Oracle Separation of BQP and PH

Ran Raz*
Princeton University
Princeton, NJ, USA
ran.raz.mail@gmail.com

Avishay Tal†
Stanford University
Stanford, CA, USA
avishay.tal@gmail.com

ABSTRACT

We present a distribution $\mathcal{D}$ over inputs in $\{\pm 1\}^{2^N}$, such that:

(1) There exists a quantum algorithm that makes one (quantum) query to the input, and runs in time $O(\log N)$, that distinguishes between $\mathcal{D}$ and the uniform distribution with advantage $\Omega(1/\log N)$.
(2) No Boolean circuit of quasipoly(N) size and constant depth distinguishes between $\mathcal{D}$ and the uniform distribution with advantage better than polylog(N)/$\sqrt{N}$.

By well known reductions, this gives a separation of the classes Promise-**BQP** and Promise-**PH** in the *black-box* model and implies an oracle O relative to which $\mathbf{BQP}^O \not\subseteq \mathbf{PH}^O$.

CCS CONCEPTS

• **Theory of computation** → **Circuit complexity**; **Oracles and decision trees**; **Quantum complexity theory**; *Complexity classes.*

KEYWORDS

BQP, oracle separation, polynomial hierarchy, bounded depth circuits

ACM Reference Format:
Ran Raz and Avishay Tal. 2019. Oracle Separation of BQP and PH. In *Proceedings of the 51st Annual ACM SIGACT Symposium on the Theory of Computing (STOC '19), June 23–26, 2019, Phoenix, AZ, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3313276.3316315

1 INTRODUCTION

Can polynomial-time quantum algorithms be simulated by classical algorithms in the polynomial-time hierarchy?

In this paper, we show that in the *black-box* model (also known as *query-complexity* or *decision-tree complexity*), the answer is negative. That is, in the black-box model, the class **BQP** of (promise[1]) problems that can be solved by bounded-error polynomial-time quantum algorithms, is not contained in the class **PH**, the (classical) polynomial-time hierarchy.

More precisely, we give an explicit black-box (promise) problem, that can be solved by a polynomial-time quantum algorithm with only one query, but cannot be solved by a classical algorithm in the polynomial-time hierarchy.

It is well known that this implies an oracle O relative to which **BQP** is not contained in **PH**. Previously, even an oracle separation of **BQP** and **AM** was not known.

1.1 Motivation and Related Work

The black-box model has played a central role in the study of quantum computational complexity. For example, Shor's algorithm for factoring [Sho97] builds on Simon's black-box algorithm for finding periodicity [Sim94] and Grover's database search algorithm is stated and proved directly in the black box model [Gro96].

The relative power of **BQP** and classical complexity classes, in the black box model, has been studied in numerous works, starting with the celebrated paper by Bernstein and Vazirani that defined the class **BQP** and founded the field of quantum computational complexity [BV97]. Bernstein and Vazirani proved that in the black-box model $\mathbf{BQP} \not\subseteq \mathbf{BPP}$ [BV97]. They also extended their proof to show that in the black-box model $\mathbf{BQP} \not\subseteq \mathbf{MA}$ and conjectured that in the black-box model $\mathbf{BQP} \not\subseteq \mathbf{PH}$ (private communication with the authors). A full (different) proof that in the black-box model $\mathbf{BQP} \not\subseteq \mathbf{MA}$ was given by Watrous [Watrous00].

Aaronson defined the *Forrelation* problem, as a candidate for separating **BQP** and **PH** in the black-box model [Aar10]. By studying a variant of the Forrelation problem, he obtained a relation problem (i.e., a problem with many valid outputs) that is solvable in the black-box model in **BQP**, but not in **PH**, thus separating the relation versions of these classes, in the black-box model. Consequently, Aaronson obtained an oracle O relative to which the relation version of **BQP** is not contained in the relation version of **PH**. That work by Aaronson has been quite influential. It led to additional results in the black box model, such as [AA15, Chen16], as well as results that are seemingly unrelated, such as [Aar11]. Followup works, such as [FSUV13, Rem16], further studied the problems of separating **BQP** and **PH** in the black-box model and obtaining an oracle O relative to which **BQP** is not contained in **PH**.

In his work [Aar10], Aaronson motivated the study of oracle/black-box separation of **BQP** and **PH** in various ways. First,

*Research supported by the Simons Collaboration on Algorithms and Geometry and by the National Science Foundation grant No. CCF-1412958.
†Research supported by a Motwani Postdoctoral Fellowship and by NSF grant CCF-1763311.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
STOC '19, June 23–26, 2019, Phoenix, AZ, USA
© 2019 Association for Computing Machinery.
ACM ISBN 978-1-4503-6705-9/19/06…$15.00
https://doi.org/10.1145/3313276.3316315

[1] In our entire discussion of black-box complexity classes, we consider complexity classes of promise problems, rather than decision problems. Nevertheless, separations of classes of promise problems in the black-box model imply oracle separations of the corresponding classes of decision problems in the "real" world (see [Aar10]; Footnote 4).

he views such a separation as a formal evidence for the possibility that **BQP** $\not\subseteq$ **PH** in the real world. Second, he argues that oracle separations played a role in many of the central developments in complexity theory and an even more central and decisive role in quantum computing. Third, he argues that the black-box (query complexity) model is a natural and well motivated model in its own right. Finally, he mentions that such a separation also implies a separation of the classes **BQLOGTIME** (the class of promise problems decidable by quantum algorithms that have random access to an N-bit input, and run in time $O(\log N)$) and $\mathbf{AC^0}$,[2] in the real world. (We refer the reader to [Aar10, FSUV13] for further details).

1.2 Our Results

Let $\mathcal{D}, \mathcal{D}'$ be two probability distributions over a finite set X. We say that an algorithm A distinguishes between $\mathcal{D}$ and $\mathcal{D}'$ with advantage ε if

$$\varepsilon = \left| \Pr_{x\sim\mathcal{D}}[A \text{ accepts } x] - \Pr_{x'\sim\mathcal{D}'}[A \text{ accepts } x'] \right|.$$

The following is our main result:

Theorem 1.1. *There exists an explicit distribution $\mathcal{D}$ over inputs in $\{\pm1\}^{2N}$, such that:*

1. *There exists a quantum algorithm that makes one query to the input, and runs in time $O(\log N)$, that distinguishes between $\mathcal{D}$ and the uniform distribution with advantage $\Omega(1/\log N)$.*
2. *No Boolean circuit of size* quasipoly(N) *and constant depth*[3] *distinguishes between $\mathcal{D}$ and the uniform distribution with advantage better than* polylog$(N)/\sqrt{N}$.

We can amplify the advantage of the quantum algorithm in Theorem 1.1 by making polylog(N) sequential repetitions and obtain the following result:

Theorem 1.2. *There exists an explicit distribution $\mathcal{D}_1$ over inputs in $\{\pm1\}^{N_1}$, such that:*

1. *There exists a quantum algorithm that makes* polylog(N_1) *queries to the input, and runs in time* polylog(N_1)*, that distinguishes between $\mathcal{D}_1$ and the uniform distribution with probability $1 - 2^{-\text{polylog}(N_1)}$.*
2. *No Boolean circuit of size* quasipoly(N_1) *and constant depth distinguishes between $\mathcal{D}_1$ and the uniform distribution with advantage better than* polylog$(N_1)/\sqrt{N_1}$.

We can also amplify the advantage of the quantum algorithm in Theorem 1.1, using only one quantum query (by standard amplification techniques) and obtain the following result:

Theorem 1.3. *There exist explicit distributions $\mathcal{D}_2$ and $\widetilde{U}$ over inputs in $\{\pm1\}^{N_2}$, such that:*

1. *There exists a quantum algorithm that makes one query to the input, and runs in time $O(\log N_2)$, that distinguishes between $\mathcal{D}_2$ and $\widetilde{U}$ with probability $1 - 2^{-(\log N_2)^{\Omega(1)}}$.*
2. *No Boolean circuit of size* quasipoly(N_2) *and constant depth distinguishes between $\mathcal{D}_2$ and $\widetilde{U}$ with advantage better than $2^{-(\log N_2)^{\Omega(1)}}$.*

[2] Recall that $\mathbf{AC^0}$ refers to Boolean circuits of polynomial size and constant depth.
[3] In fact, our lower bounds on Boolean circuits may be extended up to sub-exponential size constant depth circuits. See Theorem 7.4 for the exact dependency on the size and depth.

By the standard and straightforward relation between $\mathbf{AC^0}$ and **PH** [FSS84] (by replacing every $\forall$ by $\wedge$ gate and every $\exists$ by $\vee$ gate), we have the following corollary:

Corollary 1.4. *In the black-box model, Promise-***BQP** $\not\subseteq$ *Promise-***PH**.

By the relation between black-box separations and oracle separations, we have the following corollary. We include its proof in the appendix for completeness.

Corollary 1.5. *There exists an oracle O relative to which* $\mathbf{BQP}^O \not\subseteq \mathbf{PH}^O$.

Finally, an immediate corollary of our main theorems (for details, see [Aar10, FSUV13]):

Corollary 1.6. *Promise-***BQLOGTIME** $\not\subseteq$ *Promise-*$\mathbf{AC^0}$.

1.3 Techniques

Our distribution $\mathcal{D}$ is a variant of Aaronson's Forrelation distribution [Aar10], but differs from it in a way that turned out to be crucial in our analysis.

The quantum algorithm for distinguishing between $\mathcal{D}$ and the uniform distribution was suggested by [Aar10, AA15].

The hard part of our result is the lower bound for bounded depth circuits distinguishing between $\mathcal{D}$ and the uniform distribution. For this part, we use Fourier analysis. We use Tal's tail bounds on the Fourier spectrum of bounded depth circuits [Tal17] (that builds on a long line of works, in particular [LMN93, Hås14]). We also use the fascinating recent approach of Chattopadhyay, Hatami, Hosseini and Lovett for constructing pseudorandom distributions by considering a random walk that makes small steps, where each step is sampled from a pseudorandom distribution that takes values in $[-1,1]^N$ [CHHL18]. In particular, their (simple yet powerful) Claim 3.3 is crucial for our proof.

2 PROOF OUTLINE

2.1 The Distribution $\mathcal{D}$

Our distribution $\mathcal{D}$ is a variant of Aaronson's Forrelation distribution, but differs from it in a way that turned out to be crucial in our analysis.

