$ is bounded by α^k (that is, the sum of the absolute values of all Fourier coefficients at level $2k$ is bounded by α^k), cannot compute the XOR of k independent copies of the Forrelation function with advantage better than $O\left(\frac{\alpha^k}{N^{k/2}}\right)$. This is a strengthening of a result of [8], that gave a similar statement for $k = 1$, using the technique of [15]. We give several applications of our result. In particular, we obtain the following separations:

Quantum versus Classical Communication Complexity. We give the first example of a partial Boolean function that can be computed by a simultaneous-message quantum protocol with communication complexity $\text{polylog}(N)$ (where Alice and Bob also share $\text{polylog}(N)$ EPR pairs), and such that, any classical randomized protocol of communication complexity at most $\tilde{o}(N^{1/4})$, with any number of rounds, has quasipolynomially small advantage over a random guess. Previously, only separations where the classical protocol has polynomially small advantage were known between these models [10, 12].

Quantum Query Complexity versus Bounded Depth Circuits. We give the first example of a partial Boolean function that has a quantum query algorithm with query complexity $\text{polylog}(N)$, and such that, any constant-depth circuit of quasipolynomial size has quasipolynomially small advantage over a random guess. Previously, only separations where the constant-depth circuit has polynomially small advantage were known [15].

2012 ACM Subject Classification Theory of computation \rightarrow Communication complexity; Theory of computation \rightarrow Pseudorandomness and derandomization; Theory of computation \rightarrow Oracles and decision trees

Keywords and phrases Forrelation, Quasipolynomial, Separation, Quantum versus Classical, Xor

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2021.52

Category RANDOM



© Uma Girish, Ran Raz, and Wei Zhan;

licensed under Creative Commons License CC-BY 4.0

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2021).

Editors: Mary Wootters and Laura Sanità; Article No. 52; pp. 52:1–52:14



Leibniz International Proceedings in Informatics

LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Related Version *Full Version*: <https://arxiv.org/abs/2007.03631>

Full Version: <https://eccc.weizmann.ac.il/report/2020/101/>

Funding *Uma Girish*: Simons Collaboration on Algorithms and Geometry, by a Simons Investigator Award and by the National Science Foundation grants No. CCF-1714779, CCF-2007462.

Ran Raz: Simons Collaboration on Algorithms and Geometry, by a Simons Investigator Award and by the National Science Foundation grants No. CCF-1714779, CCF-2007462.

Wei Zhan: Simons Collaboration on Algorithms and Geometry, by a Simons Investigator Award and by the National Science Foundation grants No. CCF-1714779, CCF-2007462.

Acknowledgements We would like to thank Avishay Tal for very helpful conversations. We would also like to thank Chin Ho Lee for pointing out a simpler proof of Lemma 28.

1 Introduction

Several recent works used Fourier analysis to prove lower bounds for computing (variants of) the Forrelation (partial) function of [1, 2], in various models of computation and communication [15, 8, 12]. These works show that for many computational models, when analyzing the success probability of computing the Forrelation function, it's sufficient to bound the contribution of Fourier coefficients at level 2, ignoring all other Fourier coefficients [15, 8]. This holds for any computational model that is closed under restrictions and is proved by analyzing the Forrelation distribution as a distribution resulting from a certain random walk, rather than analyzing it directly.

While this is a powerful technique, it could only be used to bound computations of the Forrelation function with advantage (over a random guess) larger than $\approx 1/\sqrt{N}$, that is, computations with success probability larger than $\approx 1/2 + 1/\sqrt{N}$. Roughly speaking, this is because the bound on the Fourier coefficients at level 2 of the Forrelation function is $\approx O(1/\sqrt{N})$. We note that while ruling out protocols with advantage larger than $1/\sqrt{N}$ is satisfactory in many cases, an advantage of $1/\sqrt{N}$ is often viewed as non-negligible and it is often desirable to rule out protocols with negligible (sub-polynomially small) advantage as well.

In this work, we study the XOR of k independent copies of the Forrelation function of [15] (where $k < o(N^{1/50})$). We show that for many computational models, when analyzing the success probability of computing the XOR of k independent copies of the Forrelation function, it's sufficient to bound the contribution of Fourier coefficients at level $2k$, ignoring all other Fourier coefficients. Our proof builds on the techniques of [15], and followup works [8, 12], by analyzing a “product” of k random walks, one for each of the independent copies of the Forrelation function. This can be viewed as a random walk with a k -dimensional time variable.

Consequently, we obtain a very general lower bound that shows that any family of Boolean functions that is closed under restrictions, whose Fourier mass at level $2k$ is bounded by α^k (that is, for every function in the family, the sum of the absolute values of all Fourier coefficients at level $2k$ is bounded by α^k), cannot compute the XOR of k independent copies of the Forrelation function with advantage better than $O\left(\frac{\alpha^k}{N^{k/2}}\right)$, that is, with success probability larger than $\frac{1}{2} + O\left(\frac{\alpha^k}{N^{k/2}}\right)$. This is a strengthening of a result of [8], that gave a similar statement for $k = 1$, using the technique of [15]. While bounding the advantage of protocols for the XOR of k independent copies of a problem is often non-trivial, our result gives a very general way to do that for the special case of Forrelation.

We note that the requirement that the family of Boolean functions is closed under restrictions is satisfied by essentially all non-uniform computational models. The requirement of having a good bound on the Fourier mass at level $2k$ is satisfied by several central and well-studied computational models (see for example [7] for a recent discussion). In particular, we focus in this work on three such models: communication complexity, query complexity (decision trees) and bounded-depth circuits. We note that our result is valid for any $k < N^c$, for some constant $c > 0$, and hence it can be used to prove lower bounds for circuits/protocols with exponentially small advantage, in all these models. However, for the applications of separating quantum and classical computational models, we take k to be poly-logarithmic in N , so that we have quantum protocols of poly-logarithmic cost. We use our main theorem to give several separations between quantum and classical computational models.

