Oracle Separation of BQP and the Polynomial Hierarchy

Halley Goldberg and Shi Hao Liu

December 6, 2019

1 Introduction and Motivation of the Subject

This survey addresses the relationship between BQP, the class of problems efficiently solvable by quantum computers, and the polynomial hierarchy ('PH'), a classical generalization of P and NP. In 2018, Ran Raz and Avishay Tal proved the existence of an oracle O relative to which BQP $\not\subseteq$ PH [6]. This was a major breakthrough, as the existence of such an oracle had been an open question since the early days of quantum computing. Their work builds upon prior work of Scott Aaronson, who first suggested a variant of O as a candidate for such a separation. We aim to give a readable survey of the work leading up to and concepts involved in Raz and Tal's result.

Before getting into the specifics of the proof, we ought to explain why the question it answers is interesting and worthwhile to begin with. Some may feel that the relationship between BQP and PH relative to an oracle doesn't tell us much about their relationship in the actual, unrelativized world. It is well-known that oracle results can seem contradictory; for example, there exist oracles relative to which P $\neq$ NP as well as oracles relative to which P = NP. Further, even if BQP $\not\subseteq$ PH is true in the unrelativized sense, we are unlikely to have a proof of it anytime soon, for if we prove that, we have also proved that P $\neq$ PSPACE.

Aaronson explains why oracle-relative results of this kind are worth pursuing nonetheless [1]. First of all, he argues that the so-called 'query complexity' model is well-motivated in its own right, as it does represesnt a legitimate sense in which one kind of computation can have capabilities surpassing those of another. Knowledge as to what classical resources are needed to simulate quantum computation is interesting in and of itself. On a more pragmatic note, results in query complexity often serve as stepping-stones to more fundamental developments in complexity theory, and this is especially true in the history of quantum complexity theory.

So what is the significance of Raz and Tal's result in particular? For one, it is consistent with quantum computers exceeding classical computers, in terms of what they can efficiently compute, even in the case that P = NP. In other words, it is evidence that quantum computing could 'survive a collapse of PH'. Another upshot of the result is that it could open up a new place to look for quantum algorithms. This holds even if NP-complete problems lie outside of BQP. If BQP is not contained in PH, then we certainly need not limit ourselves to NP-intermediate problems, since quantum computation might be suitable for different kinds of problems lying outside of PH altogether [1].

2 Definitions and Results

We begin by recalling the main complexity classes involved in Raz and Tal's paper: PH, BQP, and AC^0. These definitions are standard and may be found in any textbook on complexity theory.

PH: The polynomial hierarchy, or PH, is a generalization of P and NP. Formally, for $i \geq 1$, a language $\mathcal{L}$ is in Σ_i^p iff there exists a polytime TM M and a polynomial q such that for all x,

$$x \in \mathcal{L} \iff \exists u_1 \in \{0,1\}^{q(|x|)} \forall u_2 \in \{0,1\}^{q(|x|)} ... Q_i u_i \in \{0,1\}^{q(|x|)} M(x, u_1, ..., u_i) = 1,$$

where Q_i denotes $\forall$ or $\exists$ depending on whether i is even or odd.

Then $\mathrm{PH} = \cup_i \Sigma_i^p$.

BQP: Bounded-error quantum polynomial time, or BQP, is the set of problems solvable by a quantum computer in polynomial time. Formally, a language $\mathcal{L}$ is in BQP iff there exists a polynomial-time uniform family of quantum circuits $\{Q_n \mid n \in \mathbb{N}\}$ such that:

1. for all $n \in \mathbb{N}$, Q_n takes n input qubits and outputs 1 bit
2. for all $x \in \mathcal{L}$, $\Pr(Q_{|x|} \text{ accepts } x) \geq \frac{2}{3}$
3. for all $x \notin \mathcal{L}$, $\Pr(Q_{|x|} \text{ rejects } x) \geq \frac{2}{3}$

$\mathbf{AC}^0$: A language $\mathcal{L}$ is in AC^0 iff it can be decided by a family of Boolean circuits $\{C_n \mid n \in \mathbb{N}\}$, where each C_n has $\mathrm{poly}(n)$ size and constant depth. Moreover, each C_n consists of unbounded fan-in AND and OR gates and NOT gates occurring only at the leaves.

Now, let U_N be the uniform distribution over $\{\pm 1\}^N$, and let $\mathcal{D}$ be some other distribution over $\{\pm 1\}^N$. We say that an algorithm A distinguishes between U_N and $\mathcal{D}$ with advantage ε iff

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

That is, the difference between the probabilities of A accepting a sample drawn from U_N and a sample drawn from D respectively is equal to ε.

Finally, we mention that much of the paper takes place in the "black box" model, also known as the "query complexity" model. In this model, the input $x \in \{\pm 1\}^N$ is accessed via queries to a black box. In particular, classical algorithms are allowed to apply the mapping $i \to x_i$ for unit cost, where x_i is the i^{th} bit of x. Quantum algorithms are allowed to apply the unitary transformation $|i, w\rangle \to x_i |i, w\rangle$ for unit cost; that is, the bit x_i is encoded in a quantum state as its phase.