Let $n \in \mathbb{N}$ and $N = 2^n$. Let $\varepsilon = 1/(24 \cdot \ln N)$. We define a probability distribution $\mathcal{G}'$ over $\mathbb{R}^N \times \mathbb{R}^N$ as follows: Sample $x_1, \ldots, x_N \sim \mathcal{N}(0, \varepsilon)$ independently. Let $y = H_N \cdot x$ (where H_N is the Hadamard transform). Output $z = (x, y)$. Note that $\mathcal{G}'$ is a multivariate gaussian distribution with zero-means and covariance matrix

$$\varepsilon \cdot \begin{pmatrix} I_N & H_N \\ H_N & I_N \end{pmatrix}.$$

Aaronson had a similar distribution (with $\varepsilon = 1$) and obtained from it a distribution over $\{\pm1\}^{2N}$, by replacing each z_i by $\text{sgn}(z_i)$. Instead, our distribution $\mathcal{D}$ is defined as follows: We draw $z \sim \mathcal{G}'$ and truncate each z_i to the interval $[-1,1]$, by applying the function $\text{trnc}(z_i) := \min(1, \max(-1, z_i))$. Then, in order to obtain values in $\{\pm1\}$, independently for each $i \in [2N]$ we draw $z'_i = 1$ with probability $\frac{1+\text{trnc}(z_i)}{2}$ and $z'_i = -1$ with probability $\frac{1-\text{trnc}(z_i)}{2}$. We output $z' \in \{\pm1\}^{2N}$.

Note that since ε is sufficiently small, we have that with high probability, every z_i is already in the interval $[-1,1]$, to begin with. Thus, the truncation operation occurs with negligible probability and we show that it could be essentially ignored in the analysis (See Sec. 5). The more important point is that, assuming that $z_i \in [-1,1]$, we take $z'_i = 1$ with probability $\frac{1+z_i}{2}$ and $z'_i = -1$ with probability $\frac{1-z_i}{2}$, rather than taking z'_i to be the sign of z_i.

The reason for defining $\mathcal{D}$ as above is that we can prove the following fact, that turned out to be crucial for our analysis: Let $F : \mathbb{R}^{2N} \to \mathbb{R}$ be any multilinear function that maps $[-1,1]^{2N}$ to $[-1,1]$. Then, F has similar expectation under $\mathcal{G}'$ and under $\mathcal{D}$. In fact, denoting $\mathrm{trnc}(z) := (\mathrm{trnc}(z_1), \ldots, \mathrm{trnc}(z_{2N}))$, we show that

$$\mathop{\mathbf{E}}_{z' \sim \mathcal{D}}[F(z')] = \mathop{\mathbf{E}}_{z \sim \mathcal{G}'}[F(\mathrm{trnc}(z))]$$

and since, as mentioned above, truncation occurs with negligible probability, we can prove that

$$\mathop{\mathbf{E}}_{z' \sim \mathcal{D}}[F(z')] \approx \mathop{\mathbf{E}}_{z \sim \mathcal{G}'}[F(z)].$$

Thus, in large parts of the proof, we can analyze the distribution $\mathcal{G}'$, rather than $\mathcal{D}$.

2.2 The Quantum Algorithm

The quantum algorithm for distinguishing between $\mathcal{D}$ and the uniform distribution is simple and is similar to [Aar10, AA15]. Since the constraint $y = H_N \cdot x$ is linear, the support of the distribution $\mathcal{G}'$ is an N-dimensional linear subspace $\mathcal{H} \subset \mathbb{R}^{2N}$. A vector $z \sim \mathcal{D}$ will be, on average, closer to $\mathcal{H}$ than a random vector. Thus, intuitively, the algorithm just needs to accept with higher probability vectors that are closer to $\mathcal{H}$.

Given an input $z = (x,y) \in \{\pm 1\}^{2N}$, the algorithm can generate, by one quantum query, the quantum state $\Psi = \frac{1}{\sqrt{2N}} \sum_{i=1}^{2N} z_i |i\rangle$. By applying an appropriate unitary transformation and measuring the state, the algorithm can accept with higher probability when the distance between z and $\mathcal{H}$ is relatively small, and thus distinguish between $\mathcal{D}$ and the uniform distribution. It was shown in [Aar10, AA15] how to implement the appropriate unitary transformation efficiently, by $O(\log N)$ quantum gates (using the fact that the Hadamard transform can be computed using $O(\log N)$ quantum gates). The algorithm distinguishes between $\mathcal{D}$ and the uniform distribution with advantage that is proportional to ε.

2.3 The AC^0 Lower Bound

The hard part of our result is the lower bound for bounded depth circuits distinguishing between $\mathcal{D}$ and the uniform distribution. Our proof uses Fourier analysis; Tal's tail bounds for the Fourier coefficients of bounded depth circuits [Tal17] and a recent approach and claim by Chattopadhyay, Hatami, Hosseini and Lovett [CHHL18].

Let $A : \{\pm 1\}^{2N} \to \{\pm 1\}$ be a Boolean circuit of quasi-polynomial size and constant depth. For a vector $z \in \mathbb{R}^{2N}$, we denote by $A(z)$ the value of the multilinear extension of A on z. The multilinear extension $A : \mathbb{R}^{2N} \to \mathbb{R}$ can be written as $A(z) = \sum_{S \subseteq [2N]} \widehat{A}(S) \cdot \prod_{i \in S} z_i$, where $\widehat{A}(S)$ are the Fourier coefficients of A. Observe that $A(\vec{0}) = \widehat{A}(\emptyset) = \mathbf{E}_{u \sim U_{2N}}[A(u)]$. Tal's tail bounds imply that for all $k \in \mathbb{N}$, we have

$$\sum_{S \subseteq [2N] : |S| = k} |\widehat{A}(S)| \leq (\mathrm{polylog}(N))^k.$$

We need to prove that A cannot distinguish between the distributions $\mathcal{D}$ and U_{2N}. That is, we need to prove that

$$\mathop{\mathbf{E}}_{z \sim \mathcal{D}}[A(z)] \approx \mathop{\mathbf{E}}_{u \sim U_{2N}}[A(u)].$$

Since $\mathbf{E}_{u \sim U_{2N}}[A(u)] = A(\vec{0})$ and since, as mentioned above, $\mathbf{E}_{z \sim \mathcal{D}}[A(z)] \approx \mathbf{E}_{z \sim \mathcal{G}'}[A(z)]$, it will be sufficient to prove that

$$\left| \mathop{\mathbf{E}}_{z \sim \mathcal{G}'}[A(z)] - A(\vec{0}) \right|$$

is small.

Denote by $\widehat{\mathcal{G}'}(S)$ the moments of the distribution $\mathcal{G}'$. That is, $\widehat{\mathcal{G}'}(S) = \mathbf{E}_{z \sim \mathcal{G}'}[\prod_{i \in S} z_i]$. (We use moments, rather than Fourier coefficients, because the distribution is over the reals. In some sense, these moments play the role of Fourier coefficients in our proof). The values of $\widehat{\mathcal{G}'}(S)$ are well known. In particular, $\widehat{\mathcal{G}'}(S) = 0$ when $|S|$ is odd, and we have a closed formula for the case that $|S|$ is even.

Similarly to Plancherel's theorem, it is easy to bound

$$\left| \mathop{\mathbf{E}}_{z \sim \mathcal{G}'}[A(z)] - A(\vec{0}) \right| = \left| \sum_{\emptyset \neq S \subseteq [2N]} \widehat{A}(S) \cdot \widehat{\mathcal{G}'}(S) \right|$$

$$\leq \sum_{k \geq 1} \sum_{|S| = k} |\widehat{A}(S)| |\widehat{\mathcal{G}'}(S)| = \sum_{k \geq 2} \sum_{|S| = k} |\widehat{A}(S)| |\widehat{\mathcal{G}'}(S)| \qquad (1)$$

(where the last equality is since $\widehat{\mathcal{G}'}(S) = 0$ when $|S|$ is odd). It turns out that when plugging in the bounds that we have on the moments $|\widehat{\mathcal{G}'}(S)|$ and the bounds that we have on $\sum_{|S|=k} |\widehat{A}(S)|$, the terms that correspond to small k-s are small, but the bounds that we get on terms that correspond to large k-s (i.e. $k \geq \sqrt{N}$) are too large. Thus, it is not clear how to use this expression to prove the desired result.

Here we use the approach of [CHHL18]. Let $t := N$ and $p := 1/\sqrt{t} = N^{-1/2}$. Rather than sampling $z \sim \mathcal{G}'$, we sample independently $z^{(1)}, \ldots, z^{(t)} \sim \mathcal{G}'$. For $i = 0, \ldots, t$, let $z^{\leq (i)} = p \cdot (z^{(1)} + \ldots + z^{(i)})$. We think of $z^{\leq (1)}, \ldots, z^{\leq (t)}$ as a random walk that makes small steps, where each step is a small constant p times a random variable that is distributed according to the original distribution $\mathcal{G}'$. (We note that in the limit $p \to 0$, that walk would be a Brownian motion).

[CHHL18] used a similar random walk[4] in order to converge to the discrete cube $\{\pm 1\}^{2N}$. In that sense, their walk was used in order to obtain better and better distributions, that is, distributions that are closer and closer to $\{\pm 1\}^{2N}$. Their motivation was the construction of pseudorandom generators and hence they tried to minimize the number of steps, as they needed fresh random bits for each step. Instead, our motivation is separating BQP from PH in the relativized world, and we use our random walk just in order to analyze the original distribution $\mathcal{G}'$. The crucial point is that since $\mathcal{G}'$ is a multivariate gaussian distribution, the distribution of $z^{\leq (t)}$ is exactly $\mathcal{G}'$. Thus, in our case, the random walk does not give a better distribution; it gives the exact same distribution and

[4] One difference is that we take a simple random walk whereas [CHHL18] adaptively scale each step according to the current location of the walk, in order to get closer to $\{\pm 1\}^{2N}$.

it is used because it gives a powerful way to analyze the original distribution, as described next.

Similarly to Claim 3.3 of [CHHL18], we can prove that if for every Boolean circuit of a certain size and depth and for some b,

$$\left| \mathop{\mathbf{E}}_{z\sim\mathcal{G}'}[A(pz)] - A(\vec{0}) \right| \leq b,$$

then for every $i \in [t]$,

$$\left| \mathbf{E}[A(z^{\leq(i)})] - A(\vec{0}) \right| \leq O(i \cdot b),$$

and in particular for $i = t$,

$$\left| \mathop{\mathbf{E}}_{z\sim\mathcal{G}'}[A(z)] - A(\vec{0}) \right| = \left| \mathbf{E}[A(z^{\leq(t)})] - A(\vec{0}) \right| \leq O(t \cdot b) = O(p^{-2} \cdot b).$$

Since multiplication of a random variable by a factor of $p < 1$, reduces all moments of order k by a factor of p^k, similarly to Equation (1), we have

$$\left| \mathop{\mathbf{E}}_{z\sim\mathcal{G}'}[A(pz)] - A(\vec{0}) \right| \leq \sum_{k\geq 2} p^k \sum_{|S|=k} |\widehat{A}(S)||\widehat{\mathcal{G}'}(S)|.$$

Thus,

$$\left| \mathop{\mathbf{E}}_{z\sim\mathcal{G}'}[A(z)] - A(\vec{0}) \right| \leq O\Big(p^{-2} \cdot \sum_{k\geq 2} p^k \sum_{|S|=k} |\widehat{A}(S)||\widehat{\mathcal{G}'}(S)|\Big).$$

The last expression is (up to a multiplicative constant) the same as Equation (1), except that all terms were reduced by a factor of p^{k-2}. Since we took p to be very small, all terms except for $k = 2$ can essentially be ignored and plugging in the bounds on $|\widehat{\mathcal{G}'}(S)|$ and the bounds on $\sum_{|S|=k} |\widehat{A}(S)|$, we get the desired bound.

