Oracle **BQP** vs. **PH**

Noah Singer, Lavanya Singh, Juspreet Singh Sandhu
CS 221 Final Project

May 9, 2021

1 Introduction

Given two complexity classes A and B, an *oracle separation of A from B* is a construction of an oracle $\mathcal{O}$ such that $A^{\mathcal{O}} \not\subseteq B^{\mathcal{O}}$, i.e., such that there is a problem "solvable by A-computations with an oracle for $\mathcal{O}$" but "not solvable by B-computations with an oracle for $\mathcal{O}$". Gill, Baker, and Solovay [BGS75] constructed an oracle relative to which $\mathbf{P} \subsetneq \mathbf{NP}$, and observed that if $\mathcal{O}$ is a **PSPACE**-complete function, we have $\mathbf{PSPACE} = \mathbf{P}^{\mathcal{O}} \subseteq \mathbf{NP}^{\mathcal{O}} \subseteq \mathbf{PSPACE}^{\mathcal{O}} = \mathbf{PSPACE}$. Oracle separations are both weak evidence for (unrelativized) separations and "lower bounds in a concrete computational model [(query complexity)] that is natural and well-motivated in [its] own right" [Aar10].

In the quantum world, when Bernstein and Vazirani defined the class **BQP**, they also constructed an oracle relative to which $\mathbf{BPP} \subsetneq \mathbf{BQP}$ [BV93]. Subsequently, Bennett, Bernstein, Brassard, and Vazirani [Ben+97] constructed an oracle relative to which $\mathbf{NP} \cap \mathbf{coNP} \not\subseteq \mathbf{BQP}$. Aaronson [Aar10] conjectured the existence of an oracle separating **BQP** from **PH**. Earlier, a series of works initiated by Furst, Saxe, and Sipser [FSS81; Yao85; Hås86] had constructed oracles relative to which $\mathbf{PSPACE} \not\subseteq \mathbf{PH}$ by "scaling down" and instead proving circuit lower bounds for the class $\mathbf{AC}^0$. Aaronson [Aar10] proposed to construct an oracle separating **BQP** from **PH** by "scaling down" and instead constructing pseudorandom distributions for $\mathbf{AC}^0$.

Raz and Tal [RT19] recently constructed such an oracle, and the goal of this paper is to outline the major ideas in this line of work. In §2, we describe the "scaling down" paradigm; it suffices to construct a distribution $\mathcal{D}$ which is "more pseudorandom" for $\mathbf{AC}^0$ circuits than for **BQLOGTIME** algorithms. Then, we present the necessary analysis in two steps. In §3, we present the distribution $\mathcal{D}$, which will be a truncated Gaussian and its Fourier transform, as well as its quantum distinguishing circuit. In §4, we present a simpler proof due to Wu [Wu20] of a lower bound on $\mathbf{AC}_0$ circuits distinguishing $\mathcal{D}$; the proof views $\mathcal{D}$ as the result of brownian motion and combines tools from stochastic calculus with bounds due to Tal [Tal17] on second-level fourier coefficients of $\mathbf{AC}^0$ circuits.

2 Lower bounds for $\mathbf{AC}^0$ and oracle separations of PH

The class **PH** is closely related to $\mathbf{AC}^0$ circuits; informally, the constant number of alternating $\exists$ and $\forall$ quantifiers in a **PH** computation graph correspond to a constant-depth circuit built from

$\vee$ and $\wedge$ gates, respectively. This relationship may be leveraged to translate hardness results for $\mathbf{AC}^0$ circuits into hardness results for **PH** oracle classes by encoding a length-2^k input for the $\mathbf{AC}^0$ task as a function from $\{0,1\}^k \to \{0,1\}$ computed by an oracle [FSS81]. The following lemma instantiates this paradigm for the **BQP** vs. **PH** problem; essentially, it "scales down" the **BQP** vs. **PH** problem by "taking the logarithm of both sides" to become **BQLOGTIME** vs. $\mathbf{AC}^0$. A proof can be found in [RT19, Appendix A], but the lemma is attributed to Aaronson [Aar10] and Fefferman, Shaltiel, Umans, and Viola [Fef+12].