The main results of Raz and Tal's paper are as follows:

Theorem 1.1. *There exists a distribution $\mathcal{D}$ over $\{\pm 1\}^{2N}$ such that: (i) there exists a quantum algorithm that makes 1 query to the input and distinguishes between U_{2N} and $\mathcal{D}$ with advantage $\Omega(1/\log N)$, and (ii) no $\mathbf{AC^0}$ circuit distinguishes between U_{2N} and $\mathcal{D}$ with advantage better than $\mathrm{polylog}(N)/\sqrt{N}$.*

Theorem 1.2. *There exists a distribution $\mathcal{D}$ over $\{\pm 1\}^N$ such that: (i) there exists a quantum algorithm that makes polylog(N) queries to the input and distinguishes between U_{2N} and $\mathcal{D}$ with advantage $1 - 2^{-polylog(N)}$, and (ii) no* $\mathbf{AC^0}$ *circuit distinguishes between U_{2N} and $\mathcal{D}$ with advantage better than* $\text{polylog}(N)/\sqrt{N}$.

Corollary 1.5. *There exists an oracle O such that* $\mathbf{BQP}^O \not\subseteq \mathbf{PH}^O$.

Informally, Theorems 1.1 and 1.2 state that there exists a distribution $\mathcal{D}$ that is easy for a quantum algorithm to distinguish from uniform, but hard for every classical circuit to distinguish. Theorem 1.2 is obtained from Theorem 1.1 through standard amplification techniques, which we do not cover here. See [6] for a proof. Corollary 1.5 is obtained from Theorem 1.2 as explained in the next section. After that section, we show how Raz and Tal proved Theorem 1.1.

3 Reframing the Problem

Claim . *Theorem 1.2 implies Corollary 1.5.*

We began this survey by characterizing our problem as an oracle separation problem: can BQP machines with access to a certain oracle solve problems that PH machines with access to that same oracle cannot solve? In this section, we sketch some of the concepts involved in reframing this question in terms of AC^0 circuits, as in Theorems 1.1 and 1.2. More formal proofs of the claim can be found in [4] as well as the appendix to [6].

One key idea was noticed in the early 1980s by Furst, Saxe, and Sipser, among others [5]. Let $N = 2^n$, let $x_n \in \{\pm 1\}^{2N}$, and suppose we want to compute some Boolean function $f : \{\pm 1\}^{2N} \to \{\pm 1\}$.

If there exists a $\mathbf{PH}$ *oracle machine M taking input 1^n and computing $f(x_n)$, then there exists an* $\mathbf{AC^0}$ *circuit taking input x_n and computing $f(x_n)$.*

The high-level reasoning behind this statement is as follows. Given fixed M, n, and f, our goal is to construct a Boolean circuit computing $f(x_n)$. We first convert M to another PH machine M_1 that only makes one oracle query, which can be done without loss of generality. Then, we build up a Boolean circuit corresponding to the computational branches of M_1. That is, we convert the alternating universal and existential quantifiers into alternating layers of AND and OR gates respectively, and we take the oracle's responses to all possible queries as input to the circuit at its leaves [5]. Note that since the length of the input x_n is exponential in n, M can only access indices of x_n via oracle queries. The size of the associated circuit is quasipolynomial in N.

How does the above equivalence apply to Raz and Tal's result? Well, for each $n \in \mathbb{N}$, draw x_n from U_{2N} with probability $1/2$, and draw x_n from $\mathcal{D}$ with probability $1/2$. Now define the language

$$\mathcal{L} = \{1^n \mid x_n \text{ was drawn from the distribution } \mathcal{D}\}$$

On one hand, Theorem 1.2 (i) implies that there exists a BQP oracle machine deciding $\mathcal{L}$ correctly on all but finitely many inputs. This machine essentially acts according to the quantum algorithm given in section 6 below [6]. Theorem 1.2 (ii) states, on the other hand, that no AC^0 circuit can distinguish well between U_{2N} and $\mathcal{D}$. By the equivalence of PH^O and AC^0, the probability that a fixed PH oracle machine decides $\mathcal{L}$ correctly on input 1^n must be low. The last remaining step is to show that the probability over oracles O

that a PH^O machine decides $\mathcal{L}$ correctly *for all* n is equal to 0. We omit this analysis, which can be found in the appendix to [6].

4 The Distributions

Intuitively, the goal is to come up with a distribution $\mathcal{D}$ such that $\mathcal{D}$ and U_{2N} can be told apart by a quantum algorithm but not by classical Boolean circuits. Aaronson was the first to suggest the appropriate distribution [1], though Raz and Tal alter it in some ways to suit their subsequent analysis.

Let $N = 2^n$ for $n \in \mathbb{N}$, and let $\varepsilon = 1/(24 \ln N)$. We start by defining distributions $\mathcal{G}$ over $\mathbb{R}^{2N}$:

1. Independently sample $x_1, ..., x_N \sim \mathcal{N}(0, 1)$ (ie. the normal distribution with mean 0 and variance 1) and let $x = x_1, ..., x_N$.
2. Let $y = H_N \cdot x$, where H_N is the Hadamard transform.
3. Output $z = (x, y)$.