3 PRELIMINARIES

We denote by U_N the uniform distribution over $\{\pm1\}^N$. We denote by I_N the identity matrix of order N.

We shall use the following standard bound on the gaussian distribution: for any positive $x \in \mathbb{R}$, we have $\mathbf{Pr}[|\mathcal{N}(0,1)| \geq x] \leq e^{-x^2/2}$.

We consider Boolean circuits consisting of unbounded fan-in AND, OR gates applied to input variables and their negation. We only consider circuits with one output. The size of a circuit is the number of gates it contains. The depth is defined as the length of the longest path (in edges) from any input to the output.

3.1 The Hadamard Transformation

Let $n \in \mathbb{N}$ and $N = 2^n$. The Hadamard transform $H = H_N \in \mathbb{R}^{N\times N}$ is defined as follows. For $i, j \in [N]$,

$$H_{i,j} = N^{-1/2} \cdot (-1)^{\langle i,j\rangle},$$

where $\langle i, j\rangle$ denotes the inner product between the binary representations of $(i-1)$ and $(j-1)$. It is well known that H_N is orthonormal and symmetric and thus $H_N \cdot H_N = I_N$.

3.2 Quantum Query

A quantum query to an input $z \in \{\pm1\}^{2N}$ performs the diagonal unitary transformation $\mathcal{U}_z$, defined by $|i, w\rangle \to z_i|i, w\rangle$, where $i \in [2N]$ and w represents the auxiliary workspace that does not participate in the query.

4 THE DISTRIBUTION $\mathcal{D}$

4.1 The Forrelation Distribution and its Variant

Let $n \in \mathbb{N}$ and $N = 2^n$. We assume that n is sufficiently large. We follow Aaronson suggestion [Aar10], that defined a distribution $\mathcal{F}$ on $\{\pm1\}^N \times \{\pm1\}^N$ (called Forrelation) that is sampled as follows:

(1) Sample $x_1, \ldots, x_N \sim \mathcal{N}(0,1)$ independently.
(2) Let $y = H_N \cdot x$ (where H_N is the Hadamard transform).
(3) Output $(\mathrm{sgn}(x), \mathrm{sgn}(y))$.

We define a probability distribution $\mathcal{G}$ over $\mathbb{R}^N \times \mathbb{R}^N$ using the same process, but without taking signs. That is, $\mathcal{G}$ is sampled as follows:

(1) Sample $x_1, \ldots, x_N \sim \mathcal{N}(0,1)$ independently.
(2) Let $y = H_N \cdot x$ (where H_N is the Hadamard transform).
(3) Output $z = (x, y)$.

Observe that $\mathcal{G}$ is a multivariate gaussian distribution with zero-means and covariance matrix

$$\begin{pmatrix} I_N & H_N \\ H_N & I_N \end{pmatrix}.$$

It is thus clear that $\mathcal{G}$ is symmetric in x and y. Note that $x_1, \ldots, x_N$ are independent random variables. Similarly $y_1, \ldots, y_N$ are independent.

For $S, T \subseteq [N]$, we wish to analyze the "Fourier coefficients" of $\mathcal{G}$, defined as

$$\widehat{\mathcal{G}}(S,T) \triangleq \mathop{\mathbf{E}}_{(x,y)\sim\mathcal{G}}\left[\prod_{i\in S} x_i \cdot \prod_{j\in T} y_j\right]$$

(these are actually the moments of $\mathcal{G}$). In the next claim, we bound $\widehat{\mathcal{G}}(S,T)$.

Claim 4.1. *Let $S, T \subseteq [N]$ and $i, j \in [N]$. Let $k_1 = |S|$, $k_2 = |T|$. Then,*

(1) $\widehat{\mathcal{G}}(\{i\},\{j\}) = N^{-1/2} \cdot (-1)^{\langle i,j\rangle}$.
(2) $\widehat{\mathcal{G}}(S,T) = 0$ *if* $k_1 \neq k_2$.
(3) $|\widehat{\mathcal{G}}(S,T)| \leq k! \cdot N^{-k/2}$ *if* $k = k_1 = k_2$.
(4) $|\widehat{\mathcal{G}}(S,T)| \leq 1$ *for all* S, T.

Proof. All items rely on the fact that $\mathcal{G}$ is a multivariate gaussian distribution with zero-means and covariance matrix

$$\begin{pmatrix} I_N & H_N \\ H_N & I_N \end{pmatrix}.$$

The first item is trivial as it is actually an entry in the covariance matrix above.

The second and third items rely on Isserlis' Theorem [Iss1918] (See also http://en.wikipedia.org/wiki/Isserlis'_theorem), stating that in a zero-mean multivariate gaussian distribution $Z_1, \ldots, Z_{2N}$, for distinct $i_1, \ldots, i_{2k} \in [2N]$, we have $\mathbf{E}[Z_{i_1} \cdots Z_{i_{2k-1}}] = 0$ and $\mathbf{E}[Z_{i_1} \cdots Z_{i_{2k}}] = \sum \prod \mathbf{E}[Z_{i_r} Z_{i_\ell}]$, where the notation $\sum \prod$ means summing over all distinct ways of partitioning $Z_{i_1}, \ldots, Z_{i_{2k}}$ into pairs and each summand is the product of the k pairs. In our case, if $|S| \neq |T|$, in any partition of the elements of S and T into pairs, we will have a pair in which either the two entries are from the left half or the two entries are from the right half, however the covariance of any such pair is 0. This gives the second item. For the third item, we note that if $|S| = |T| = k$ there are $k!$ partitions of the elements of S and T into pairs such that each pair contains

exactly one variable from each half. The covariance of each pair is $\pm N^{-1/2}$. Thus, we are summing $k!$ numbers which are $\pm N^{-k/2}$, which gives $|\widehat{\mathcal{G}}(S,T)| \le k! \cdot N^{-k/2}$.

The last item is a simple application of Cauchy-Schwarz inequality:

$$\widehat{\mathcal{G}}(S,T) = \mathbf{E}\left[\prod_{i\in S} x_i \prod_{j\in T} y_j\right] \le \sqrt{\mathbf{E}\left[\prod_{i\in S} x_i^2\right] \cdot \mathbf{E}\left[\prod_{j\in T} y_j^2\right]}$$
$$= \sqrt{\prod_{i\in S} \mathbf{E}[x_i^2] \cdot \prod_{j\in T} \mathbf{E}[y_j^2]} = 1. \qquad \square$$

4.2 The Distribution $\mathcal{D}$

Let $n \in \mathbb{N}$. Let $N = 2^n$. We assume that n is sufficiently large. Recall that $\mathcal{G}$ is a multivariate gaussian distribution over $\mathbb{R}^{2N}$ with zero-means and covariance matrix

$$\begin{pmatrix} I_N & H_N \\ H_N & I_N \end{pmatrix}.$$

Let $\varepsilon = 1/(24 \cdot \ln N)$. We define a probability distribution $\mathcal{G}'$ over $\mathbb{R}^{2N}$, sampled as follows: Sample $z \sim \mathcal{G}$ and output $\sqrt{\varepsilon} \cdot z$. Note that $\mathcal{G}'$ is a multivariate gaussian distribution over $\mathbb{R}^{2N}$ with zero-means and covariance matrix

$$\varepsilon \cdot \begin{pmatrix} I_N & H_N \\ H_N & I_N \end{pmatrix}.$$

Let $\mathrm{trnc}(a) := \min(1, \max(-1, a))$ be the function that given a real number, truncates it to the interval $[-1, 1]$.

The Distribution $\mathcal{D}$: We draw $z \sim \mathcal{G}'$ and take $\mathrm{trnc}(z) := (\mathrm{trnc}(z_1), \ldots, \mathrm{trnc}(z_{2N}))$. Then, independently for each $i \in [2N]$ we draw $z'_i = 1$ with probability $\frac{1+\mathrm{trnc}(z_i)}{2}$ and $z'_i = -1$ with probability $\frac{1-\mathrm{trnc}(z_i)}{2}$. We output $z' \in \{\pm 1\}^{2N}$.

Observe that conditioned on the value of $z \sim \mathcal{G}'$, we have that $z'_1, \ldots, z'_{2N}$ are independent and for each $i \in [2N]$ the expected value of z'_i equals $\mathrm{trnc}(z_i)$.

5 MULTILINEAR FUNCTIONS ON $\mathcal{D}$

In this section, we show that any multilinear function $F : \mathbb{R}^{2N} \to \mathbb{R}$ that maps $[-1, 1]^{2N}$ to $[-1, 1]$ has similar expectation under $\mathcal{G}'$ and under $\mathcal{D}$. Let $F : \mathbb{R}^{2N} \to \mathbb{R}$ be a multilinear function, defined by

$$F(z) = \sum_{S\subseteq[2N]} \widehat{F}(S) \cdot \prod_{i\in S} z_i,$$

where $\widehat{F}(S) \in \mathbb{R}$. First we show that

$$\mathop{\mathbf{E}}_{z'\sim\mathcal{D}}[F(z')] = \mathop{\mathbf{E}}_{z\sim\mathcal{G}'}[F(\mathrm{trnc}(z))]. \tag{2}$$

For the proof of Equation (2), recall that $z' \sim \mathcal{D}$ can be generated as follows: Draw $z \sim \mathcal{G}'$. Then, independently for each $i \in [2N]$, draw $z'_i = 1$ with probability $\frac{1+\mathrm{trnc}(z_i)}{2}$ and $z'_i = -1$ with probability $\frac{1-\mathrm{trnc}(z_i)}{2}$. Equation (2) holds since conditioned on the value of z, the expected value of $F(z')$ equals $F(\mathrm{trnc}(z))$. To see this, we use linearity of expectation and the definition of $\mathcal{G}', \mathcal{D}$:

$$\mathbf{E}[F(z') \mid z] = \mathbf{E}\left[\sum_{S\subseteq[2N]} \widehat{F}(S) \cdot \prod_{i\in S} z'_i \;\middle|\; z\right]$$
$$= \sum_{S\subseteq[2N]} \widehat{F}(S) \cdot \prod_{i\in S} \mathbf{E}[z'_i \mid z]$$
$$= \sum_{S\subseteq[2N]} \widehat{F}(S) \cdot \prod_{i\in S} \mathrm{trnc}(z_i)$$
$$= F(\mathrm{trnc}(z)).$$

Thus, we need to bound the difference between $\mathbf{E}_{z\sim\mathcal{G}'}[F(\mathrm{trnc}(z))]$ and $\mathbf{E}_{z\sim\mathcal{G}'}[F(z)]$. Note that whenever $z \in [-1, 1]^{2N}$, there is no difference between $F(z)$ and $F(\mathrm{trnc}(z))$, and we only need to bound the difference when z is outside $[-1, 1]^{2N}$. The next claim bounds the value of $|F(z)|$ when z is outside $[-1, 1]^{2N}$.