1.1 Communication Complexity

Quantum versus classical separations in communication complexity have been studied for more than two decades in numerous works. We briefly summarize the history of quantum advantage in communication complexity of partial functions, that is most relevant for us: First, Buhrman, Cleve and Wigderson proved an exponential separation between zero-error simultaneous-message quantum communication complexity (without entanglement) and classical deterministic communication complexity [4]. For the bounded-error model, Raz showed an exponential separation between two-way quantum communication complexity and two-way randomized communication complexity [14]. Gavinsky et al (building on Bar-Yossef et al [3]) gave an exponential separation between one-way quantum communication complexity and one-way randomized communication complexity [11]. Klartag and Regev gave an exponential separation between one-way quantum communication complexity and two-way randomized communication complexity [16]. The state of the art separation, by Gavinsky, gave an exponential separation between simultaneous-message quantum communication complexity (with entanglement) and two-way randomized communication complexity [10]. An alternative proof for Gavinsky's result was recently given by [12], as a followup to [15, 8], and had the additional desired property that in the quantum protocol, the time complexity of all the players is poly-logarithmic.

Our Result

In all these works, the lower bounds for classical communication complexity only hold when the advantage of the protocol (over a random guess) is more than $\approx 1/\sqrt{N}$, that is, the success probability is larger than $\approx 1/2 + 1/\sqrt{N}$.

In this work, we give a partial Boolean function that can be computed by a simultaneous-message quantum protocol with communication complexity $\text{polylog}(N)$ (where Alice and Bob also share $\text{polylog}(N)$ EPR pairs), and such that, any classical randomized protocol of communication complexity at most $\tilde{o}(N^{1/4})$, with any number of rounds, has quasipolynomially small advantage over a random guess. This qualitatively matches the results of [10, 12] and has the additional desired property that the lower bound for the classical communication protocol holds for quasipolynomially small advantage, rather than polynomially small advantage. Moreover, as in [12], the quantum protocol in our upper bound has the additional property of being *efficiently implementable*, in the sense that it can be described by quantum circuits of size $\text{polylog}(N)$, with oracle access to the inputs.

To prove this result we use the XOR of k independent copies of the Forrelation function, lifted to communication complexity using XOR as the gadget [13], as in [12]. The quantum upper bound is simple. For the classical lower bound, we use ideas from [12] to bound the

level- $2k$ Fourier mass. This, along with our main theorem implies the desired separation. Our bounds for the level- $2k$ Fourier mass may be interesting in their own right and are proved in Section 7.

Related Work

We note that an exponential separation between **two-way** quantum communication complexity and two-way randomized communication complexity, with quasipolynomially small advantage, can be proved by a combination of several previous results, as follows:

Start with an existing separation between quantum and classical query complexity, such as the one of [2]. Use Drucker's XOR lemma for randomized decision tree [9] to get a separation between quantum and classical query complexity, where the classical protocol has quasipolynomially small advantage. Finally, use the recent lifting theorem of [5] to lift the result to communication complexity. To the best of our knowledge, this separation was not previously observed.

It follows from these works that there exists a function computable in the quantum two-way model in communication complexity $\text{polylog}(N)$, for which randomized protocols of cost $\tilde{o}(\sqrt{N})$ have at most quasipolynomially small advantage. While the lower bound is for cost $\tilde{o}(\sqrt{N})$ protocols, which is quantitatively stronger than our lower bound for cost $\tilde{o}(N^{1/4})$ protocols, the quantum upper bound in this result seems to require two rounds of communication, while our function is computable in the simultaneous model when Alice and Bob share entanglement.

1.2 Bounded Depth Circuits

Separations of quantum query complexity and bounded-depth classical circuit complexity have been studied in the context of oracle separations of the classes BQP and PH. An example of a partial Boolean function (Forrelation) that has a quantum query algorithm with query complexity $\text{polylog}(N)$, and such that, any constant-depth circuit of quasipolynomial size has polynomially small advantage over a random guess, was given in [15]. This result implied an oracle separation of the classes BQP and PH.

Here, we give the first example of a partial Boolean function (XOR of k copies of Forrelation) that has a quantum query algorithm with query complexity $\text{polylog}(N)$, and such that, any constant-depth circuit of quasipolynomial size has **quasipolynomially** small advantage over a random guess.

For the proof, we use our main theorem, together with Tal's bounds on the level- $2k$ Fourier mass of bounded-depth circuits [17].

1.3 Decision Trees

The query complexity model (also known as black box model or decision-tree complexity) has played a central role in the study of quantum computational complexity. Quantum advantages in query complexity (decision trees) have been demonstrated for partial functions in various settings and numerous works. For example, Aaronson and Ambainis [2] showed that the Forrelation problem can be solved by one quantum query, while its randomized query complexity is $\Omega(\sqrt{N}/\log N)$.

For classical randomized query complexity, there is a known XOR lemma, proved by Drucker [9]. In particular, Theorem 1.3 of [9], along with the result of [2] gives a partial function (XOR of $\text{polylog}(N)$ copies of Forrelation) that can be computed by a quantum query algorithm with $\text{polylog}(N)$ queries, while every classical randomized algorithm that makes $\tilde{o}(N^{1/2})$ queries, has quasipolynomially small advantage.

Our main theorem implies a different proof for this result, using Tal’s recent bounds on the level- $2k$ Fourier mass of decision trees [18].

1.4 The Main Theorem

Our functions are obtained by taking an XOR of several copies of a variant of the Forrelation problem, as defined in [15].

Let $N = 2^n$ for sufficiently large $n \in \mathbb{N}$. Let $k \in \mathbb{N}$ be a parameter. We assume that $k = o(N^{1/50})$. Let $\epsilon = \frac{1}{60k^2 \ln N}$ be a parameter.