From $\mathcal{G}$, define the distribution $\mathcal{G}'$ over $\mathbb{R}^{2N}$ as follow:

1. Sample $z \sim \mathcal{G}$.
2. Output $\sqrt{\varepsilon} z$.

Observe that $\mathcal{G}'$ can be equivalently defined by following the definition of $\mathcal{G}$ but sample $x_1, ..., x_N$ from $\mathcal{N}(0, \varepsilon)$ instead.

From $\mathcal{G}'$, we define another distribution $\mathcal{D}$:

1. Sample z from $\mathcal{G}'$.
2. Replace each z_i with $\mathrm{trnc}(z_i)$, where trnc is a function that truncates its input to the interval $[-1, +1]$, i.e. $\mathrm{trnc}(z_i) = \min(1, \max(-1, z_i))$.
3. Set $z'_i = 1$ with probability $\frac{1+\mathrm{trnc}(z_i)}{2}$ and $z'_i = -1$ with probability $\frac{1-\mathrm{trnc}(z_i)}{2}$.
4. Output $z' \in \{\pm 1\}^{2N}$.

Note that in Aaronson's original distribution [1], the main differences from $\mathcal{D}$ were that $\varepsilon = 1$ and each z'_i was taken to be $\mathrm{sgn}(z_i) \in \{\pm 1\}$, where sgn denotes the sign function. He called it a 'forrelated' distribution (ie. 'Fourier' + 'correlated'), since y is tightly correlated with the Fourier distribution of x, though x and y individually are both uniformly random.

5 Multilinear Functions

In this section we describe some properties about multilinear function. Along the way we will see how these type of functions correlates to $\mathbf{AC^0}$ circuits.

It is well-known that every $\mathbf{AC^0}$ circuit can be well-approximated by multilinear (low-degree) polynomials over $\mathbb{R}$. For every $A : \{-1,1\}^m \to \{-1,1\}$, there is an unique multilinear real polynomial

$$\tilde{A}(x) = \sum_{S \subseteq [m]} \hat{A}(S) \prod_{i \in S} x_i$$

called the *multilinear extension of* A, that agrees with A over $\{-1,1\}^m$. The terms $\hat{A}(S)$ are called the *Fourier coefficients of* f. As a side note, truncating $\tilde{A}$ to its degree-k parts for some small k gives us a low-degree approximation of A. To analyze $\mathbf{AC^0}$ circuits, it is therefore sufficient for us to analyze the behaviours of their multilinear extensions. For any Boolean circuit $A : \{\pm 1\}^{2N} \to \{\pm 1\}$ of quasi-polynomial size and constant depth, we will ambiguously denote by A both the circuit itself and its multilinear extension.

Observe that, by the linearity of expectation, $A(\vec{0}) = \hat{A}(\emptyset) = \mathbb{E}_{u \sim U_{2N}}[A(u)]$. To show that A cannot distinguish between $\mathcal{D}$ and the uniform distribution U_{2N}, we need to prove that

$$\mathbb{E}_{z \sim D}[A(z)] \approx \mathbb{E}_{u \sim U_{2N}}[A(u)] = A(\vec{0})$$

However, it is enough for us to show that $\left|\mathbb{E}_{z \sim \mathcal{G}'}[A(z)] - A(\vec{0})\right|$ is small, due to the following claim.

Claim. *Any multilinear function $f : [-1,1]^{2N} \to [-1,1]$ has very similar expectation under $\mathcal{G}'$ and under $\mathcal{D}$.*

In fact, any multilinear function has the *same* expectation under $\mathcal{D}$ and the truncated variant of $\mathcal{G}'$. That is,

$$\mathbb{E}_{z' \sim \mathcal{D}}[A(z')] = \mathbb{E}_{z \sim \mathcal{G}'}[A(\mathrm{trnc}(z))]$$

Therefore, the above claim suggests that the quantity $|\mathbb{E}_{z \sim \mathcal{G}'}[A(z) - A(\mathrm{trnc}(z))]|$ is small. To state this more precisely, consider the following more generalized claim.

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

$$\mathbb{E}_{z \sim \mathcal{G}'}\left[|A(\mathrm{trnc}(z_0 + pz)) - A(z_0 + pz)|\right] \leq 8N^{-2}$$

This allows the subsequent analysis to essentially ignore the fact that we truncated to the interval $[-1,1]$ in the definition of D. We omit the proof of Claim 5.3, for it is long and rather straightforward. Interested readers may refer to Section 5 of [6].

6 The Quantum Algorithm

Claim . *There exists a quantum algorithm making 1 query and running in time $O(\log N)$ that distinguishes $\mathcal{D}$ from U_{2N} with advantage $\Omega(1/\log N)$.*

Informally, we would like to show that distinguishing U_{2N} and $\mathcal{D}$ is easy for quantum circuits. This is part (i) of Theorem 1.1. Raz and Tal did not give a quantum algorithm explicitly in their paper, but they used properties of an algorithm first developed by Aaronson in [1] and subsequently improved by Aaronson and Ambainis in [2]. The algorithm begins in the state $|0\rangle^{\otimes n}|0\rangle$ and proceeds as follows:

1) Apply the Hadamard gate H to the last (control) qubit.