Claim 5.1. *Let $F : \mathbb{R}^{2N} \to \mathbb{R}$ be a multilinear function that maps $\{\pm 1\}^{2N}$ to $[-1, 1]$. Let $z = (z_1, \ldots, z_{2N}) \in \mathbb{R}^{2N}$. Then, $|F(z)| \le \prod_{i=1}^{2N} \max(1, |z_i|)$.*

Proof. Recall that two multilinear functions that agree on $\{\pm 1\}^{2N}$ must be equal as functions on all $\mathbb{R}^{2N}$. Thus, we can write $F(x)$ as

$$F(x) = \sum_{w\in\{\pm1\}^{2N}} F(w) \cdot \prod_{i=1}^{2N} \frac{x_i w_i + 1}{2},$$

since these two expressions are multilinear and they agree on $\{\pm 1\}^{2N}$. By our assumption, for any fixed $w \in \{\pm 1\}^{2N}$, $F(w) \in [-1, 1]$. Thus, the value of $|F(z)|$ is at most

$$|F(z)| \le \sum_{w\in\{\pm1\}^{2N}} |F(w)| \cdot \left|\prod_{i=1}^{2N} \frac{z_i w_i + 1}{2}\right|$$
$$\le \sum_{w\in\{\pm1\}^{2N}} \prod_{i=1}^{2N} \frac{|z_i w_i + 1|}{2}$$
$$= \prod_{i=1}^{2N} \left(\frac{|z_i + 1|}{2} + \frac{|-z_i + 1|}{2}\right)$$
$$= \prod_{i=1}^{2N} \max(1, |z_i|). \qquad \square$$

We got that the value of $|F(z)|$ is bounded by $\prod_i \max(1, |z_i|)$. The following claim bounds the latter times the indicator that $z \ne \mathrm{trnc}(z)$.

Claim 5.2. $\mathbf{E}_{z\sim\mathcal{G}'}\left[\prod_{i=1}^{2N} \max(1, |z_i|) \cdot 1_{\{z\ne\mathrm{trnc}(z)\}}\right] \le 4 \cdot N^{-2}$.

Proof. Let $z = (x, y) \sim \mathcal{G}'$. For every sequence of non-negative integers $a = (a_1, \ldots, a_N)$, we consider the event

$$\forall i \in [N] : a_i \le |x_i| \le a_i + 1,$$

denoted by $\mathcal{E}_a$. For every sequence of non-negative integers $b = (b_1, \ldots, b_N)$, we consider the event

$$\forall i \in [N] : b_i \le |y_i| \le b_i + 1,$$

denoted by $\mathcal{E}'_b$. Since $x_1, \ldots, x_N$ are independent (by the definition of $\mathcal{G}'$), we have

$$\mathbf{Pr}[\mathcal{E}_a] \le \prod_{i=1}^{N} \mathbf{Pr}[|\mathcal{N}(0, \varepsilon)| \ge a_i] \le \prod_{i=1}^{N} e^{-a_i^2/(2\varepsilon)},$$

and similarly $\mathbf{Pr}[\mathcal{E}'_b] \le \prod_{i=1}^{N} e^{-b_i^2/(2\varepsilon)}$. We thus have

$$
\begin{aligned}
(*) &= \mathop{\mathbf{E}}_{z\sim\mathcal{G}'}\left[\prod_{i=1}^{2N}\max(1,|z_i|)\cdot 1_{\{z\neq\mathrm{trnc}(z)\}}\right]\\
&= \mathop{\mathbf{E}}_{(x,y)\sim\mathcal{G}'}\left[\prod_{i=1}^{N}\max(1,|x_i|)\cdot\prod_{i=1}^{N}\max(1,|y_i|)\cdot 1_{\{(x,y)\neq\mathrm{trnc}(x,y)\}}\right]\\
&\le \sum_{\substack{a\in\mathbb{N}^N,b\in\mathbb{N}^N,\\(a,b)\neq 0^{2N}}}\mathbf{Pr}[\mathcal{E}_a\wedge\mathcal{E}'_b]\cdot\prod_{i=1}^{N}(1+a_i)\cdot\prod_{i=1}^{N}(1+b_i)\\
&\le \sum_{\substack{a\in\mathbb{N}^N,b\in\mathbb{N}^N,\\(a,b)\neq 0^{2N}}}\min\left\{\mathbf{Pr}[\mathcal{E}_a],\mathbf{Pr}[\mathcal{E}'_b]\right\}\cdot\prod_{i=1}^{N}(1+a_i)\cdot\prod_{i=1}^{N}(1+b_i)\\
&\le \sum_{\substack{a\in\mathbb{N}^N,\\b\in\mathbb{N}^N,\\(a,b)\neq 0^{2N}}}\sqrt{\mathbf{Pr}[\mathcal{E}_a]\cdot\mathbf{Pr}[\mathcal{E}'_b]}\cdot\prod_{i=1}^{N}(1+a_i)\cdot\prod_{i=1}^{N}(1+b_i). \qquad (3)
\end{aligned}
$$

Since $1+a_i \le e^{a_i} \le e^{a_i^2/(8\varepsilon)}$ for $\varepsilon < 1/8$, we get

$$
\begin{aligned}
\sqrt{\mathbf{Pr}[\mathcal{E}_a]}\cdot\prod_{i=1}^{N}(1+a_i) &\le e^{-\sum_i a_i^2/(4\varepsilon)}\cdot\prod_{i=1}^{N}(1+a_i)\\
&\le e^{-\sum_i a_i^2/(8\varepsilon)}.
\end{aligned}
$$

Similarly, $\sqrt{\mathbf{Pr}[\mathcal{E}'_b]}\cdot\prod_{i=1}^{N}(1+b_i) \le e^{-\sum_i b_i^2/(8\varepsilon)}$. We plug these estimates in Expression (3):

$$
\begin{aligned}
(*) &\le \sum_{\substack{a\in\mathbb{N}^N,b\in\mathbb{N}^N,\\(a,b)\neq 0^{2N}}} e^{-\sum_i a_i^2/(8\varepsilon)}\cdot e^{-\sum_i b_i^2/(8\varepsilon)}\\
&\le \sum_{\substack{a\in\mathbb{N}^N,b\in\mathbb{N}^N,\\(a,b)\neq 0^{2N}}} e^{-\sum_i a_i/(8\varepsilon)}\cdot e^{-\sum_i b_i/(8\varepsilon)}\\
&\le \sum_{k=1}^{\infty} e^{-k/(8\varepsilon)}\cdot\left|\left\{(a,b): a\in\mathbb{N}^N, b\in\mathbb{N}^N, \sum_i a_i+b_i = k\right\}\right|\\
&\le \sum_{k=1}^{\infty} e^{-k/(8\varepsilon)}\cdot\binom{2N+k-1}{k}\\
&\le \sum_{k=1}^{\infty} e^{-k/(8\varepsilon)}\cdot(2N)^k \le \sum_{k=1}^{\infty} N^{-3k}\cdot(2N)^k \le 4\cdot N^{-2}. \qquad \square
\end{aligned}
$$

The next claim shows that a multilinear function has very similar expectation under $\mathcal{G}'$ and under the truncated variant of $\mathcal{G}'$. For the application in Section 7, we generalize the claim a bit to include any shift by a constant vector $z_0 \in [-p_0, p_0]^{2N}$ and any multiplication by a positive constant $p \in \mathbb{R}$, as long as $p + p_0 \le 1$. We shall later use the claim with $p_0 = 1/2, p \le 1/2$ in Section 7 and $(p_0, p) = (0, 1)$ in Section 6.

Claim 5.3. *Let $0 \le p, p_0$ such that $p + p_0 \le 1$. Let $F : \mathbb{R}^{2N} \to \mathbb{R}$ be a multilinear function that maps $\{\pm1\}^{2N}$ to $[-1, 1]$. Let $z_0 \in [-p_0, p_0]^{2N}$. Then,*

$$\mathop{\mathbf{E}}_{z\sim\mathcal{G}'}[|F(\mathrm{trnc}(z_0 + p\cdot z)) - F(z_0 + p\cdot z)|] \le 8\cdot N^{-2}.$$

Proof. Let $\mathcal{E}$ be the event that $(\mathrm{trnc}(z_0 + p\cdot z) \neq z_0 + p\cdot z)$. Note that $\mathcal{E}$ implies the event $z \neq \mathrm{trnc}(z)$ since $p + p_0 \le 1$. Using Claim 5.1, we get

$$
\begin{aligned}
&\mathop{\mathbf{E}}_{z\sim\mathcal{G}'}[|F(\mathrm{trnc}(z_0 + p\cdot z)) - F(z_0 + p\cdot z)|]\\
&\le \mathop{\mathbf{E}}_{z\sim\mathcal{G}'}[(1+|F(z_0 + p\cdot z)|)\cdot 1_{\mathcal{E}}]\\
&\le \mathop{\mathbf{E}}_{z\sim\mathcal{G}'}[(1+|F(z_0 + p\cdot z)|)\cdot 1_{\{z\neq\mathrm{trnc}(z)\}}]\\
&\le \mathop{\mathbf{E}}_{z\sim\mathcal{G}'}\left[\left(1+\prod_{i=1}^{2N}\max(1,|(z_0)_i + p\cdot z_i|)\right)\cdot 1_{\{z\neq\mathrm{trnc}(z)\}}\right]\\
&\le \mathop{\mathbf{E}}_{z\sim\mathcal{G}'}\left[2\cdot\prod_{i=1}^{2N}\max(1,|(z_0)_i + p\cdot z_i|)\cdot 1_{\{z\neq\mathrm{trnc}(z)\}}\right].
\end{aligned}
$$

However, $\prod_{i=1}^{2N}\max(1,|(z_0)_i + p\cdot z_i|) \le \prod_{i=1}^{2N}\max(1, p_0 + p|z_i|) \le \prod_{i=1}^{2N}\max(1,|z_i|)$. Using Claim 5.2, we get

$$
\begin{aligned}
&\mathop{\mathbf{E}}_{z\sim\mathcal{G}'}[|F(\mathrm{trnc}(z_0 + p\cdot z)) - F(z_0 + p\cdot z)|]\\
&\le \mathop{\mathbf{E}}_{z\sim\mathcal{G}'}\left[2\cdot\prod_{i=1}^{2N}\max(1,|z_i|)\cdot 1_{\{z\neq\mathrm{trnc}(z)\}}\right] \le 8\cdot N^{-2}. \qquad \square
\end{aligned}
$$

6 QUANTUM ALGORITHM DISTINGUISHING $\mathcal{D}$ AND U_{2N}

Let Q be the 1-query algorithm for Forrelation by Aaronson and Ambainis [AA15]. By [AA15, Prop.6], on a given input $x \in \{\pm1\}^N, y \in \{\pm1\}^N$, the algorithm Q accepts with probability $(1 + \varphi(x, y))/2$, where

$$\varphi(x, y) := \frac{1}{N}\cdot\sum_{i\in[N],j\in[N]} x_i\cdot H_{i,j}\cdot y_j.$$

In other words, if the algorithm outputs a $\{\pm1\}$ value, then its expected value is exactly $\varphi(x, y)$. Observe that φ is a homogeneous polynomial of degree 2 in the input variables.