Let H_N denote the $N \times N$ normalized Hadamard matrix whose entries are either $-\frac{1}{\sqrt{N}}$ or $\frac{1}{\sqrt{N}}$. Let

$$f_{\text{orr}}(z) := \frac{1}{N} \langle z_2, H_N z_1 \rangle$$

denote the *Forrelation* of a vector $z = (z_1, z_2)$, where $z_1, z_2 \in \mathbb{R}^N$. The **Forrelation Decision Problem** is the partial Boolean function $F : \{-1, 1\}^{2N} \rightarrow \{-1, 1\}$ defined at $z \in \{-1, 1\}^{2N}$ by

$$F(z) := \begin{cases} -1 & \text{if } f_{\text{orr}}(z) \geq \epsilon/2 \\ 1 & \text{if } f_{\text{orr}}(z) \leq \epsilon/4 \\ \text{undefined} & \text{otherwise} \end{cases}$$

The \oplus^k **Forrelation Decision Problem** $F^{(k)} : \{-1, 1\}^{2kN} \rightarrow \{-1, 1\}$ is defined as the XOR of k independent copies of F . More precisely, for every $z_1, \dots, z_k \in \{-1, 1\}^{2N}$, let

$$F^{(k)}(z_1, \dots, z_k) := \prod_{j=1}^k F(z_j).$$

For our separation results, we take the function $F^{(k)}$, where $k = \lceil \log^2 N \rceil$. For our communication complexity separation we take the lift of $F^{(k)}$ with XOR as the gadget. The quantum upper bounds in all these separation results are quite simple. Moreover, all the quantum algorithms in our upper bounds have the additional advantage of being *efficiently implementable*, in the sense that they can be described by quantum circuits of size $\text{polylog}(N)$, with oracle access to the inputs.

Our main contribution is the classical lower bound. Towards this, our main theorem provides an upper bound on the maximum correlation of $F^{(k)}$ with any family of Boolean functions, in terms of the maximum level- $2k$ Fourier mass of a function in the family.

► **Main Theorem (Informal).** *There exist two distributions, $\sigma_0^{(k)}$ and $\sigma_1^{(k)}$, on the NO and YES instances of $F^{(k)}$, respectively, with the following property. Let \mathcal{H} be a family of Boolean functions, each of which maps $\{-1, 1\}^{2kN}$ into $[-1, 1]$. Assume that \mathcal{H} is closed under restrictions. For $H \in \mathcal{H}$, let $L_{2k}(H) := \sum_{|S|=2k} |\widehat{H}(S)|$. Let $\alpha \in \mathbb{R}$ be such that $\alpha^k := \sup_{H \in \mathcal{H}} (L_{2k}(H), 1)$. Then, for every $H \in \mathcal{H}$,*

$$\left| \mathbb{E}_{z \sim \sigma_0^{(k)}} [H(z)] - \mathbb{E}_{z \sim \sigma_1^{(k)}} [H(z)] \right| \leq O\left(\frac{\alpha^k}{N^{k/2}}\right)$$

Our main theorem implies that functions in \mathcal{H} cannot correlate with $F^{(k)}$ by more than $\frac{1}{2} + O\left(\frac{\alpha^k}{N^{k/2}}\right)$. For the applications, we instantiate \mathcal{H} with the class of functions computed by classical protocols of small cost.

1.5 Overview of Proof of the Main Theorem for $k = 2$

Our proof builds on the techniques of [15], and followup works [8, 12], which, in turn, used a key idea from [7]. We will now give an overview of the proof of the Main Theorem for the special case $k = 2$, where one can already see most of the key ideas.

We start by recalling the hard distributions for $k = 1$, as in [15]. The **distribution \mathcal{U} on no instances** of F is the uniform distribution U_{2N} on $\{-1, 1\}^{2N}$. It can be shown that a bit string drawn uniformly at random almost always has low Forrelation. The **distribution \mathcal{G} on yes instances** of F is the Gaussian distribution with mean 0 and covariance matrix $\epsilon \begin{bmatrix} \mathbb{I}_N & H_N \\ H_N & \mathbb{I}_N \end{bmatrix}$. It can be shown that a vector drawn from this distribution almost always has high Forrelation (at least $\epsilon/2$). Although \mathcal{G} is not a distribution over $\{-1, 1\}^{2N}$, this can be fixed (by probabilistically rounding the values) and we ignore this issue in the proof overview.

Our hard distributions for $k \geq 2$ are obtained by naturally lifting these distributions. The **distribution μ_0 on no instances** of $F^{(2)}$ is $\frac{1}{2}(\mathcal{U} \times \mathcal{U} + \mathcal{G} \times \mathcal{G})$. The **distribution μ_1 on yes instances** is $\frac{1}{2}(\mathcal{U} \times \mathcal{G} + \mathcal{G} \times \mathcal{U})$. It can be shown that these distributions indeed have almost all their mass on the YES and NO instances of $F^{(2)}$, respectively.

Throughout this proof, we identify functions in \mathcal{H} with their unique multilinear extensions. Using this identification, it follows that for all $H \in \mathcal{H}$ and $z_0 \in \mathbb{R}^{4N}$, we have $\mathbb{E}_{z \sim \mathcal{U}}[H(z_0 + (z, 0))] = \mathbb{E}_{z \sim \mathcal{U}}[H(z_0 + (0, z))] = \mathbb{E}_{z \sim \mathcal{U}^2}[H(z_0 + z)] = H(z_0)$.

Bounding the Advantage of H in Distinguishing $p \cdot \mu_0$ and $p \cdot \mu_1$, for Small p

As in [15, 8], in order to show that functions in \mathcal{H} can't distinguish between μ_0 and μ_1 , we first show that they can't distinguish between $p \cdot \mu_0$ and $p \cdot \mu_1$, for small p . We show that for every $H \in \mathcal{H}$, and $p \leq \frac{1}{2N}$,

$$\begin{aligned} \left| \mathbb{E}_{z \sim p \cdot \mu_0} [H(z)] - \mathbb{E}_{z \sim p \cdot \mu_1} [H(z)] \right| &\triangleq \frac{1}{2} \left| \mathbb{E}_{\substack{z_1 \sim p \cdot \mathcal{G} \\ z_2 \sim p \cdot \mathcal{G}}} [H(z_1, z_2) - H(z_1, 0) - H(0, z_2) + H(0, 0)] \right| \\ &\leq p^4 \cdot O\left(\frac{L_4(H)}{N}\right) + O(p^6 N^{1.5}) \end{aligned}$$

This claim is analogous to Claim 20 from [8]. For sufficiently small p , the second term in the R.H.S. of the inequality is negligible, compared to the first term. To prove this inequality, we use the Fourier expansion of H in the L.H.S. and bound the difference between the moments of $p \cdot \mu_0$ and $p \cdot \mu_1$. We show that $p \cdot \mu_0$ and $p \cdot \mu_1$ agree on moments of degree less than 4, so these moments don't contribute to the difference. We then show that the contribution of the moments of degree 4 is $L_4(H) \cdot O\left(\frac{p^4}{N}\right)$ and the contribution of moments of higher degrees is $O(p^6 N^{1.5})$.