Lemma 1 (Scaling down for **BQP** vs. **PH**)**.** *Suppose that (for every N) there exists a distribution $\mathcal{D}$ on $\{0,1\}^N$, such that:*

1. ***Classical lower bound:*** *For every function $f : \{0,1\}^N \to \{0,1\}$ computable by an $\mathbf{AC}_0$ circuit,*
$$|\mathbb{E}[f(\mathcal{D})] - \mathbb{E}[f(\mathcal{U}_N)]| \leq O\left(\frac{\text{polylog}(N)}{\sqrt{N}}\right).$$
2. ***Quantum upper bound:*** *There exists an efficient quantum algorithm Q that runs in time $O(\log(N))$ such that,*
$$|\mathbb{E}[Q(\mathcal{D})] - \mathbb{E}[Q(\mathcal{U}_N)]| \geq \Omega\left(\frac{1}{\log(N)}\right).$$

Then, there exists an oracle $\mathcal{O}$ such that $\mathbf{BQP}^{\mathcal{O}} \not\subseteq \mathbf{PH}^{\mathcal{O}}$.

In what follows, we define the distribution $\mathcal{D}$ in Definition 3, and prove the quantum upper and classical lower bounds in Theorem 1 and Theorem 3, respectively.

3 Quantum upper bound

In this section, we define the distribution $\mathcal{D}$ and prove the appropriate quantum upper bound (Theorem 1 below). We introduce the quantum circuit first to motivate the definition of $\mathcal{D}$.

3.1 Efficient quantum circuit

Aaronson [Aar10] showed that the 2-query quantum circuit in Figure 1 distinguishes a certain distribution $(F, G) \sim \mathcal{F}$ which they called the *forrelation* distribution. Aaronson and Ambainis [AA15] optimize this to a single-query circuit[1].

Lemma 2 (Optimized forrelation circuit)**.** *[AA15] give a single-query quantum circuit such that, for every $F, G : \{0,1\}^n \to \{0,1\}$ and $N = 2^n$, the circuit accepts with probability $\frac{1+\phi(F,G)}{2}$, where*

$$\phi(F,G) = \frac{1}{N} \sum_{i,j \in \{0,1\}^N} F(i) H_{i,j} G(j).$$

This circuit distinguishes a distribution $\mathcal{D}$ if $E_{(F,G)\sim\mathcal{D}}[\phi(F,G)]$ and $E_{(F,G)\sim\mathcal{U}_{2N}}[\phi(F,G)]$ differ sufficiently[2]. This circuit takes $\log N$ time, so it acts as the quantum algorithm in Condition (2) of Lemma 1. $\phi(F,G)$ measures the correlation between F and G in the Fourier space, motivating the choice of a Gaussian and its Fourier transform as a separating distribution.

[1] The optimized circuit first prepares a control qubit in the state $|+\rangle$. If the control is $|1\rangle$, it applies $H^{\oplus N} \to U_F$. If the control qubit is $|0\rangle$, it applies $H^{\oplus N} \to U_F \to H^{\oplus N}$. It accepts if the control qubit is in the state $|+\rangle$

[2] $\mathcal{U}_{2N}$ denotes the uniform distribution over $\{\pm 1\}^{2N}$

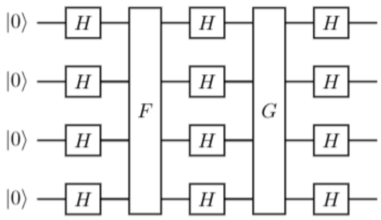

Figure 1: The naive forrelation circuit for $n = 4$, $N = 2^4$

3.2 The distribution $\mathcal{D}$

Raz and Tal [RT19] construct the distribution $\mathcal{D}$ by modifying the forrelation distribution $\mathcal{F}$ [Aar10] in a way that turns out to be crucial for the classical analysis.

Definition 1 ($\mathcal{G}$: Alternate forrelation distribution). *$z = (F, G) \sim \mathcal{G}$ where $F \sim \mathcal{N}(0,1)^N$ and $G = H_N F$.*[3]

Aaronson uses the sign of $\mathcal{G}$ to create a discrete forrelation distribution $\mathcal{F}$ (i.e., over $\{\pm 1\}$). Raz and Tal create a different discrete distribution $\mathcal{D}$ by transforming $\mathcal{G}$, then truncating.

Definition 2 (The distribution $\mathcal{G}'$). *Let $\epsilon = 1/(24 \ln N)$. $z' \sim \mathcal{G}'$ if $z' = \sqrt{\epsilon} z$ where $z \sim \mathcal{G}$.*

$\mathcal{G}'$ becomes a discrete distribution by truncation:

Definition 3 (The distribution $\mathcal{D}$). *$z' \sim \mathcal{D}$ if, for $z \sim \mathcal{G}'$, $\Pr[z'_i = 1] = \frac{1+trunc(z_i)}{2}$ where $trunc(z_i)$ truncates z_i to the interval $[-1, 1]$.*

Truncation is tricky to analyze, so Raz and Tal will analyze $\mathcal{G}'$ instead of $\mathcal{D}$. $\mathcal{G}'$ is a multi-variate Guassian distribution over $\mathbb{R}^{2N}$ with covariance matrix $\epsilon \cdot \begin{pmatrix} I_N & H_N \\ H_N & I_N \end{pmatrix}$.