Claim 6.1. $\mathbf{E}_{(x,y)\sim U_{2N}}[\varphi(x, y)] = 0.$

Proof. For any $i, j \in [N]$, we have $\mathbf{E}[x_i y_j] = 0$, under the uniform distribution. By linearity of expectation, $\mathbf{E}[\varphi(x, y)] = 0$. $\square$

Claim 6.2. $\mathbf{E}_{(x,y)\sim\mathcal{G}'}[\varphi(x, y)] = \varepsilon.$

Proof. Using the fact that $\mathbf{E}_{(x,y)\sim\mathcal{G}'}[x_i\cdot y_j] = \varepsilon\cdot H_{i,j}$, we get

$$
\begin{aligned}
\mathop{\mathbf{E}}_{(x,y)\sim\mathcal{G}'}[\varphi(x, y)] &= \tfrac{1}{N}\cdot\sum_{i\in[N],j\in[N]} H_{i,j}\cdot\mathop{\mathbf{E}}_{(x,y)\sim\mathcal{G}'}[x_i\cdot y_j]\\
&= \tfrac{1}{N}\cdot\sum_{i\in[N],j\in[N]} H_{i,j}\cdot\varepsilon\cdot H_{i,j} = \varepsilon. \qquad \square
\end{aligned}
$$

Claim 6.3. $\mathbf{E}_{(x',y')\sim\mathcal{D}}[\varphi(x', y')] \ge \varepsilon/2.$

Proof. By the multi-linearity of φ, Equation (2) and the definition of $\mathcal{G}', \mathcal{D}$ we have

$$
\begin{aligned}
&\mathop{\mathbf{E}}_{(x',y')\sim\mathcal{D}}[\varphi(x', y')]\\
&= \mathop{\mathbf{E}}_{(x,y)\sim\mathcal{G}'}[\varphi(\mathrm{trnc}(x), \mathrm{trnc}(y))]\\
&\ge \mathop{\mathbf{E}}_{(x,y)\sim\mathcal{G}'}[\varphi(x, y)] - \left|\mathop{\mathbf{E}}_{(x,y)\sim\mathcal{G}'}[\varphi(\mathrm{trnc}(x), \mathrm{trnc}(y)) - \varphi(x, y)]\right|
\end{aligned}
$$

$$\geq \varepsilon - \mathop{\mathbf{E}}_{(x,y)\sim\mathcal{G}'}[|\varphi(\mathrm{trnc}(x),\mathrm{trnc}(y)) - \varphi(x,y)|].$$

Thus it suffices to upper bound $\mathbf{E}_{(x,y)\sim\mathcal{G}'}[|\varphi(\mathrm{trnc}(x),\mathrm{trnc}(y)) - \varphi(x,y)|]$ by $\varepsilon/2$. Since $\varphi(x,y)$ is the expected value of a quantum algorithm outputting a value in $\{\pm1\}$ it is bounded in $[-1,1]$ on inputs $x,y\in\{\pm1\}^{2N}$. Since φ is multilinear, we may apply Claim 5.3 with $p_0=0$, $p=1$ and $z_0=0^{2N}$ to get

$$\mathop{\mathbf{E}}_{(x,y)\sim\mathcal{G}'}[|\varphi(\mathrm{trnc}(x),\mathrm{trnc}(y)) - \varphi(x,y)|] \leq 8\cdot N^{-2} \leq \varepsilon/2,$$

which completes the proof. □

Corollary 6.4. *There exists a quantum algorithm Q making 1-query and running in time $O(\log N)$ such that*

$$\left|\mathop{\mathbf{E}}_{z'\sim\mathcal{D}}[Q(z')] - \mathop{\mathbf{E}}_{u\sim U_{2N}}[Q(u)]\right| \geq \varepsilon/2.$$

7 $\mathcal{D}$ FOOLS BOUNDED DEPTH CIRCUITS

In the following, for a Boolean function $A:\{\pm1\}^{2N}\to\{\pm1\}$ and a vector $x\in\mathbb{R}^{2N}$, we denote by $A(x)$ the value of the multilinear extension of A on x. The multilinear extension $A:\mathbb{R}^{2N}\to\mathbb{R}$ can be written as

$$A(x) = \sum_{S\subseteq[2N]} \widehat{A}(S)\cdot\prod_{i\in S} x_i. \tag{4}$$

Observe that $A(\vec{0}) = \widehat{A}(\emptyset) = \mathbf{E}_{x\sim U_{2N}}[A(x)]$.

We use the following result of Tal [Tal17]:

Lemma 7.1 ([Tal17, Thm. 37])**.** *There exists a universal constant $c>0$ such that the following holds. Let $A:\{\pm1\}^{2N}\to\{\pm1\}$ be a Boolean circuit with at most s gates and depth at most d. Then, for all $k\in\mathbb{N}$, we have $\sum_{S\subseteq[2N]:|S|=k}|\widehat{A}(S)| \leq (c\cdot\log s)^{(d-1)k}$.*

For two vectors $R,Q\in\mathbb{R}^{2N}$ we denote by $R\circ Q\in\mathbb{R}^{2N}$ their point-wise product, that is $(R\circ Q)_i = R_i\cdot Q_i$ for all $i\in[2N]$.

Claim 7.2. *Let $p\leq 1/2$. Let $A:\{\pm1\}^{2N}\to\{\pm1\}$ be a Boolean circuit of size at most s and depth at most d, such that $\sqrt{\varepsilon}p\cdot(c\cdot\log s)^{d-1}\leq 1/2$. Let $P\in[-p,p]^{2N}$. Then,*

$$\left|\mathop{\mathbf{E}}_{z\sim\mathcal{G}'}[A(P\circ z)] - A(\vec{0})\right| \leq 3\varepsilon\cdot p^2\cdot(c\cdot\log s)^{2(d-1)}\cdot N^{-1/2}.$$

Proof. By Equation (4) and since $A(\vec{0}) = \widehat{A}(\emptyset)$,

$$\begin{aligned}
&\left|\mathop{\mathbf{E}}_{z\sim\mathcal{G}'}[A(P\circ z)] - A(\vec{0})\right| \\
&= \left|\mathop{\mathbf{E}}_{z\sim\mathcal{G}'}\left[\sum_{\emptyset\neq S\subseteq[2N]}\widehat{A}(S)\cdot\prod_{i\in S}P_i\cdot z_i\right]\right| \\
&= \left|\sum_{\emptyset\neq S\subseteq[2N]}\widehat{A}(S)\cdot\prod_{i\in S}P_i\cdot\mathop{\mathbf{E}}_{z\sim\mathcal{G}'}\left[\prod_{i\in S}z_i\right]\right| \\
&= \left|\sum_{\emptyset\neq S\subseteq[2N]}\widehat{A}(S)\cdot\prod_{i\in S}P_i\cdot\widehat{\mathcal{G}'}(S)\right| \\
&\leq \sum_{\emptyset\neq S\subseteq[2N]}|\widehat{A}(S)|\cdot p^{|S|}\cdot\sqrt{\varepsilon}^{|S|}\cdot|\widehat{\mathcal{G}}(S)| \\
&\leq \sum_{k=1}^{2N}(\sqrt{\varepsilon}p)^k\cdot\left(\max_{S:|S|=k}|\widehat{\mathcal{G}}(S)|\right)\cdot\sum_{S\subseteq[2N],|S|=k}|\widehat{A}(S)| \\
&\text{(by Lemma 7.1)}\quad \leq \sum_{k=1}^{2N}(\sqrt{\varepsilon}p)^k\cdot\left(\max_{S:|S|=k}|\widehat{\mathcal{G}}(S)|\right)\cdot(c\log s)^{(d-1)k} \\
&= \sum_{k=1}^{2N}q^k\cdot\left(\max_{S:|S|=k}|\widehat{\mathcal{G}}(S)|\right)
\end{aligned}$$

where $q := \sqrt{\varepsilon}\cdot p\cdot(c\log s)^{d-1}$. For odd k, Claim 4.1 gives $\max_{S:|S|=k}|\widehat{\mathcal{G}}(S)| = 0$. For $k=2\ell$, $\ell\leq\lfloor n/2\rfloor$, Claim 4.1 gives $\max_{S:|S|=2\ell}|\widehat{\mathcal{G}}(S)|\leq\ell!\cdot N^{-\ell/2}$. For $k=2\ell$, $\ell\geq\lfloor n/2\rfloor+1$, we have $\max_{S:|S|=2\ell}|\widehat{\mathcal{G}}(S)|\leq 1$. Plugging these bounds in the above expression, gives

$$\left|\mathop{\mathbf{E}}_{z\sim\mathcal{G}'}[A(P\circ z)] - A(\vec{0})\right| \leq \sum_{\ell=1}^{\lfloor n/2\rfloor}q^{2\ell}\cdot\ell!\cdot N^{-\ell/2} + \sum_{\ell=\lfloor n/2\rfloor+1}^{N}q^{2\ell}.$$

Observe that each two consecutive elements in each sequence above are decreasing by at least a factor of 2 (since $q=\sqrt{\varepsilon}p\cdot(c\cdot\log s)^{d-1}\leq 1/2$). Thus, the sum is bounded by $2\cdot q^2\cdot N^{-1/2} + 2\cdot q^{n+1} \leq 3q^2\cdot N^{-1/2}$. □

Claim 7.3. *Let $p\leq 1/4$. Let $A:\{\pm1\}^{2N}\to\{\pm1\}$ be a Boolean circuit of size s and depth d, such that $\sqrt{\varepsilon}p\cdot(c\cdot\log s)^{d-1}\leq 1/4$. Let $z_0\in[-1/2,1/2]^{2N}$. Then,*

$$\left|\mathop{\mathbf{E}}_{z\sim\mathcal{G}'}[A(z_0+p\cdot z)] - A(z_0)\right| \leq 12\varepsilon\cdot p^2\cdot(c\cdot\log s)^{2(d-1)}\cdot N^{-1/2}.$$

The proof is similar to the proof of [CHHL18, Claim 3.3], relying on the fact that restrictions of A are also Boolean circuits of size at most s and depth at most d.

Proof. Given z_0, we define a distribution $\mathcal{R}_{z_0}$ over restrictions $\rho\in\{-1,1,*\}^{2N}$, as follows. For each entry $i\in[2N]$ independently, we set $\rho_i = \mathrm{sgn}((z_0)_i)$ with probability $|(z_0)_i|$ and $\rho_i = *$ otherwise.

Define $P\in[-2p,2p]^{2N}$ by $P_i = p\cdot\frac{1}{1-|(z_0)_i|}$ for $i\in[2N]$.