Bounding the Advantage of $H(z_0 + z)$ in Distinguishing $p \cdot \mu_0$ and $p \cdot \mu_1$, for Small p

Next, as in [15, 8], we show a similar statement for the function $H(z_0 + z)$ of z , where z_0 is not too large. We show that for every $H \in \mathcal{H}$, and every $z_0 \in [-1/2, 1/2]^{2kN}$ and $p \leq \frac{1}{2N}$,

$$\begin{aligned} & \frac{1}{2} \left| \mathbb{E}_{\substack{z_1 \sim p \cdot \mathcal{G} \\ z_2 \sim p \cdot \mathcal{G}}} [H(z_0 + (z_1, z_2)) - H(z_0 + (z_1, 0)) - H(z_0 + (0, z_2)) + H(z_0)] \right| \\ & \leq p^4 \cdot O\left(\frac{L_4(H)}{N}\right) + O(p^6 N^{1.5}) \end{aligned} \tag{1}$$

The proof of this inequality is similar to the proof of Claim 19 of [8], using key ideas from [7], and relies on the multilinearity of functions in \mathcal{H} and the closure of \mathcal{H} under restrictions.

A Random Walk with Two-Dimensional Time Variable

This is the main place where our proof differs from the one of [15] and followup works [8, 12]. In all these works the Forrelation distribution was ultimately analyzed as the distribution obtained by a certain random walk. Here, we consider a product of two random walks, which can also be viewed as a random walk with two-dimensional time variable.

Let $T = 16N^4$ and $p = \frac{1}{\sqrt{T}}$. Let $z_1^{(1)}, z_2^{(1)}, \dots, z_1^{(T)}, z_2^{(T)} \sim p \cdot \mathcal{G}$ be independent samples. Let $t = (t_1, t_2)$ for $t_1, t_2 \in \{0, \dots, T\}$. Let $z^{\leq(t)} := \left(\sum_{i=1}^{t_1} z_1^{(i)}, \sum_{i=1}^{t_2} z_2^{(i)}\right)$. Note that $z^{\leq(t)}$ is distributed according to $(p\sqrt{t_1} \cdot \mathcal{G}) \times (p\sqrt{t_2} \cdot \mathcal{G})$. In particular, $z^{\leq(T,T)}$ is distributed according to $\mathcal{G} \times \mathcal{G}$. This implies that

$$(*) := \mathbb{E}_{z \sim \mu_0} [H(z)] - \mathbb{E}_{z \sim \mu_1} [H(z)] \triangleq \frac{1}{2} \mathbb{E} \left[H(z^{\leq(T,T)}) - H(z^{\leq(T,0)}) - H(z^{\leq(0,T)}) + H(0,0) \right]$$

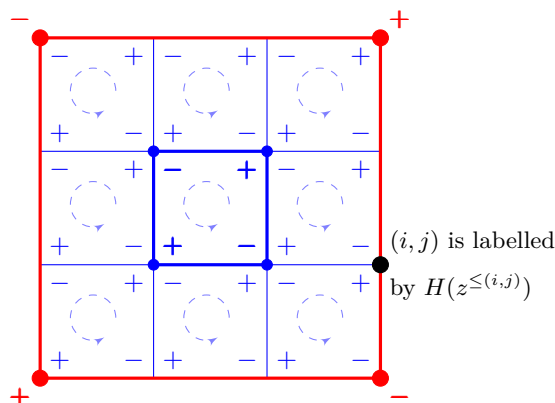
We now rewrite (*) as follows.

$$(*) = \frac{1}{2} \sum_{\substack{t_1 \in [T] \\ t_2 \in [T]}} \mathbb{E} \left[H(z^{\leq(t_1, t_2)}) - H(z^{\leq(t_1-1, t_2)}) - H(z^{\leq(t_1, t_2-1)}) + H(z^{\leq(t_1-1, t_2-1)}) \right] \tag{2}$$

The last equation follows by a two-dimensional telescopic cancellation, as depicted in Figure 1. This turns out to be a powerful observation. Note that for every fixed $t = (t_1, t_2)$, the random variable $z^{\leq(t)} - z^{\leq(t-(1,1))} \triangleq (z_1^{(t_1)}, z_2^{(t_2)})$ is distributed according to $p \cdot \mathcal{G}^2$, by construction. We can thus apply Inequality(1), setting $z_0 = z^{\leq(t-(1,1))}$. This, along with the Triangle-Inequality implies that

$$\begin{aligned} |(*)| & \leq \frac{1}{2} \sum_{\substack{t_1 \in [T] \\ t_2 \in [T]}} \left| \mathbb{E} \left[H(z^{\leq(t_1, t_2)}) - H(z^{\leq(t_1-1, t_2)}) - H(z^{\leq(t_1, t_2-1)}) + H(z^{\leq(t_1-1, t_2-1)}) \right] \right| \\ & \leq \frac{1}{2} \sum_{\substack{t_1 \in [T] \\ t_2 \in [T]}} \left(p^4 \cdot O\left(\frac{L_4(H)}{N}\right) + O(p^6 N^{1.5}) \right) \quad \text{by Inequality (1)} \\ & = O\left(\frac{L_4(H)}{N}\right) + o\left(\frac{1}{N}\right) \quad \text{since } T = 16N^4 = \frac{1}{p^2} \end{aligned}$$

This completes the proof overview for $k = 2$, albeit with many details left out.



■ **Figure 1** Consider the $(T + 1) \times (T + 1)$ grid whose vertices are indexed by $v \in (\{0\} \cup [T])^2$. Each vertex v is labelled by $H(z^{\leq(v)})$. Each rectangle has a sign on its vertices as defined in Figure 1 and the label of a rectangle is the sum of signed labels of its vertices. The sum of labels of all 1×1 rectangles equals the label of the larger $T \times T$ rectangle. This is exactly the content of Equation (2).

1.6 Organization of the Paper

In the appendix, we present a formal description of our main results. The proofs can be found in the full version of the paper.

1.7 Related Work

Independently of our result, [6] demonstrated PRGs with polylogarithmic dependence on seed length, for a large class of boolean functions. Their result builds on the framework of [7, 8, 15] and constructs improved PRGs by leveraging level- k Fourier bounds.