$|0\rangle^{\otimes n}|+\rangle = \frac{1}{\sqrt{2}}|0\rangle^{\otimes n}|0\rangle + \frac{1}{\sqrt{2}}|0\rangle^{\otimes n}|1\rangle$

2) Apply H to the first n qubits.

$$\frac{1}{\sqrt{2N}}\sum_{i\in[N]}|i\rangle|0\rangle + \frac{1}{\sqrt{2N}}\sum_{i\in[N]}|i\rangle|1\rangle$$

3) Query x in superposition conditioned on the control qubit being $|0\rangle$, and query y conditioned on the control qubit being $|1\rangle$.

$$\frac{1}{\sqrt{2N}}\sum_{i\in[N]}x_i|i\rangle|0\rangle + \frac{1}{\sqrt{2N}}\sum_{i\in[N]}y_i|i\rangle|1\rangle$$

4) Apply H to the first n qubits conditioned on the control qubit being $|1\rangle$.

$$\frac{1}{\sqrt{2N}}\sum_{i\in[N]}x_i|i\rangle|0\rangle + \frac{1}{\sqrt{2N}}\sum_{i\in[N]}\sum_{j\in[N]}y_jH_{ij}|i\rangle|1\rangle$$

Finally, measure this state in the $\{|+\rangle, |-\rangle\}$ basis, accepting iff the outcome is $|+\rangle$ [2].

$$\begin{aligned}\Pr[\text{outcome }|+\rangle] &= \sum_{i\in[N]}\left(\frac{1}{2\sqrt{N}}x_i + \frac{1}{2\sqrt{N}}\sum_{j\in[N]}(y_jH_{ij}\right)^2 \\ &= \frac{1}{4N}\sum_{i\in[N]}\left(x_i^2 + 2\sum_{j\in[N]}H_{ij}x_iy_j + \left(\sum_{j\in[N]}y_jH_{ij}\right)^2\right) \\ &= \frac{1}{2} + \frac{1}{2N}\sum_{i\in[N]}\sum_{j\in[N]}H_{ij}x_iy_j\end{aligned}$$

Define $\varphi(xy) := \frac{1}{N}\sum_{i\in[N]}\sum_{j\in[N]}H_{ij}x_iy_j$, so $\Pr[\text{outcome}\,|+\rangle] = \frac{1+\varphi(x,y)}{2}$. Note that φ is a multilinear function.

Now that we have defined the quantum algorithm, we need to show that it successfully distinguishes between (x, y) drawn from U_{2N} and (x, y) drawn from $\mathcal{D}$, as stated in the claim above.

First of all, it is clear that $E_{(x,y)\sim U_{2N}}[\Phi(x, y)] = 0$, by linearity of expectation and the fact that $E[x_iy_j] = 0$ in the uniform distribution. So when (x, y) is drawn from the uniform distribution, the quantum algorithm returns each possible output with probability $\frac{1}{2}$.

It remains to show that when (x, y) is drawn from the distribution $\mathcal{D}$, we have $E[\varphi(x, y)] \in \Omega(1/\log N)$. Raz

and Tal begin by showing this in the case where (x, y) is drawn from $\mathcal{G}'$, as $\mathcal{G}'$ is a much 'nicer' distribution to analyze than $\mathcal{D}$. They then use the definitions of $\mathcal{G}'$ and $\mathcal{D}$, as well as some properties of multilinear functions (see the previous section), to give a lower bound in the case of $\mathcal{D}$.

In particular, since $E_{(x,y)\sim\mathcal{G}'}[x_i y_j] = \varepsilon \cdot H_{ij}$, it is straightforward to show that $E_{(x,y)\sim\mathcal{G}'}[\varphi(x, y)] = \varepsilon$. Then:

$E_{(x',y')\sim\mathcal{D}}[\varphi(x', y')] = E_{(x,y)\sim\mathcal{G}'}[\varphi(\text{trnc}(x), \text{trnc}(y))]$ (by multilinearity of φ)

$\geq E_{(x',y')\sim\mathcal{D}}[\varphi(x', y')] - |E_{(x',y')\sim\mathcal{D}}[\varphi(\text{trnc}(x), \text{trnc}(y)) - \varphi(x', y')]|$ (by definition of $\mathcal{G}', \mathcal{D}$)

$\geq \varepsilon/2 \in \Omega(1/\log N)$ (by multilinearity of φ)

As a corollary, we get the claim that opens this section [6].

7 The Circuit Lower Bound

We continue our discussion on $\mathbf{AC^0}$ circuits. In this section, we wish to establish the following:

Theorem 7.4. *Let $A : \{\pm 1\}^{2N} \to \{\pm 1\}$ be a Boolean circuit of size s and depth d. Then,*

$$\left|\mathbb{E}_{z'\sim\mathcal{D}}[A(z')] - A(\vec{0})\right| \leq 32\varepsilon(c \log s)^{2(d-1)} N^{-1/2}$$

As an immediate consequence, we get the following corollary, which is the wanted result:

Corollary 7.5. *Let $A : \{\pm 1\}^{2N} \to \{\pm 1\}$ be a Boolean circuit of size $\exp\left(\log^{O(1)}(N)\right)$ and depth $O(1)$. Then,*

$$|\mathbb{E}_{z'\sim\mathcal{D}}[A(z')] - \mathbb{E}_{u\sim U_{2N}}[A(u)]| \leq \text{polylog}(N)/\sqrt{N}$$

Before we move on, let us note some property of $\hat{\mathcal{G}}(S, T) = \mathbb{E}_{(x,y)\sim(G)}\left[\left(\prod_{i\in S} x_i\right) \cdot \left(\prod_{j\in T} y_j\right)\right]$, the moments of $\mathcal{G}$.