Let $\rho\sim\mathcal{R}_{z_0}$. Next, for any vector $z\in\mathbb{R}^{2N}$, we define a vector $\widetilde{z} = \widetilde{z}(z,\rho)\in\mathbb{R}^{2N}$, as follows:

$$\widetilde{z}_i = \begin{cases}\rho_i & \text{if } \rho_i\in\{\pm1\}\\ P_i\cdot z_i & \text{otherwise}\end{cases}$$

Thus, for a fixed $z\in\mathbb{R}^{2N}$, the vector $\widetilde{z}$ is a random variable that depends on ρ. We show that for any fixed $z\in\mathbb{R}^{2N}$, the distribution of the random variable $\widetilde{z}$ is a product distribution (over inputs in $\mathbb{R}^{2N}$), and the expectation of $\widetilde{z}$ is the vector $z_0+p\cdot z$. Indeed, each coordinate $\widetilde{z}_i$ is independent of the other coordinates, and its expected value is

$$\mathop{\mathbf{E}}_{\rho\sim\mathcal{R}_{z_0}}[\widetilde{z}_i] = |(z_0)_i|\cdot\mathrm{sgn}((z_0)_i) + (1-|(z_0)_i|)\cdot P_i\cdot z_i = (z_0)_i + p\cdot z_i.$$

Hence, since A is multi-linear and $\widetilde{z}$ has a product distribution, by Equation (4), $\mathbf{E}_{\rho\sim\mathcal{R}_{z_0}}[A(\widetilde{z})] = A(z_0+p\cdot z)$.

Let $z\sim\mathcal{G}'$. We get

$$\begin{aligned}
&\left|\mathop{\mathbf{E}}_{z\sim\mathcal{G}'}[A(z_0+p\cdot z)] - A(z_0)\right| \\
&= \left|\mathop{\mathbf{E}}_{z\sim\mathcal{G}'}\mathop{\mathbf{E}}_{\rho\sim\mathcal{R}_{z_0}}\left[A(\widetilde{z}(z,\rho)) - A(\widetilde{z}(\vec{0},\rho))\right]\right| \\
&\leq \mathop{\mathbf{E}}_{\rho\sim\mathcal{R}_{z_0}}\left[\left|\mathop{\mathbf{E}}_{z\sim\mathcal{G}'}\left[A(\widetilde{z}(z,\rho))\right] - A(\widetilde{z}(\vec{0},\rho))\right|\right]
\end{aligned}$$

However, for any fixed ρ, we have $A(\widetilde{z}(z,\rho)) = A_\rho(P \circ z)$, where A_ρ is attained from A by fixing the coordinates that were fixed in ρ, according to ρ. Thus,

$$\left|\mathop{\mathbf{E}}_{z\sim\mathcal{G}'}[A(z_0 + p\cdot z)] - A(z_0)\right| \le \mathop{\mathbf{E}}_{\rho\sim\mathcal{R}_{z_0}}\left[\left|\mathop{\mathbf{E}}_{z\sim\mathcal{G}'}[A_\rho(P\circ z)] - A_\rho(\vec{0})\right|\right]$$

and we may apply Claim 7.2 to get that for any fixed ρ we have

$$\left|\mathop{\mathbf{E}}_{z\sim\mathcal{G}'}[A_\rho(P\circ z)] - A_\rho(\vec{0})\right| \le 3\cdot\varepsilon\cdot(2p)^2\cdot(c\log s)^{2(d-1)}\cdot N^{-1/2}$$

using the fact that $P \in [-2p,2p]^{2N}$ and the assumption $\sqrt{\varepsilon}p(c\cdot\log s)^{d-1} \le 1/4$. □

Theorem 7.4. *Let $A : \{\pm1\}^{2N} \to \{\pm1\}$ be a Boolean circuit of size s and depth d. Then, $\left|\mathbf{E}_{z'\sim\mathcal{D}}[A(z')] - A(\vec{0})\right| \le 32\varepsilon\cdot(c\cdot\log s)^{2(d-1)}\cdot N^{-1/2}$.*

Proof. First, we can assume without loss of generality that $\sqrt{\varepsilon}\cdot(c\cdot\log s)^{d-1} \le \frac{1}{4}\cdot N^{1/4}$, as otherwise the claim is vacuous (as the LHS is at most 2 and the RHS is bigger than 2).

Let $t := N$, $p := 1/\sqrt{t} = N^{-1/2}$. Note that $\sqrt{\varepsilon}p(c\cdot\log s)^{d-1} \le \frac{1}{4}\cdot N^{-1/4} \le 1/4$. Let $z^{(1)},\ldots,z^{(t)} \sim \mathcal{G}'$. For $i = 0,\ldots,t$, let $z^{\le(i)} = p\cdot(z^{(1)}+\ldots+z^{(i)})$. The main observation is that $z^{\le(t)} \sim \mathcal{G}'$. This is since the distribution of $z^{\le(t)}$ is a multivariate gaussian distribution with the same expectation and the same covariance matrix as $\mathcal{G}'$. Thus, by Equation (2), it will be sufficient to bound

$$\left|\mathbf{E}[A(\mathrm{trnc}(z^{\le(t)}))] - A(\vec{0})\right|.$$

We do so by induction, by the triangle inequality: for $i = 0,\ldots,t-1$, we will show

$$\left|\mathbf{E}[A(\mathrm{trnc}(z^{\le(i+1)}))] - \mathbf{E}[A(\mathrm{trnc}(z^{\le(i)}))]\right| \le 12\cdot\varepsilon p^2\cdot(c\cdot\log s)^{2(d-1)}\cdot N^{-1/2} + 12\cdot N^{-2}.$$

For $i \in \{0,\ldots,t-1\}$, let E_i be the event that $z^{\le(i)} \in [-1/2,1/2]^{2N}$. Since (for $i \ge 1$) $\frac{z^{\le(i)}}{p\sqrt{i}} \sim \mathcal{G}'$, each j-th entry in $z^{\le(i)}$ is distributed $\mathcal{N}(0,p^2i\varepsilon)$ and we have

$$\mathbf{Pr}[|(z^{\le(i)})_j| \ge 1/2] \le \mathbf{Pr}[|\mathcal{N}(0,\varepsilon)| \ge 1/2] \le e^{-1/(8\varepsilon)} \le N^{-3}.$$

By the union bound, we have $\mathbf{Pr}[E_i] \ge 1 - 2N\cdot(N^{-3}) = 1 - 2N^{-2}$.

By Claim 7.3, used with $z_0 = z^{\le(i)}$, we have that conditioned on the event E_i,

$$\left|\mathbf{E}[A(z^{\le(i+1)})|E_i] - \mathbf{E}[A(z^{\le(i)})|E_i]\right| \le 12\varepsilon\cdot p^2\cdot(c\cdot\log s)^{2(d-1)}\cdot N^{-1/2}.$$

We wish to show a similar bound on the truncated version of $z^{\le(i+1)}$. Note that conditioned on E_i, we have $z^{\le(i)} = \mathrm{trnc}(z^{\le(i)})$, but this is not necessarily the case for $z^{\le(i+1)}$. Using Claim 5.3 with $p_0 = 1/2$ and $p \le 1/2$ we get $\left|\mathbf{E}\left[|A(\mathrm{trnc}(z^{\le(i+1)})) - A(z^{\le(i+1)})| \;\middle|\; E_i\right]\right| \le 8\cdot N^{-2}$. By the triangle inequality we get

$$\begin{aligned}&\left|\mathbf{E}[A(\mathrm{trnc}(z^{\le(i+1)}))|E_i] - \mathbf{E}[A(\mathrm{trnc}(z^{\le(i)}))|E_i]\right|\\ &\le \left|\mathbf{E}[A(\mathrm{trnc}(z^{\le(i+1)}))|E_i] - \mathbf{E}[A(z^{\le(i+1)})|E_i]\right|\\ &\quad + \left|\mathbf{E}[A(z^{\le(i+1)})|E_i] - \mathbf{E}[A(\mathrm{trnc}(z^{\le(i)}))|E_i]\right|\\ &\le 8\cdot N^{-2} + 12\varepsilon\cdot p^2\cdot(c\cdot\log s)^{2(d-1)}\cdot N^{-1/2}.\end{aligned}$$

When E_i does not hold, the difference between $A(\mathrm{trnc}(z^{\le(i+1)}))$ and $A(\mathrm{trnc}(z^{\le(i)}))$ is at most 2 since A maps $[-1,1]^{2N}$ to $[-1,1]$. Thus,

$$\begin{aligned}&\left|\mathbf{E}[A(\mathrm{trnc}(z^{\le(i+1)}))] - \mathbf{E}[A(\mathrm{trnc}(z^{\le(i)}))]\right|\\ &\le \left|\mathbf{E}[A(\mathrm{trnc}(z^{\le(i+1)}))|E_i] - \mathbf{E}[A(\mathrm{trnc}(z^{\le(i)}))|E_i]\right| + 2\cdot\mathbf{Pr}[\neg E_i]\\ &\le 12\varepsilon\cdot p^2\cdot(c\cdot\log s)^{2(d-1)}\cdot N^{-1/2} + 12\cdot N^{-2}\end{aligned}$$

By Equation (2) and the triangle inequality,

$$\begin{aligned}&\left|\mathop{\mathbf{E}}_{z'\sim\mathcal{D}}[A(z')] - A(\vec{0})\right|\\ &= \left|\mathop{\mathbf{E}}_{z\sim\mathcal{G}'}[A(\mathrm{trnc}(z))] - A(\vec{0})\right| = \left|\mathbf{E}[A(\mathrm{trnc}(z^{\le t}))] - A(\vec{0})\right|\\ &\le \sum_{i=0}^{t-1}\left|\mathbf{E}[A(\mathrm{trnc}(z^{\le(i+1)}))] - \mathbf{E}[A(\mathrm{trnc}(z^{\le(i)}))]\right|\\ &\le t\cdot 12\varepsilon\cdot p^2\cdot(c\cdot\log s)^{2(d-1)}\cdot N^{-1/2} + 12t\cdot N^{-2}\\ &= 12\varepsilon\cdot(c\cdot\log s)^{2(d-1)}\cdot N^{-1/2} + 12\cdot N^{-1}\\ &\le 32\varepsilon\cdot(c\cdot\log s)^{2(d-1)}\cdot N^{-1/2},\end{aligned}$$

which completes the proof. □

The following is an immediate corollary to Theorem 7.4.

Corollary 7.5. *Let $A : \{\pm1\}^{2N} \to \{\pm1\}$ be a Boolean circuit of size $\exp(\log^{O(1)}(N))$ and depth $O(1)$. Then, $\left|\mathbf{E}_{z'\sim\mathcal{D}}[A(z')] - \mathbf{E}_{u\sim U_{2N}}[A(u)]\right| \le \mathrm{polylog}(N)/\sqrt{N}$.*

8 PROOFS OF THE MAIN THEOREMS

8.1 Proof of Theorem 1.1

Proof. The first part of the theorem is Corollary 6.4. The second part is Corollary 7.5. □

8.2 Proof of Theorem 1.2

Proof. Let $\delta \in (0,1)$ be such that $\delta \ge 2^{-\mathrm{polylog}(N)}$. Let $m = 32\cdot\frac{\ln(1/\delta)}{\varepsilon^2} = \mathrm{polylog}(N)$. We take $N_1 = 2Nm$. Note that $\mathrm{polylog}(N_1) = \mathrm{polylog}(N)$. We take $\mathcal{D}_1 = \mathcal{D}^{\otimes m}$, that is, the distribution over $\{\pm1\}^{2N\cdot m}$, generated by taking a concatenation of m independent random variables with distribution $\mathcal{D}$. The first part of the theorem follows by the following claim.