References

- 1 Scott Aaronson. BQP and the Polynomial Hierarchy. In *STOC 2010*, 2010.
- 2 Scott Aaronson and Andris Ambainis. Forrelation: A Problem That Optimally Separates Quantum from Classical Computing. In *STOC 2015*, 2015.
- 3 Ziv Bar-Yossef, T. S. Jayram, and Iordanis Kerenidis. Exponential Separation of Quantum and Classical One-Way Communication Complexity. In *STOC 2004*, 2004.
- 4 Harry Buhrman, Richard Cleve, and Avi Wigderson. Quantum vs. Classical Communication and Computation. In *STOC 1998*, 1998.
- 5 Arkadev Chattopadhyay, Yuval Filmus, Sajin Koroth, Or Meir, and Toniann Pitassi. Query-To-Communication Lifting for BPP Using Inner Product. In *ICALP 2019*, 2019.
- 6 Eshan Chattopadhyay, Jason Gaitonde, Chin Ho Lee, Shachar Lovett, and Abhishek Shetty. Fractional Pseudorandom Generators from Any Fourier Level. *CoRR*, abs/2008.01316, 2020.
- 7 Eshan Chattopadhyay, Pooya Hatami, Kaave Hosseini, and Shachar Lovett. Pseudorandom Generators from Polarizing Random Walks. In *CCC 2018*, 2018.
- 8 Eshan Chattopadhyay, Pooya Hatami, Shachar Lovett, and Avishay Tal. Pseudorandom Generators from the Second Fourier Level and Applications to AC0 with Parity Gates. In *ITCS 2019*, 2019.
- 9 Andrew Drucker. Improved Direct Product Theorems for Randomized Query Complexity. In *CCC 2011*, 2011.
- 10 Dmitry Gavinsky. Entangled Simultaneity versus Classical Interactivity in Communication Complexity. In *STOC 2016*, 2016.

- 11 Dmitry Gavinsky, Julia Kempe, Iordanis Kerenidis, Ran Raz, and Ronald de Wolf. Exponential Separation for One-Way Quantum Communication Complexity, with Applications to Cryptography. In *STOC 2007*, 2007.
- 12 Uma Girish, Ran Raz, and Avishay Tal. Quantum versus Randomized Communication Complexity, with Efficient Players. In *ITCS 2021*, 2021.
- 13 Ran Raz. Fourier Analysis for Probabilistic Communication Complexity. In *Computational Complexity Journal 1995*, 1995.
- 14 Ran Raz. Exponential Separation of Quantum and Classical Communication Complexity. In *STOC 1999*, 1999.
- 15 Ran Raz and Avishay Tal. Oracle separation of BQP and PH . In *STOC 2019*, 2019.
- 16 Oded Regev and Boaz Klartag. Quantum One-Way Communication can be Exponentially Stronger than Classical Communication. In *STOC 2011*, 2011.
- 17 Avishay Tal. Tight Bounds on the Fourier Spectrum of AC0. In *CCC 2017*, 2017.
- 18 Avishay Tal. Towards Optimal Separations between Quantum and Randomized Query Complexities. In *FOCS 2020*, 2020.

A Formal Description of the Main Results

Notation

For $n \in \mathbb{N}$, we use $[n]$ to denote the set $\{1, 2, \dots, n\}$. We typically use N to refer to 2^n . For a set $S \subseteq [n]$, let $\bar{S} := [n] \setminus S$ denote the complement of S . For sets $S \subseteq [n], T \subseteq [m]$, we typically use $S \times T := \{(s, t) : s \in S, t \in T\}$ denote the set product of S and T . Sometimes, we use the notation (S, T) . Note that the map $(i, j) \rightarrow m(i-1) + j$ is a bijection between $[n] \times [m]$ and $[nm]$. Using this identification, $S \times T$ is a subset of $[nm]$. We identify subsets $S \subseteq [n]$ with their $\{0, 1\}$ indicator vector, that is, the vector $S \in \{0, 1\}^n$ such that for each $j \in [n]$, $S_j = 1$ if and only if $j \in S$.

Let $v \in \mathbb{R}^n$. For $i \in [n]$, we refer to the i -th coordinate of v by v_i or $v(i)$. For $x, y \in \mathbb{R}^n$, let $x \cdot y \in \mathbb{R}^n$ be the pointwise product between x and y . This is the vector whose i -th coordinate is $x_i y_i$, for every $i \in [n]$. Let $\langle x, y \rangle$ denote the real inner product between x and y . For $x, y \in \{0, 1\}^n$, let $\langle x, y \rangle_2 := \sum_{i=1}^n x_i y_i \pmod 2$ denote the mod 2 inner product between x and y . We use \mathbb{I}_n to denote the $n \times n$ identity matrix. We use 0 to denote the zero vector in arbitrary dimensions.

Distributions

For a probability distribution D , let $x \sim D$ denote a random variable x sampled according to D . For distributions D_1 and D_2 , we use $D_1 \times D_2$ to denote the product distribution defined by sampling (x, y) where $x \sim D_1$ and $y \sim D_2$ are sampled independently. For $n \in \mathbb{N}$ and a distribution D , let D^n denote the product of n distributions, each of which is D . Let $\mu \in \mathbb{R}^n$ be a vector and $\Sigma \in \mathbb{R}^{n \times n}$ be a positive semi-definite matrix. We use $\mathcal{N}(\mu, \Sigma)$ to refer to the n -dimensional Gaussian distribution with mean μ and covariance matrix Σ . Let U_n denote the uniform distribution on $\{-1, 1\}^n$. For a distribution D over \mathbb{R}^n and $a \in \mathbb{R}^n$, let $a + D$ refer to the distribution obtained by sampling $z \sim D$ and returning $z + a$. For $P \in \mathbb{R}^n$ and a distribution D over \mathbb{R}^n , let $P \cdot D$ denote the distribution obtained by sampling $x \sim D$ and returning $P \cdot x$. For $p \in \mathbb{R}$, we use $p \cdot D$ to denote the distribution obtained by sampling $x \sim D$ and returning px . For $I \subseteq [n]$, let $\hat{D}(I) := \mathbb{E}_{z \sim D} [\prod_{i \in I} z_i]$ refer to the I -th moment of D .