3.2.1 Quantum circuit distinguishing $\mathcal{U}$ and $\mathcal{D}$

Theorem 1 (Quantum algorithm distinguishing $\mathcal{U}$ and $\mathcal{D}$). *Let Q be the 1-query quantum algorithm described by Aaronson and Ambainis.*

$$\left| \mathop{\mathbb{E}}_{(F,G)\sim\mathcal{D}}[Q(F,G)] - \mathop{\mathbb{E}}_{(F,G)\sim\mathcal{U}_{2N}}[Q(F,G)] \right| \geq \epsilon/2.$$

$\mathbb{E}_{(F,G)\sim\mathcal{U}_{2N}}[\phi(F,G)] = 0$ follows by linearity of expectation because $\forall i, j \in [N]$, $\mathbb{E}[F(i)G(j)] = 0$ since $F(i)$ and $G(j)$ are independent. To simplify their analysis of $\mathbb{E}_{(F,G)\sim\mathcal{D}}[\phi(F,G)]$, Raz and Tal show a lemma that allows them to replace $\mathcal{D}$ with $\mathcal{G}'$.

[3] $H_N \in \mathbb{R}^{NxN}$ is the Hadamard transform, where $H_N[i][j] = N^{-1/2}(-1)^{<i,j>}$. A quantum circuit can compute H_N in $\log(N)$ time by applying a Hadamard gate to each input qubit. $H_N \cdot F$ computes the discrete Fourier transform of F.

Lemma 3 (Multilinear functions on $\mathcal{D}$ and $\mathcal{G}$)**.** *Consider positive p and p_0 such that $p + p_0 = 1$. Let $F : \mathbb{R}^{2N} \to \mathbb{R}$ be a multilinear function[4] that maps $\{-1,1\}^{2N}$ to $[-1,1]$. Let $z_0 \in [-p_0, p_0]^{2N}$. Then,*

$$\mathop{\mathbb{E}}_{z \in \mathcal{G}'} [|F(trunc(z_0 + p \cdot z)) - F(z_0 + p \cdot z)|] \leq 8 \cdot N^{-2}.$$

For small ϵ, $F(x)$ is in $[-1,1]$ with high probability, so truncation has little impact on the expectation.

Proof of Theorem 1. $\mathbb{E}_{(F,G)\sim\mathcal{G}'}[\phi(F,G)] = \epsilon$ by the definition of $\mathcal{G}'$ and the fact that the Hadamard transform is its own inverse. If $p_0 = 0, p = 1$, then by Lemma 3,

$$\left| \mathop{\mathbb{E}}_{(F,G)\sim\mathcal{G}'} [\phi(F,G)] - \mathop{\mathbb{E}}_{(F,G)\sim\mathcal{D}} [\phi(F,G)] \right| \leq 8 \cdot N^{-2}.$$

This yields $E_{(F,G)\sim\mathcal{D}}[\phi(F,G)] \geq \epsilon/2$, completing the proof. □

The choice of ϵ is Raz and Tal's key modification to Aaronson's Forrelation distribution $\mathcal{F}$. Condition (2) of Lemma 1 requires that ϵ is large (i.e. $\Omega(\frac{1}{\log(N)})$) so that the quantum algorithm has sufficient advantage in distinguishing $\mathcal{D}$. On the other hand, in order to replace $\mathcal{D}$ with $\mathcal{G}'$ in the classical and quantum analysis, ϵ must be small enough that truncation is rare and Lemma 3 holds. Raz and Tal's choice of $\epsilon = \Theta(\frac{1}{\log(N)})$ satisfies both these conditions.

4 Classical lower bound

In this section, we describe the lower bound (see Theorem 3 below) on the ability of $\mathbf{AC}^0$ circuits to distinguish the forrelation distribution $\mathcal{D}$ (see Definition 3) from the uniform distribution on N bits; first, we review some mathematical preliminaries.

4.1 Facts from boolean function analysis

Definition 4 (Fourier expansion)**.** *Every boolean function $f : \{\pm 1\}^n \to \{\pm 1\}$ can be written as a linear sum of parity functions,*

$$f = \sum_{S \subseteq [n]} \hat{f}(S) \chi_S ,$$

where $\chi_S = \prod_{i \in S} x_i$.