Claim 8.1. *There exists a quantum algorithm Q_1 making $O(m)$ queries and running in time $O(m\cdot\log N)$, such that, $\mathbf{Pr}_{z\sim\mathcal{D}_1}[Q_1(z)$ accepts$] \ge 1-\delta$ and $\mathbf{Pr}_{z\sim U_{N_1}}[Q_1(z)$ accepts$] \le \delta$.*

Proof. The claim is proved by amplifying the advantage of the quantum algorithm in Theorem 1.1, by making $m = \mathrm{polylog}(N)$ sequential repetitions, with fresh quantum states for each repetition. Since the repetitions are sequential with fresh quantum states, the output of each repetition is an independent random variable. Thus, by Chernoff's bound, the probability that the algorithm successfully distinguishes between $\mathcal{D}_1$ and the uniform distribution is close to 1.

Formally, let $z = (z^{(1)},\ldots,z^{(m)}) \in \{\pm1\}^{m\cdot2N}$, where each $z^{(i)} \in \{\pm1\}^{2N}$. We run the quantum algorithm Q from Theorem 1.1 on each $z^{(i)}$ sequentially, and take a measurement after each run.

Given the results of the m runs, denoted $r \in \{\pm 1\}^m$, we compute $S = \sum_{i=1}^m r_i$ and accept if and only if $S \geq m \cdot \varepsilon/4$. The algorithm runs in time $O(m \cdot \log N)$ as it runs m times the algorithm Q and then performs an addition of m bits. The algorithm makes m queries to the input.

On a uniform input, the string r is distributed uniformly at random on $\{\pm 1\}^m$, thus by Chernoff's bound the probability that the algorithm accepts is at most $e^{-m(\varepsilon/4)^2/2} \leq \delta$.

On input $z \sim \mathcal{D}_1$, the string r is distributed as a product distribution of random variables taking values in $\{\pm 1\}$ with expectation $\geq \varepsilon/2$. By Chernoff's bound, the probability that the algorithm accepts is at least $1 - e^{-m(\varepsilon/4)^2/2} \geq 1 - \delta$. □

The second part of the theorem follows by the following claim.

Claim 8.2. *Let A be any Boolean circuit of size s and depth d. Then,*

$$\left| \mathop{\mathbf{E}}_{z \sim \mathcal{D}_1}[A(z)] - \mathop{\mathbf{E}}_{u \sim U_{N_1}}[A(u)] \right| \leq m \cdot 32\varepsilon \cdot (c \cdot \log s)^{2(d-1)} \cdot N^{-1/2}.$$

Proof. We apply a hybrid argument. We define $m+1$ hybrids: $U_{2Nm} = H_0, H_1, \ldots, H_m = \mathcal{D}^{\otimes m}$, by taking $H_i = \mathcal{D}^{\otimes i} \otimes (U_{2N})^{\otimes(m-i)}$. We bound the difference between $\mathbf{E}_{z \sim H_{i-1}}[A(z)]$ and $\mathbf{E}_{z' \sim H_i}[A(z')]$, for $i \in \{1, \ldots, m\}$. Let $z = (z^{(1)}, \ldots, z^{(i-1)}, u^{(i)}, u^{(i+1)}, \ldots, u^{(m)}) \sim H_{i-1}$. Let $z' = (z^{(1)}, \ldots, z^{(i-1)}, z^{(i)}, u^{(i+1)}, \ldots, u^{(m)}) \sim H_i$. For a partial assignment to all m parts except the i-th part, $a_{-i} = (a^{(1)}, \ldots, a^{(i-1)}, a^{(i+1)}, \ldots, a^{(m)})$, we denote by $A|_{a_{-i}}(w)$ the value of A on $(a^{(1)}, \ldots, a^{(i-1)}, w, a^{(i+1)}, \ldots, a^{(m)})$. By an averaging argument, there exists a string a_{-i}, such that,

$$\begin{aligned} &\left| \mathop{\mathbf{E}}_{z' \sim H_i}[A(z')] - \mathop{\mathbf{E}}_{z \sim H_{i-1}}[A(z)] \right| \\ &\leq \left| \mathop{\mathbf{E}}_{z^{(i)} \sim \mathcal{D}}[A|_{a_{-i}}(z^{(i)})] - \mathop{\mathbf{E}}_{u^{(i)} \sim U_{2N}}[A|_{a_{-i}}(u^{(i)})] \right|. \end{aligned}$$

For any fixed string a_{-i}, the restricted function $A|_{a_{-i}}$ is a Boolean circuit of size at most s and depth at most d. Thus, we can apply Theorem 7.4 to get that

$$\begin{aligned} &\left| \mathop{\mathbf{E}}_{z^{(i)} \sim \mathcal{D}}[A|_{a_{-i}}(z^{(i)})] - \mathop{\mathbf{E}}_{u^{(i)} \sim U_{2N}}[A|_{a_{-i}}(u^{(i)})] \right| \\ &\qquad \leq 32\varepsilon \cdot (c \cdot \log s)^{2(d-1)} \cdot N^{-1/2}. \end{aligned}$$

The proof follows by a triangle inequality. □

□

8.3 Proof of Theorem 1.3

Proof. Let $\delta \in (0,1)$ be such that $\delta = 2^{-\mathrm{polylog}(N)}$. Let $m = 32 \cdot \frac{\ln(1/\delta)}{\varepsilon^2} = \mathrm{polylog}(N)$. Let $N' = (2N)^m$. Note that $\mathrm{polylog}(N') = \mathrm{polylog}(N)$.

Define the distribution $\mathcal{D}'$ over $\{\pm 1\}^{N'}$, generated as follows: Let $z^{(1)}, \ldots, z^{(m)}$ be m independent random variables with distribution $\mathcal{D}$. Output $z := z^{(1)} \otimes \ldots \otimes z^{(m)}$, where $\otimes$ denotes tensor product, that is, for $i_1, \ldots, i_m \in [2N]$, $z_{i_1,\ldots,i_m} = \prod_{j=1}^m z^{(j)}_{i_j}$.

Define the distribution U' over $\{\pm 1\}^{N'}$, generated as follows: Let $u^{(1)}, \ldots, u^{(m)}$ be m independent random variables with distribution U_{2N}. Output $u := u^{(1)} \otimes \ldots \otimes u^{(m)}$, that is, for $i_1, \ldots, i_m \in [2N]$, $u_{i_1,\ldots,i_m} = \prod_{j=1}^m u^{(j)}_{i_j}$.

Claim 8.3. *There exists a quantum algorithm Q' making one query and running in time $\mathrm{polylog}(N)$, such that, $\mathbf{Pr}_{z \sim \mathcal{D}'}[Q'(z)$ accepts$] \geq 1 - \delta$ and $\mathbf{Pr}_{z \sim U'}[Q'(z)$ accepts$] \leq \delta$.*

Proof. Let Q be the one-query quantum algorithm from Theorem 1.1, for distinguishing between $\mathcal{D}$ and U_{2N}. Intuitively, the algorithm Q' will run m unentangled copies of Q in parallel.

Recall that for an input $z \in \{\pm 1\}^{2N}$, a quantum query to the input performs the diagonal unitary transformation $\mathcal{U}_z$, defined by

$$|i, w\rangle \to z_i |i, w\rangle,$$

where $i \in [2N]$ and w represents the auxiliary workspace that does not participate in the query. Thus, for an input $z = z^{(1)} \otimes \ldots \otimes z^{(m)} \in \{\pm 1\}^{N'}$, a quantum query to the input performs the diagonal unitary transformation $\mathcal{U}_z$, defined by

$$|i_1, \ldots, i_m, w\rangle \to z_{i_1,\ldots,i_m} |i_1, \ldots, i_m, w\rangle,$$

where $i_1, \ldots, i_m \in [2N]$ and w represents the auxiliary workspace. Since $z_{i_1,\ldots,i_m} = \prod_{j=1}^m z^{(j)}_{i_j}$,

$$z_{i_1,\ldots,i_m} |i_1, \ldots, i_m, w\rangle = \left(z^{(1)}_{i_1} |i_1\rangle\right) \otimes \ldots \otimes \left(z^{(m)}_{i_m} |i_m\rangle\right) \otimes |w\rangle.$$

Thus, for an input $z = z^{(1)} \otimes \ldots \otimes z^{(m)} \in \{\pm 1\}^{N'}$,

$$\mathcal{U}_z = \mathcal{U}_{z^{(1)}} \otimes \ldots \otimes \mathcal{U}_{z^{(m)}},$$

where $\otimes$ represents tensor product of operators.

Note that the algorithm Q' is promised that the input $z \in \{\pm 1\}^{N'}$ satisfies $z = z^{(1)} \otimes \ldots \otimes z^{(m)}$, for some $z^{(1)}, \ldots, z^{(m)} \in \{\pm 1\}^{2N}$, as this is the case in both distributions $\mathcal{D}'$ and U'.

Assume that the algorithm Q applies a query transformation on the state $|\Psi\rangle$. The algorithm Q' will prepare m unentangled copies of $|\Psi\rangle$ (by applying m times the procedure run by Q to prepare $|\Psi\rangle$) and obtain the state $|\Psi\rangle_1 \otimes \ldots \otimes |\Psi\rangle_m$. Next, Q' applies a query transformation on that state and, assuming that $z = z^{(1)} \otimes \ldots \otimes z^{(m)}$, obtains the state

$$\left(\mathcal{U}_{z^{(1)}} |\Psi\rangle_1\right) \otimes \ldots \otimes \left(\mathcal{U}_{z^{(m)}} |\Psi\rangle_m\right).$$

Finally, Q' takes the same measurement as Q, on each of the m unentangled states separately. Since the states are unentangled, the measurements give independent results.

Given the results of the m measurements, denoted $r \in \{\pm 1\}^m$, we compute $S = \sum_{i=1}^m r_i$ and accept if and only if $S \geq m \cdot \varepsilon/4$. The algorithm Q' runs in time $\mathrm{polylog}(N)$ as it runs m times the algorithm Q. The algorithm makes one query to the input.

On input $z \sim U'$, the string r is distributed uniformly at random on $\{\pm 1\}^m$, thus by Chernoff's bound the probability that the algorithm accepts is at most $e^{-m(\varepsilon/4)^2/2} \leq \delta$.