Fourier Analysis

We refer to $\{-1, 1\}^n$ as the Boolean hypercube in n dimensions. Let $\mathcal{F} := \{f : \{-1, 1\}^n \rightarrow \mathbb{R}\}$ denote the real vector space of all Boolean functions on n variables. There is an inner product on this space as follows. For $f, g \in \mathcal{F}$, let $\langle f, g \rangle := \mathbb{E}_{x \sim U_n}[f(x)g(x)]$. For every $S \subseteq [n]$, there is a character function $\chi_S : \{-1, 1\}^n \rightarrow \{-1, 1\}$ defined at $x \in \{-1, 1\}^n$ by $\chi_S(x) := \prod_{i \in S} x_i$. The set of character functions $\{\chi_S\}_{S \subseteq [n]}$ forms an orthonormal basis for \mathcal{F} . For $f \in \mathcal{F}$ and $S \subseteq [n]$, let $\hat{f}(S) := \langle f, \chi_S \rangle$ denote the S -th Fourier coefficient of f . Note that for all $f \in \mathcal{F}$, we have $f = \sum_{S \subseteq [n]} \hat{f}(S) \chi_S$. For $f \in \mathcal{F}$, the multilinear extension of f is the unique multilinear polynomial $\tilde{f} : \mathbb{R}^n \rightarrow \mathbb{R}$ which agrees with f on $\{-1, 1\}^n$. For every $S \subseteq [n]$, the multilinear extension of χ_S is the monomial $\prod_{i \in S} x_i$. This implies that the multilinear extension of $f \in \mathcal{F}$ is $\sum_{S \subseteq [n]} \hat{f}(S) \prod_{i \in S} x_i$. Henceforth, we identify Boolean functions with their multilinear extensions. With this identification, it can be shown that functions in \mathcal{F} which map $\{-1, 1\}^n$ into $[-1, 1]$ also map $[-1, 1]^n$ into $[-1, 1]$. For $f, g \in \mathcal{F}$, let $f * g \in \mathcal{F}$ be defined at $z \in \{-1, 1\}^n$ by $(f * g)(z) := \mathbb{E}_{x \sim U_n}[f(x)g(x \cdot z)]$. It can be shown that for all $S \subseteq [n]$, we have $\widehat{f * g}(S) = \hat{f}(S)\hat{g}(S)$.

Level- k Fourier Mass

For $f \in \mathcal{F}$ and $k \in \{0, \dots, n\}$, let $L_k(f) := \sum_{|S|=k} |\hat{f}(S)|$ denote the level- k Fourier mass of f . For a family $\mathcal{H} \subseteq \mathcal{F}$ of Boolean functions, let $L_k(\mathcal{H}) := \sup_{H \in \mathcal{H}} L_k(H)$.

A.1 The Forrelation Problem

Let $k, N \in \mathbb{N}$ be parameters, where $N = 2^n$ for some $n \in \mathbb{N}$. We assume that $k = o(N^{1/50})$. Fix a parameter $\epsilon = \frac{1}{60k^2 \ln N}$. Let \mathcal{U} refer to U_{2N} .

Hadamard Matrix

The Hadamard matrix H_N of size N is an $N \times N$ matrix. The rows and columns are indexed by strings a and b respectively where $a, b \in \{0, 1\}^n$ and the (a, b) -th entry of H_N is defined to be $\frac{1}{\sqrt{N}}(-1)^{\langle a, b \rangle_2}$. Equivalently,

$$H_N(a, b) := \begin{cases} \frac{-1}{\sqrt{N}} & \text{if } \sum_{i=1}^n a_i b_i \equiv 1 \pmod{2} \\ \frac{\pm 1}{\sqrt{N}} & \text{if } \sum_{i=1}^n a_i b_i \equiv 0 \pmod{2} \end{cases}$$

The Forrelation Function

The Forrelation Function $f_{\text{orr}} : \mathbb{R}^{2N} \rightarrow \mathbb{R}$ is defined as follows. Let $z \in \mathbb{R}^{2N}$ and $x, y \in \mathbb{R}^N$ be such that $z = (x, y)$. Then,

$$f_{\text{orr}}(z) := \frac{1}{N} \langle x, H_N y \rangle$$

The \oplus^k Forrelation Decision Problem

► **Definition 1** (The \oplus^k Forrelation Decision Problem). *The Forrelation Decision Problem is the partial Boolean function $F : \{-1, 1\}^{2N} \rightarrow \{-1, 1\}$ defined as follows. For $z \in \{-1, 1\}^{2N}$,*

let

$$F(z) := \begin{cases} -1 & \text{if } \text{forr}(z) \geq \epsilon/2 \\ 1 & \text{if } \text{forr}(z) \leq \epsilon/4 \\ \text{undefined} & \text{otherwise} \end{cases}$$

The \oplus^k Forrelation Decision Problem $F^{(k)} : \{-1, 1\}^{2kN} \rightarrow \{-1, 1\}$ is defined as the XOR of k independent copies of F . To be precise, for every $z_1, \dots, z_k \in \{-1, 1\}^{2N}$, let

$$F^{(k)}(z_1, \dots, z_k) := \prod_{j=1}^k F(z_j)$$

The Gaussian Forrelation Distribution \mathcal{G}

► **Definition 2.** Let \mathcal{G} denote the Gaussian distribution over \mathbb{R}^{2N} defined by the following process.

1. Sample $x_1, \dots, x_N \sim \mathcal{N}(0, \epsilon)$ independently.
2. Let $x = (x_1, \dots, x_N)$ and $y = H_N x$.
3. Output (x, y) .

The distribution \mathcal{G} can be equivalently expressed as $\mathcal{N}\left(0, \epsilon \begin{bmatrix} \mathbb{I}_N & H_N \\ H_N & \mathbb{I}_N \end{bmatrix}\right)$.