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

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

The first item is an entry in the covariance matrix of $\mathcal{G}$, which is $\begin{pmatrix} I_N & H_N \\ H_N & I_N \end{pmatrix}$. The second item follows from Isserlis' Theorem, stating that $\mathbb{E}[z_{i_1} ... z_{i_{2k-1}}] = 0$ and $\mathbb{E}[z_{i_1} ... z_{i_{2k}}] = \sum\prod \mathbb{E}\left[z_{i_r} z_{i_\ell}\right]$ for any $k \leq N$ and distinct $i_1, ..., i_{2k} in [2N]$, where $\sum\prod$ means summing over all distinct ways of partitioning $z_{i_1}, ..., z_{i_{2k}}$ into pairs and each summand is the product of the k pairs. The third item follows from the fact that there are

$k!$ partitions of elements of S and T into pairs such that each pair contains exactly one variable from each half, and the fact that the covariant of each such pair is $\pm N^{-1/2}$. The last item can be easily proven using Cauchy-Schwarz inequality.

To get a picture of how the analysis would go, let us unpack the quantity $\mathbb{E}_{z' \sim \mathcal{D}}[A(z')] - A(\vec{0})$.

$$\begin{aligned}
\mathbb{E}_{z \sim \mathcal{G}'}[A(z)] - \mathbb{E}_{u \sim U_{2N}}[A(u)] &= \sum_{S \subseteq [2N]} \hat{A}(S) \left(\mathbb{E}_{z \sim \mathcal{G}'} \left[\prod_{i \in S} z_i \right] - \mathbb{E}_{u \sim U_{2N}} \left[\prod_{i \in S} u_i \right] \right) \\
&= \sum_{S \subseteq [2N], |S| \geq 1} \hat{A}(S) \cdot \mathbb{E}_{z \sim \mathcal{G}'} \left[\prod_{i \in S} z_i \right] \\
&= \sum_{\ell=1}^{N} \sum_{S \subseteq [2N], |S| = 2\ell} \hat{A}(S) \cdot \mathbb{E}_{z \sim \mathcal{G}'} \left[\prod_{i \in S} z_i \right] \qquad (1) \\
&= \sum_{\ell=1}^{N} \sum_{S \subseteq [2N], |S| = 2\ell} \hat{A}(S) \cdot \varepsilon^{\ell} \cdot \hat{\mathcal{G}}(S) \qquad (2)
\end{aligned}$$

Equation (1) follows from the second item in Claim 4.1. Since the third and fourth items in Claim 4.1 give us a bound on $\mathbb{E}_{z \sim G} \left[\prod_{i \in S} z_i \right]$, to obtain the result, it remains for us to get a bound on the Fourier coefficients $\hat{A}(S)$.

7.1 Tal's Tail Bounds on Fourier Coefficients

The analysis made extensive use of the tail bounds on the Fourier coefficients of multilinear extensions of any quasi-polynomial size and constant depth circuit.

Lemma 7.1. *There exists a universal constant $c > 0$ such that the following holds. Let $A : \{\pm 1\}^{2N} \to \{\pm 1\}$ be a Boolean circuit with at most s gates and depth at most d. Then, for all $k \in \mathbb{N}$,*

$$\sum_{S \subseteq [2N] : |S| = k} \left| \hat{A}(S) \right| \leq (c \log s)^{(d-1)k}$$

The quantity translates to $(\text{polylog}(N))^k$ when s is quasi-polynomial in N and d is constant. This bound was proven by Tal in 2014 in [7] and it was primarily motivated by the proving Theorem 7.4. This is the first place where the Raz and Tal paper goes substantially beyond the work of Aaronson and others. It is tempting to take the bounds in Lemma 7.1 and Claim 4.1 and plug them into formula (2) derived above, hoping to obtain the result directly. In fact, this is what they initially tried to do. Unfortunately, the bound fails to give us the desired result when we add in the terms corresponding to large k.

7.2 Random Walks

Instead, Raz and Tal took a different approach, inspired by the work [3] of Chattopadhyay, Hatami, Hosseini, and Lovett. We view $z \sim \mathcal{G}'$ as the result of a random walk $\frac{1}{\sqrt{t}} \cdot \left(z^{(1)} + ... + z^{(t)} \right)$, where $z^{(1)}, ..., z^{(t)} \sim \mathcal{G}'$ are t vectors sampled independently from $\mathcal{G}'$, for some t. One could check that the walk produces a distribution having the same mean and variance as $\mathcal{G}'$. The proof is then carried out using a hybrid argument. That is, we define the i-th hybrid as $H_t = \frac{1}{\sqrt{t}} \cdot \left(z^{(1)} + ... + z^{(i)} \right)$, and show that, for each $i = 0, ..., t-1$,

$$\left| \mathbb{E}\left[A(H_{i+1}) \right] - \mathbb{E}\left[A(H_i) \right] \right| \leq \frac{\delta}{t} \cdot \text{polylog}(N)$$

for some suitable δ which would give us the result.