On input $z \sim \mathcal{D}'$, the string r is distributed as a product distribution of random variables taking values in $\{\pm 1\}$ with expectation $\geq \varepsilon/2$. By Chernoff's bound, the probability that the algorithm accepts is at least $1 - e^{-m(\varepsilon/4)^2/2} \geq 1 - \delta$. □

Claim 8.4. *Let A be any Boolean circuit of size* quasipoly(N) *and constant depth. Then,*

$$\left| \mathop{\mathbf{E}}_{z\sim\mathcal{D}'}[A(z)] - \mathop{\mathbf{E}}_{u\sim U'}[A(u)] \right| \leq \text{polylog}(N)\cdot N^{-1/2}.$$

Proof. The proof is by a reduction to Claim 8.2. Let A be any Boolean circuit of size quasipoly(N) and constant depth, and denote

$$\alpha = \left| \mathop{\mathbf{E}}_{z\sim\mathcal{D}'}[A(z)] - \mathop{\mathbf{E}}_{u\sim U'}[A(u)] \right|.$$

We will construct a Boolean circuit A' of size quasipoly(N) and constant depth, such that,

$$\left| \mathop{\mathbf{E}}_{z\sim\mathcal{D}_1}[A'(z)] - \mathop{\mathbf{E}}_{u\sim U_{N_1}}[A'(u)] \right| = \alpha,$$

where $\mathcal{D}_1, N_1$ are as in the proof of Theorem 1.2. The proof hence follows by Claim 8.2.

The circuit A' gets as input $z = (z^{(1)},\ldots,z^{(m)}) \in \{\pm1\}^{m\cdot 2N}$, computes $z' := z^{(1)}\otimes\cdots\otimes z^{(m)}$ and outputs $A(z')$. Note that by the definitions of $\mathcal{D}_1, U_{N_1}, \mathcal{D}', U'$, if $z\sim\mathcal{D}_1$, then $z'\sim\mathcal{D}'$ and if $z\sim U_{N_1}$, then $z'\sim U'$. Thus,

$$\left| \mathop{\mathbf{E}}_{z\sim\mathcal{D}_1}[A'(z)] - \mathop{\mathbf{E}}_{u\sim U_{N_1}}[A'(u)] \right| = \alpha.$$

Note that each bit in z' is the XOR of m bits in the inputs $(z^{(1)},\ldots,z^{(m)})$. Since $m = \text{polylog}(N)$, the computation $z' := z^{(1)}\otimes\cdots\otimes z^{(m)}$ can be done by a circuit of size poly(N) and constant depth. Thus, A' is a Boolean circuit of size quasipoly(N) and constant depth. Therefore, by Claim 8.2, $\alpha \leq \text{polylog}(N)\cdot N^{-1/2}$. □

We are now ready to define the distributions $\mathcal{D}_2$ and $\widetilde{U}$ and complete the proof of the theorem. Note that in Claim 8.3, the running time of the algorithm Q' is polylog(N) and not $O(\log N')$ as needed. Nevertheless, this is easy to fix by a padding argument. Assume that the running time of Q' is at most $(\log N)^c$, where c is a constant. Let $N_2 = N' + 2^{(\log N)^c}$. Let $\mathcal{D}_2$ be $\mathcal{D}'$, padded by $2^{(\log N)^c}$ ones, and let $\widetilde{U}$ be U', padded by $2^{(\log N)^c}$ ones. Note that $\text{polylog}(N_2) = \text{polylog}(N)$, that is, $N = 2^{(\log N_2)^{\Omega(1)}}$.

The theorem thus follows by Claim 8.3 and Claim 8.4. □

A ORACLE SEPARATION RESULT

In this section, we prove that using the distribution $\mathcal{D}_1$, one can construct an oracle O such that $\mathbf{BQP}^O \not\subseteq \mathbf{PH}^O$. The proof was essentially given in the work of Aaronson [Aar10] and Fefferman et al. [FSUV13, Section 2.6] (based on [BG81]). We repeat it here for completeness.

Proof of Corollary 1.5. We view the oracle O as encoding the truth-tables of Boolean functions of different input lengths. As in the proof of Theorem 1.2, for each $n\in\mathbb{N}$, let $N = 2^n$, $\varepsilon = \frac{1}{24\ln(N)}$, $\delta = \frac{1}{n^2}$, $m = 32\cdot\lceil\frac{\ln(1/\delta)}{\varepsilon^2}\rceil$ and $N_1 = 2N\cdot m$. Note that N_1 is a function of n and we denote it also as $N_1(n)$. With probability $1/2$ we draw $x_n\in\{\pm1\}^{N_1}$ from the uniform distribution U_{N_1}, and with probability $1/2$ we draw $x_n\in\{\pm1\}^{N_1}$ from the distribution $\mathcal{D}_1$. We interpret $x_n\in\{\pm1\}^{N_1}$ as a Boolean function $f_n : \{\pm1\}^{\lceil\log(N_1(n))\rceil}\to\{\pm1\}$ that describes the oracle O restricted to strings of length $\lceil\log(N_1(n))\rceil$ (note that $\lceil\log(N_1(n))\rceil$ is strictly increasing in n). Let L be the unary language consisting of all 1^n for which x_n was drawn from the distribution $\mathcal{D}_1$.

Using Claim 8.1, we show that there exists a $\mathbf{BQP}^O$ machine M that decides L on all but finitely many values of n. The machine M on input 1^n would run the quantum algorithm Q_1 from Claim 8.1 on the oracle string provided by O of length $N_1(n)$ and would accept/reject according to Q_1. Note that this is a $\mathbf{BQP}$ machine since Q_1 runs in polylog(N) = poly(n) time. We show that with high probability over the choices of O, the machine M decides L correctly on all but finitely many inputs. Indeed, for sufficiently large n:

(1) If $1^n\in L$, then x_n was sampled from $\mathcal{D}_1$, and the probability that Q_1 accepts x_n is at least $1-1/n^2$.
(2) If $1^n\notin L$, then x_n was sampled from U_{N_1}, and the probability that Q_1 accepts x_n is at most $1/n^2$.

We see that in both cases the probability (over the choices of O and the randomness of M's measurements) that M^O decides L correctly on 1^n is at least $1-1/n^2$. Let $n_0\in\mathbb{N}$ be sufficiently large. Then,

$$\Pr_{M,O}[M^O \text{ decides } L \text{ correctly on } 1^n \text{ for all } n\geq n_0] \geq \prod_{n\geq n_0}(1-1/n^2) \geq 0.9,$$

(where $\mathbf{Pr}_{M,O}$ denotes the probability over the choices of O and the randomness of M's measurements). By averaging, with probability at least 0.5 over the choice of O, we have

$$\Pr_{M}[M^O \text{ decides } L \text{ correctly on } 1^n \text{ for all } n\geq n_0] \geq 0.8.$$

On the other hand, for any fixed $\mathbf{PH}$ machine A and fixed oracle O, let $E_n(A,O)$ be the event that A^O decides L correctly on 1^n. By Theorem 1.2 for sufficiently large n, we have $\mathbf{Pr}_O[E_n(A,O)]\leq 0.51$, since we may reinterpret A^O on 1^n as a Boolean circuit of size at most $2^{\text{poly}(n)}$ = quasipoly(N) and constant depth. By independence of O on different input lengths, and the fact that A can only ask queries of length poly(n) on input 1^n, we get that there are infinitely many input lengths $n_1, n_2, \ldots$ such that for each $i\in\mathbb{N}$, $\mathbf{Pr}_O[E_{n_{i+1}}(A,O)|E_{n_1}(A,O)\wedge\ldots\wedge E_{n_i}(A,O)]\leq 0.51$. We get that

$$\Pr_{O}[E_1(A,O)\wedge E_2(A,O)\wedge\ldots] = 0,$$

and since there are countably many $\mathbf{PH}$ machines we have

$$\Pr_{O}[\exists A : E_1(A,O)\wedge E_2(A,O)\wedge\ldots] = 0.$$

Overall, we got that with probability at least 0.5 over the choice of O, M^O decides L correctly on 1^n for all $n\geq n_0$, and no $\mathbf{PH}^O$ machine decides L correctly on 1^n for all $n\geq 1$. Thus, there exists an oracle O where both events happen. Fixing the oracle O, we may hardwire the values of L on 1^n for $n<n_0$ to M, making it a $\mathbf{BQP}^O$ machine that decides L correctly on 1^n for all $n\geq 1$. □

ACKNOWLEDGEMENTS

We would like to thank Scott Aaronson, Shalev Ben-David, Mika Göös, Johan Håstad, Pooya Hatami, and Toni Pitassi for very helpful discussions.

REFERENCES

[Aar10] Scott Aaronson: BQP and the polynomial hierarchy. STOC 2010: 141-150

[Aar11] Scott Aaronson: A Counterexample to the Generalized Linial-Nisan Conjecture. CoRR abs/1110.6126 (2011)

[AA15] Scott Aaronson, Andris Ambainis: Forrelation: A Problem that Optimally Separates Quantum from Classical Computing. STOC 2015: 307-316

[BG81] Charles H. Bennett, John Gill: Relative to a Random Oracle A, P^A != NP^A != co-NP^A with Probability 1. SIAM J. Comput. 10(1): 96-113 (1981)

[BV97] Ethan Bernstein, Umesh V. Vazirani: Quantum Complexity Theory. SIAM J. Comput. 26(5): 1411-1473 (1997)

[Chen16] Lijie Chen: A Note on Oracle Separations for BQP. CoRR abs/1605.00619 (2016)

[CHHL18] Eshan Chattopadhyay, Pooya Hatami, Kaave Hosseini, Shachar Lovett: Pseudorandom Generators from Polarizing Random Walks. Electronic Colloquium on Computational Complexity (ECCC) 25: 15 (2018)

[FSS84] Merrick L. Furst, James B. Saxe, Michael Sipser: Parity, Circuits, and the Polynomial-Time Hierarchy. Mathematical Systems Theory 17(1): 13-27 (1984)

[FSUV13] Bill Fefferman, Ronen Shaltiel, Christopher Umans, Emanuele Viola: On Beating the Hybrid Argument. Theory of Computing 9: 809-843 (2013)

[Gro96] Lov K. Grover: A Fast Quantum Mechanical Algorithm for Database Search. STOC 1996: 212-219

[Hås14] Johan Håstad: On the Correlation of Parity and Small-Depth Circuits. SIAM J. Comput. 43(5): 1699-1708 (2014)

[Iss1918] Leon Isserlis: On a Formula for the Product-Moment Coefficient of any Order of a Normal Frequency Distribution in any Number of Variables. Biometrika, 12(1): 134-139 (1918)

[LMN93] Nathan Linial, Yishay Mansour, Noam Nisan: Constant Depth Circuits, Fourier Transform, and Learnability. J. ACM 40(3): 607-620 (1993)

[Rem16] Zachary Remscrim: The Hilbert Function, Algebraic Extractors, and Recursive Fourier Sampling. FOCS 2016: 197–208

[Sho97] Peter W. Shor: Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. SIAM J. Comput. 26(5): 1484-1509 (1997)

[Sim94] Daniel R. Simon: On the Power of Quantum Computation. SIAM J. Comput. 26(5): 1474-1483 (1997)

[Tal17] Avishay Tal: Tight Bounds on the Fourier Spectrum of AC0. Computational Complexity Conference 2017: 15:1-15:31

[Watrous00] John Watrous: Succinct quantum proofs for properties of finite groups. FOCS 2000: 537-546