A.2 Hard Distributions over \mathbb{R}^{2kN}

Let \mathcal{P}, \mathcal{Q} be two probability distributions on the domain $\mathbb{D} := \mathbb{R}^{2N}$. Let $S \subseteq [k]$. We define $\mathcal{P}^S \mathcal{Q}^{\bar{S}}$ to be the distribution on \mathbb{D}^k defined by sampling $x = (x_1, \dots, x_k)$ where $x_1, \dots, x_k \in \mathbb{D}$ are sampled as follows.

$$\text{For each } j \in [k], \text{ independently sample } \begin{cases} x_j \sim \mathcal{P} & \text{if } j \in S \\ x_j \sim \mathcal{Q} & \text{if } j \in \bar{S} \end{cases}$$

► **Definition 3.** Let \mathcal{G} be the distribution in Definition 2 and $\mathcal{U} = U_{2N}$. Define a pair of distributions $\mu_0^{(k)}, \mu_1^{(k)}$ on \mathbb{R}^{2kN} as follows.

$$\mu_0^{(k)} := \frac{1}{2^{k-1}} \sum_{\substack{S \subseteq [k] \\ |S| \text{ is even}}} \mathcal{G}^S \mathcal{U}^{\bar{S}} \quad \text{and} \quad \mu_1^{(k)} := \frac{1}{2^{k-1}} \sum_{\substack{S \subseteq [k] \\ |S| \text{ is odd}}} \mathcal{G}^S \mathcal{U}^{\bar{S}}$$

A.3 Rounding Distributions to the Boolean Hypercube

Let $\text{trnc} : \mathbb{R} \rightarrow [-1, 1]$ denote the truncation function, whose action on $a \in \mathbb{R}$ is given by

$$\text{trnc}(a) = \begin{cases} \text{sign}(a) & \text{if } a \notin [-1, 1] \\ a & \text{otherwise} \end{cases}$$

For $l \in \mathbb{R}$, we also use $\text{trnc} : \mathbb{R}^l \rightarrow [-1, 1]^l$ to refer to the function that applies the above truncation function coordinate-wise.

► **Definition 4.** Let μ be any distribution on \mathbb{R}^M . We define the rounded distribution $\tilde{\mu}$ on $\{-1, 1\}^M$ as follows.

1. Sample $z \sim \mu$.

52:12 Lower Bounds for XOR of Forrelations

2. For each coordinate $i \in [M]$, independently, let $z'_i = 1$ with probability $\frac{1+\text{trnc}(z_i)}{2}$ and $z'_i = -1$ with probability $\frac{1-\text{trnc}(z_i)}{2}$.
 3. Output $z' = (z'_1, \dots, z'_M)$.
- Let $z_0 \in \mathbb{R}^M$ and μ be the distribution whose support is $\{z_0\}$. We use \tilde{z}_0 to refer to $\tilde{\mu}$.

A.4 The Forrelation Distribution

Let $k \in \mathbb{N}$. Let $\tilde{\mu}_0^{(k)}$ and $\tilde{\mu}_1^{(k)}$ (respectively $\tilde{\mathcal{G}}$) be distributions over $\{-1, 1\}^{2kN}$ (respectively $\{-1, 1\}^{2N}$) generated from rounding $\mu_1^{(k)}$ and $\mu_0^{(k)}$ (respectively \mathcal{G}) according to Definition 4. Observe that we may alternatively define $\tilde{\mu}_0^{(k)}$ and $\tilde{\mu}_1^{(k)}$ as follows.

► **Definition 5.** Let \mathcal{G} be as in Definition 2 and $\mathcal{U} = U_{2N}$. Let

$$\tilde{\mu}_0^{(k)} := \frac{1}{2^{k-1}} \sum_{\substack{S \subseteq [k] \\ |S| \text{ is even}}} \tilde{\mathcal{G}}^S \mathcal{U}^{\bar{S}} \quad \text{and} \quad \tilde{\mu}_1^{(k)} := \frac{1}{2^{k-1}} \sum_{\substack{S \subseteq [k] \\ |S| \text{ is odd}}} \tilde{\mathcal{G}}^S \mathcal{U}^{\bar{S}}$$

We refer to $\tilde{\mu}_1^{(1)} \triangleq \tilde{\mathcal{G}}$ as the Forrelation Distribution.

We show that the distributions $\tilde{\mu}_1^{(k)}$ and $\tilde{\mu}_0^{(k)}$ put considerable mass on the YES and NO instances of $F^{(k)}$, respectively, where $F^{(k)}$ is the \oplus^k Forrelation Decision Problem as in Definition 1.

► **Lemma 6.** Let $\tilde{\mu}_0^{(k)}$ and $\tilde{\mu}_1^{(k)}$ be distributions as in Definition 5 and $F^{(k)}$ be the \oplus^k Forrelation Decision Problem as in Definition 1. Then,

$$\mathbb{P}_{z \sim \tilde{\mu}_0^{(k)}} [F^{(k)}(z) = 1] \geq 1 - O\left(\frac{k}{N^{6k^2}}\right) \quad \text{and} \quad \mathbb{P}_{z \sim \tilde{\mu}_1^{(k)}} [F^{(k)}(z) = -1] \geq 1 - O\left(\frac{k}{N^{6k^2}}\right)$$

A.5 Closure under Restrictions

► **Definition 7.** Let $a \in \{-1, 1, 0\}^M$. Let $\rho_a : \mathbb{R}^M \rightarrow \mathbb{R}^M$ be a restriction defined as follows. For $v \in \mathbb{R}^M$, let $\rho_a(v) \in \mathbb{R}^M$ be such that for all $j \in [M]$,

$$(\rho_a(v))(j) := \begin{cases} v(j) & \text{if } a(j) = 0 \\ a(j) & \text{otherwise} \end{cases}$$

For a function $F : \{-1, 1\}^M \rightarrow \mathbb{R}$, the restricted function $F \circ \rho_v : \{-1, 1\}^M \rightarrow \mathbb{R}$ is defined at $z \in \{-1, 1\}^M$ by $(F \circ \rho_v)(z) := F(\rho_v(z))$.

We say that a family \mathcal{H} of Boolean functions in M variables is closed under restrictions if for all restrictions $v \in \{-1, 1, 0\}^M$ and $H \in \mathcal{H}$, the restricted function $H \circ \rho_v$ is in \mathcal{H} .

B The Main Result

Let $N \in \mathbb{N}$ be a parameter describing the input size. We will assume that N is a sufficiently large power of 2. Let $k \in \mathbb{N}$. We assume that $k = o(N^{1/50})$. Let $\epsilon = \frac{1}{60k^2 \ln N}$ be the parameter defining \mathcal{G} as before.