7.3 Proving the Main Theorem

With the general ideas in mind, we finally proceed to the actual proof of the result. As described above, we analyze the behaviour of A under each consecutive pairs of hybrids. We begin with a lemma that would allow us to take care of the base case, where $i = 0$. For vectors $R, Q \in \mathbb{R}^{2N}$, define $R \circ Q \in \mathbb{R}^{2N}$ to be their point-wise product.

Claim 7.2. *Let $p \leq 1/2$. Let $A : \{\pm 1\}^{2N} \to \{\pm 1\}$ 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,*

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

Proof. Let $q = \sqrt{\varepsilon} \cdot p \cdot (c \log s)^{d-1} \leq 1/2$. Following a similar expansion as shown in the beginning of Section 7,

$$\begin{aligned}
\left|\mathbb{E}_{z\sim\mathcal{G}'}\left[A(P \circ z) - A(\vec{0})\right]\right| &= \left|\sum_{\emptyset\neq S\subseteq[2N]} \hat{A}(S) \cdot \left(\prod_{i\in S} P_i\right) \cdot \hat{\mathcal{G}'}(S)\right| \\
&\leq \sum_{\emptyset\neq S\subseteq[2N]} \left|\hat{A}(S)\right| \cdot p^{|S|} \cdot \sqrt{\varepsilon}^{|S|} \cdot \left|\hat{\mathcal{G}}(S)\right| \\
&= \sum_{k=1}^{2N} \left(\left(\sqrt{\varepsilon}p\right)^k \cdot \sum_{S\subseteq[2N],|S|=k} \left(\left|\hat{A}(S)\right| \cdot \left|\hat{\mathcal{G}}(S)\right|\right)\right) \\
&\leq \sum_{k=1}^{2N} \left(\left(\sqrt{\varepsilon}p\right)^k \cdot \left(\max_{S:|S|=k} \left|\hat{\mathcal{G}}(S)\right|\right) \cdot \sum_{S\subseteq[2N],|S|=k} \left(\left|\hat{A}(S)\right|\right)\right) \\
&\leq \sum_{k=1}^{2N} \left(\left(\sqrt{\varepsilon}p\right)^k \cdot \left(\max_{S:|S|=k} \left|\hat{\mathcal{G}}(S)\right|\right) \cdot (c\log s)^{(d-1)k}\right) && \text{(by Lemma 7.1)} \\
&= \sum_{k'=1}^{N} \left(q^{2k'} \cdot \left(\max_{S:|S|=2k'} \left|\hat{\mathcal{G}}(S)\right|\right)\right) && \text{(by Claim 4.1(2.))} \\
&= \sum_{k'=1}^{\lfloor n/2\rfloor} \left(q^{2k'} \cdot \left(\max_{S:|S|=2k'} \left|\hat{\mathcal{G}}(S)\right|\right)\right) + \sum_{k'=\lfloor n/2\rfloor+1}^{N} \left(q^{2k'} \cdot \left(\max_{S:|S|=2k'} \left|\hat{\mathcal{G}}(S)\right|\right)\right) \\
&\leq \sum_{k'=1}^{\lfloor n/2\rfloor} \left(q^{2k'} \cdot k'! \cdot N^{-k'/2}\right) + \sum_{k'=\lfloor n/2\rfloor+1} \left(q^{2k'}\right) && \text{(by Claim 4.1(3. \& 4.))} \\
&\leq 2 \cdot q^2 \cdot N^{-1/2} + 2 \cdot q^{n+1} && \text{(since } q \leq 1/2\text{)} \\
&\leq 3q^2 \cdot N^{-1/2}
\end{aligned}$$

□

The next claim takes care of the case for $1 \leq 1 \leq t-1$. Its proof makes use of the fact that $\mathbf{AC^0}$ circuits are closed under restrictions.

Claim 7.3. *Let $p \leq 1/4$. Let $A : \{\pm 1\}^{2N} \to \{\pm 1\}$ 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,*

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

Proof. The proof is based on (the brilliant technique of) random restriction. Intuitively, we want to reduce this case to the case where $i = 0$, and apply Claim 7.2. We start by defining a distribution, $\mathcal{R}_{z_0}$, for the random restriction, which depends (only) on z_0. A restriction ρ is sampled from $\mathcal{R}_{z_0}$ as follow: for each entry $i \in [2N]$,

$$\rho_i = \begin{cases} \mathrm{sng}((z_0)_i) & \text{with probability } |(z_0)_i| \\ * & \text{otherwise} \end{cases}$$

Next, we will define a (random) vector $\tilde{z}$ from ρ for which we will use to compare with $\vec{0}$. Though this is not quite right – in truth, we will compare $\tilde{z}$ with $\vec{0}\restriction_\rho$ (that is, we set each coordinate i of $\vec{0}$ on which $\rho_i \neq *$ to be ρ_i instead). We can think of this as shrinking the dimension of the hypercube and shifting the "center" to become $\vec{0}\restriction_\rho$, thus allowing us to reduce the problem to Claim 7.2. We now make this precise.

Let $\rho \sim \mathcal{R}_{z_0}$. For any $z \in \mathbb{R}^{2N}$, define the vector $\tilde{z} = \tilde{z}(z, \rho) \in \mathbb{R}^{2N}$ to be

$$\tilde{z}_i = \begin{cases} \rho_i & \text{if } \rho_i \in \{\pm 1\} \\ P_i \cdot z_i & \text{otherwise} \end{cases}$$

where $P \in [-2p, 2p]^{2N}$ is defined by $P_i = p \cdot \frac{1}{1 - |(z_0)_i|}$. Observe that each coordinate $\tilde{z}_i$ is independent of the other coordinates, and its expected value is

$$\begin{aligned} \mathbb{E}_{\rho \sim \mathcal{R}_{z_0}} [\tilde{z}_i] &= |(z_0)_i| \cdot \mathrm{sgn}((z_0)_i) + (1 - |(z_0)_i|) \cdot P_i \cdot z_i \\ &= (z_0)_i + p \cdot z_i \end{aligned}$$

Hence,

$$\begin{aligned} \mathbb{E}_{\rho \sim \mathcal{R}_{z_0}} [A(\tilde{z})] &= \sum_{S \subseteq [2N]} \hat{A}(S) \cdot \mathbb{E}_{\rho \sim \mathcal{R}_{z_0}} \left[\prod_{i \in S} \tilde{z}_i \right] \\ &= \sum_{S \subseteq [2N]} \hat{A}(S) \cdot \prod_{i \in S} \mathbb{E}_{\rho \sim \mathcal{R}_{z_0}} [(\tilde{z}_i)] && \text{(by independence)} \\ &= A(z_0 + p \cdot z) \end{aligned}$$

Let $z \sim \mathcal{G}'$, then

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

Notice that $A\restriction_\rho$ is again a multilinear function and how its $A\restriction_\rho(\vec{0})$ value, i.e. its "center" value, has shifted because of the restriction ρ. This allows us to apply Claim 7.2, replacing A with $A\restriction_\rho$, and using the fact that $P \in [-2p, 2p]^{2N}$ and the assumption $\sqrt{\varepsilon} p (c \cdot \log s)^{d-1} \leq 1/4$, we get,

$$\begin{aligned} &\leq \mathbb{E}_{\rho \sim \mathcal{R}_{z_0}} \left[3\varepsilon (2p)^2 \cdot (c \log s)^{2(d-1)} \cdot N^{-1/2} \right] \\ &\leq 3\varepsilon (2p)^2 \cdot (c \log s)^{2(d-1)} \cdot N^{-1/2} && \text{(by averaging)} \end{aligned}$$

□

It is worthwhile to re-read the proof of Claim 7.3 to understand what is really going on and the point of using the random restriction.

We conclude this section with the proof of Theorem 7.4, which sums up the results we have discussed so far and completes the puzzle.

Proof (Theorem 7.4). The proof made use of Claim 7.3 about the random walk, Claim 5.3 about $\mathcal{G}'$ versus its truncated version, and the fact that the walk at each step remains within the subcube $[-1/2, 1/2]^{2N}$ with high probability (this allows us to apply Claim 7.3!).

Let $t = N$ be the number of steps in the walk and let $p = 1/\sqrt{t}$. Without loss of generality, assume that $\sqrt{\varepsilon} \cdot (c \cdot \log s)^{d-1} \leq \frac{1}{4} \cdot N^{1/4}$ (for otherwise the inequality in the Theorem holds trivially), then

$$p \cdot \left(\sqrt{\varepsilon} \cdot (c \cdot \log s)^{d-1}\right) \leq \frac{1}{4} \cdot N^{-1/4} \leq 1/4$$

For $0 \leq i \leq t$, let H_i be as defined in Section 7.2. Note that $\frac{H_i}{p\sqrt{i}} \sim \mathcal{G}'$, and for each $j \in \{1, ..., 2N\}$, $(H_i)_j \sim \mathcal{N}(0, p^2 i\varepsilon)$. Let E_i be the event that $H_i \in [-1/2, 1/2]^{2N}$. Using the bound for Gaussian distribution,

$$\begin{aligned}
\mathbb{P}[E_i] &\geq 1 - \sum_j \mathbb{P}[|H_i| \geq 1/2] && \text{(by union bound)} \\
&\geq 1 - \sum_j \mathbb{P}[|\mathcal{N}(0, \varepsilon)| \geq 1/2] \\
&\geq 1 - 2N \cdot e^{-1/(8\varepsilon)} && \text{(by bounds on Gaussian)} \\
&\geq 1 - 2N^{-2}
\end{aligned}$$

Conditioned on E_i, we analyze the difference $|\mathbb{E}[A(\text{trnc}(H_{i+1})) \mid E_i] - \mathbb{E}[A(\text{trnc}(H_i)) \mid E_i]|$.