► **Theorem 8.** Let \mathcal{H} be a family of Boolean functions on $2kN$ variables, each of which maps $\{-1, 1\}^{2kN}$ into $[-1, 1]$. Assume that \mathcal{H} is closed under restrictions. Let $\tilde{\mu}_0^{(k)}, \tilde{\mu}_1^{(k)}$ be the distributions over $\{-1, 1\}^{2kN}$ as in Definition 5. Then, for every $H \in \mathcal{H}$,

$$\left| \mathbb{E}_{z \sim \tilde{\mu}_0^{(k)}} [H(z)] - \mathbb{E}_{z \sim \tilde{\mu}_1^{(k)}} [H(z)] \right| \leq O\left(\frac{L_{2k}(\mathcal{H})}{N^{k/2}}\right) + o\left(\frac{1}{N^{k/2}}\right)$$

► **Definition 9.** Let $\tilde{\mu}_0^{(k)}, \tilde{\mu}_1^{(k)}$ be as in Definition 5. Let $\sigma_0^{(k)}$ (respectively $\sigma_1^{(k)}$) be obtained by conditioning $\tilde{\mu}_0^{(k)}$ on being a NO (respectively YES) instance of $F^{(k)}$.

► **Corollary 10.** Under the same hypothesis as Theorem 8, for every $H \in \mathcal{H}$

$$\left| \mathbb{E}_{z \sim \sigma_0^{(k)}} [H(z)] - \mathbb{E}_{z \sim \sigma_1^{(k)}} [H(z)] \right| \leq O\left(\frac{L_{2k}(\mathcal{H})}{N^{k/2}}\right) + o\left(\frac{1}{N^{k/2}}\right)$$

B.1 Applications to Quantum versus Classical Separations

Query Complexity Separations

► **Lemma 11.** Let $D : \{-1, 1\}^{2kN} \rightarrow \{-1, 1\}$ be a deterministic decision tree of depth $d \geq 1$. Then,

$$\left| \mathbb{E}_{z \sim \sigma_0^{(k)}} [D(z)] - \mathbb{E}_{z \sim \sigma_1^{(k)}} [D(z)] \right| \leq \left(\frac{O(d \log(kN))}{N^{1/2}} \right)^k$$

► **Theorem 12.** $F^{(k)}$ can be computed in the bounded-error quantum query model with $O(k^5 \log^2 N \log k)$ queries. However, every randomized decision tree of depth $\tilde{o}(\sqrt{N})$ has a worst-case success probability of at most $\frac{1}{2} + \exp(-\Omega(k))$.

Setting $k = \lceil \log^c N \rceil$ for $c \in \mathbb{N}$ in Theorem 12 gives us an explicit family of partial functions that are computable by quantum query algorithms of cost $\tilde{O}(\log^{5c+2} N)$, however every randomized query algorithm of cost $\tilde{o}(N^{\frac{1}{2}})$ has at most $\frac{1}{2^{\Omega(\log^c N)}}$ advantage over random guessing.

Communication Complexity Separations

► **Definition 13** (The \oplus^k Forrelation Communication Problem $F^{(k)} \circ \text{XOR}$). Alice is given x and Bob is given y where $x, y \in \{-1, 1\}^{2kN}$. Let $F^{(k)}$ be as in Definition 1. Their goal is to compute the partial function $F^{(k)}(x \cdot y)$.

► **Lemma 14.** Let $C : \{-1, 1\}^{2kN} \times \{-1, 1\}^{2kN} \rightarrow \{-1, 1\}$ be any deterministic protocol of communication complexity c . Then,

$$\left| \mathbb{E}_{\substack{x \sim U_{2kN} \\ z \sim \sigma_0^{(k)}}} [C(x, x \cdot z)] - \mathbb{E}_{\substack{x \sim U_{2kN} \\ z \sim \sigma_1^{(k)}}} [C(x, x \cdot z)] \right| \leq O\left(\frac{(c + 8k)^{2k}}{N^{k/2}}\right)$$

► **Theorem 15.** $F^{(k)} \circ \text{XOR}$ can be solved in the quantum simultaneous with entanglement model with $O(k^5 \log^3 N \log k)$ bits of communication, when Alice and Bob share $O(k^5 \log^3 N \log k)$ EPR pairs. However, any randomized protocol of cost $\tilde{o}(N^{1/4})$ has a worst-case success probability of at most $\frac{1}{2} + \exp(-\Omega(k))$.

Setting $k = \lceil \log^c N \rceil$ for $c \in \mathbb{N}$ in Theorem 15 gives us an explicit family of partial functions that are computable by quantum simultaneous protocols of cost $\tilde{O}(\log^{5c+3} N)$ when Alice and Bob share $\tilde{O}(\log^{5c+3} N)$ EPR pairs, however every interactive randomized protocol of cost $\tilde{o}(N^{\frac{1}{4}})$ has at most $\frac{1}{2^{\Omega(\log^c N)}}$ advantage over random guessing.

Circuit Complexity Separations

► **Lemma 16.** *Let $C : \{-1, 1\}^{2kN} \rightarrow \{-1, 1\}$ be an AC0 circuit of depth $d \geq 1$ and size s . Then,*

$$\left| \mathbb{E}_{z \sim \sigma_0^{(k)}} [C(z)] - \mathbb{E}_{z \sim \sigma_1^{(k)}} [C(z)] \right| \leq \left(\frac{O(\log^{2d-2}(s))}{N^{1/2}} \right)^k$$

► **Theorem 17.** *The distributions $\sigma_1^{(k)}$ and $\sigma_0^{(k)}$ can be distinguished by a bounded-error quantum query protocol with $O(k^5 \log^2 N \log k)$ queries with $2/3$ advantage. However, every constant depth circuit of size $o\left(\exp\left(N^{\frac{1}{4(d-1)}}\right)\right)$ can distinguish these distributions with at most $\exp(-\Omega(k))$ advantage.*

Setting $k = \lceil \log^c N \rceil$ for $c \in \mathbb{N}$ in Theorem 17 gives us an explicit family of distributions that are distinguishable by cost $\tilde{O}(\log^{5c+2} N)$ quantum query algorithms, however every constant depth circuit of quasipolynomial size can distinguish them with at most $\frac{1}{2^{\Omega(\log^c N)}}$ advantage.