$$\begin{aligned}
\left|\mathbb{E}[A(\text{trnc}(H_{i+1})) \mid E_i] - \mathbb{E}[A(\text{trnc}(H_i)) \mid E_i]\right| &\leq \left|\mathbb{E}[A(\text{trnc}(H_{i+1})) \mid E_i] - \mathbb{E}[A(H_{i+1}) \mid E_i]\right| \\
&\quad + \left|\mathbb{E}[A(H_{i+1}) \mid E_i] - \mathbb{E}[A(\text{trnc}(H_i)) \mid E_i]\right| \\
&\leq 8 \cdot N^{-2} + 12\varepsilon \cdot p^2 \cdot (c \cdot \log s)^{2(d-1)} \cdot N^{-1/2}
\end{aligned}$$

where the last inequality is obtained from applying Claim 5.3 to the first term (with $z_0 = H_i$, $p_0 = 1/2$ and $p = 1\sqrt{t} < 1/2$, using the fact that $H_i \in [-p_0, p_0]^{2N}$) and applying Claim 7.3 to the second term. Together, since $A(\text{trnc}(H_{i+1})) + A(\text{trnc}(H_i)) \leq 2$,

$$\begin{aligned}
|\mathbb{E}[A(\text{trnc}(H_{i+1}))] - \mathbb{E}[A(\text{trnc}(H_i))]| &\leq 8 \cdot N^{-2} + 12\varepsilon \cdot p^2 \cdot (c \cdot \log s)^{2(d-1)} \cdot N^{-1/2} + 2 \cdot \Pr\left[\bar{E}\right] \\
&\leq 12 \cdot N^{-2} + 12\varepsilon \cdot p^2 \cdot (c \cdot \log s)^{2(d-1)} \cdot N^{-1/2}
\end{aligned}$$

Finally,

$$\begin{aligned}
\left|\mathbb{E}_{z' \sim \mathcal{D}}[A(z')] - A(\vec{0})\right| &\leq t\left(12 \cdot N^{-2} + 12\varepsilon \cdot p^2 \cdot (c \cdot \log s)^{2(d-1)} \cdot N^{-1/2}\right) \\
&= 12 \cdot N^{-1} + 12\varepsilon \cdot (c \cdot \log s)^{2(d-1)} \cdot N^{-1/2} \\
&\leq 32\varepsilon \cdot (c \cdot \log s)^{2(d-1)} \cdot N^{-1/2}
\end{aligned}$$

□

8 Open Questions

In this section, we indicate some questions left open by Raz and Tal's result.

1. Can we expand or generalize the class of problems that suffice for the separation? For example, does Aaronson's original distribution work? Raz and Tal give one instance of an oracle problem that can be done in BQP and not in PH, but it would be interesting to see if there are more problems like this and whether these problems have certain properties in common.

2. Can we adapt the methods to bear on the containment of NP in BQP? Whether this containment would imply the collapse of PH is currently an open question, and many who believe the hierarchy infinite would take such an implication as good evidence that NP $\not\subseteq$ BQP. One target, suggested by Scott Aaronson, would be to come up with an oracle relative to which NP $\subseteq$ BQP but PH $\not\subseteq$ BQP. This would be an enlightening result, because it would tell us that any proof of the above implication would necessarily be non-relativizing [1].

3. Can we come up with a relativized world in which P = NP but P $\neq$ BQP? In other words, can the notion that BQP survives a collapse of PH be demonstrated explicitly? This question was raised by Lance Fortnow in a blog post following Raz and Tal's result [8].

4. Is there a quantum analogue to $\mathrm{NP}^{\mathrm{BPP}} \subseteq \mathrm{BPP}^{\mathrm{NP}}$? Namely, is it true that $\mathrm{NP}^{\mathrm{BQP}} \subseteq \mathrm{BQP}^{\mathrm{NP}}$? [1]

5. The proof of this result relies on two crucial properties of $\mathbf{AC^0}$ circuits, namely the tail bound on Fourier coefficients of their multilinear extensions (specifically, for sets of size 2) and the fact that they are closed under restrictions. They argued that the result can be generalized to any class of functions satisfying these two properties. One plausible such candidate is $\mathbf{AC^0}[\otimes]$, which stands for $\mathbf{AC^0}$ circuits augmented with parity gates. It suffices to show that for all f in $\mathbf{AC^0}[\otimes]$, $\sum_{|S|=2} |\hat{f}(S)| \leq \frac{\sqrt{N}}{\mathrm{polylog}(N)}$ in order to separate **BQLogTime** from $\mathbf{AC^0}[\otimes]$.

9 References

[1] S. Aaronson, "BQP and the polynomial hierarchy," *Proceedings of the 42nd ACM symposium on Theory of computing - STOC 10*, 2010.

[2] S. Aaronson and A. Ambainis, "Forrelation: A Problem that Optimally Separates Quantum from Classical Computing," *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing - STOC 15*, 2015

[3] E. Chattopadhyay, P. Hatami, K. Hosseini, and S. Lovett, "Pseudorandom Generators from Polarizing Random Walks," *Electronic Colloquium on Computational Complexity, Report No. 15*, 2018.

[4] B. Fefferman and C. Umans, "Pseudorandom generators and the BQP vs. PH problem," 2010.

[5] M. Furst, J. B. Saxe, and M. Sipser, "Parity, circuits, and the polynomial-time hierarchy," *22nd Annual Symposium on Foundations of Computer Science*, 1981.

[6] R. Raz and A. Tal, "Oracle separation of BQP and PH," *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, 2018.

[7] A. Tal, " Tight Bounds on the Fourier Spectrum of AC0," *Electronic Colloquium on Computational Complexity, Report No. 174*, 2014.

[8] L. Fortnow, "BQP not in the Polynomial-Time Hierarchy in Relativized Worlds," *Computational Complexity* (Blog), 01-Jun-2018.