In the above definition, the characteristic function χ_S corresponding to a set $S \subseteq [n]$ is given by the product of the bits $\{x_i\}_{i \in S}$ on any given input x. These form an orthogonal basis, which can be made orthonormal by scaling every coefficient $\hat{f}(S)$ by a factor of 2^n. We define the ℓ_1-norm of the level-2 fourier coefficients of $f : \{\pm 1\}^n \to \{\pm 1\}$ as $W_1^2[f] = \sum_{S \subseteq N, |S|=2} |\hat{f}(S)|$.

[4] $F : \mathbb{R}^{2N} \to \mathbb{R}$ is a multilinear function if $F(z) = \sum_{S \subseteq [2N]} \hat{F}(S) \cdot \prod_{i \in S} z_i$ where $\hat{F}(S) \in \mathbb{R}$ and denotes the fourier coefficients of F.

4.2 Facts from stochastic calculus

Definition 5 (Brownian stochastic process)**.** *A* brownian stochastic process *is a continuous-time stochastic process* $\{B_t\}_{t\in\mathbb{R}^+}$*, such that,*

$$\begin{gathered} B_0 = 0 \ \& \ B_t \text{ is continuous almost surely.} \\ \forall s > 0,\, B_{t+s} - B_t \text{ is distributed as } \mathcal{N}(\mathbf{0},\, s\mathcal{I}^{n\times n})\,. \end{gathered} \tag{1}$$

The above states that a brownian stochastic process behaves like a continuous n-dimensional random walk starting at the origin in the limit that a large number of steps have been taken. More importantly, any truncated multi-variate gaussian — and in particular, the forrelation distribution $\mathcal{D}$, see Definition 3 — may be viewed as the distribution induced by a multi-dimensional brownian walk stopped at an appropriate time (known as the *stopping time*).

Theorem 2 (Dynkin's formula)**.** *Given a stochastic process* $\{X_t\}_{t\in\mathbb{R}^+}$ *obtained by solving a stochastic differential equation of the form,*

$$dX_t = b(X_t)dt + \sigma(X_t)dX_t\,,$$

with a stopping time τ *having finite expected value, and a twice-differentiable vector-valued function* $f : \mathbb{R}^N \to \mathbb{R}^N$ *acting on* $\{X_t\}_{t\in\mathbb{R}^+}$*, for all* $x \in \mathbb{R}^n$*, the expected value of the function of the stopped process is,*

$$\overset{x}{\mathbb{E}}[f(X_\tau)] = f(x) + \overset{x}{\mathbb{E}}\left[\int_0^\tau Af(X_s)ds\right]\,,$$

where $\mathbb{E}^x$ *is an expectation over a walk starting at initial position* x*, and*

$$Af(x) = \lim_{t\to 0} \frac{\overset{x}{\mathbb{E}}[f(X_t)] - f(x)}{t}\,.$$

Dynkin's theorem is used crucially in the proof of Theorem 4 below (since brownian motion is obtained by solving such an equation; we set $X_t = B_t$ and $x = 0$). The operator A may be seen as parameterizing the "rate of change" of the action of the function f on the stochastic process $\{X_t\}$.

4.3 BQP vs PH

It remains to prove the following theorem (for convenience, we view $\mathcal{D}$ as a distribution on N instead of $2N$ bits):

Theorem 3 (Classical lower bound)**.** *Let* $f : \{\pm1\}^N \to \{\pm1\}$ *be computable by an* $\mathbf{AC}^0$ *circuit. Then*

$$|\mathbb{E}[f(\mathcal{D})] - \mathbb{E}[f(\mathcal{U})]| \leq \frac{\text{polylog}(N)}{\sqrt{N}}.$$

The key insight in the proof by [Wu20] of this theorem is to reframe the fooling distribution $\mathcal{D}$ defined above. Now, we view it as the final distribution of an appropriate N-dimensional *brownian walk* B_t for $\epsilon = O(\frac{1}{\ln(N)})$ steps (taking the covariance matrix given in §3.2 above).[5] As

[5] We must implement some appropriate stopping condition which corresponds to truncation as well, which may slightly change the definition of $\mathcal{D}$. However, the underlying non-truncated distribution (i.e., $\mathcal{G}'$) will be the same and essentially the same quantum upper bound will hold.

shown below, if we can get a desirably small deviation in the expected behavior of a function f (under B_t and $\mathcal{U}_N$) when $W_1^2[f]$ is small, then we easily obtain that $\mathcal{D} = B_\epsilon$ is a fooling distribution for any function f computable by a $\mathbf{AC}_0$ circuit by parametrizing the deviation with an earlier result of [Tal17] which relates the gate size and depth of an $\mathbf{AC}_0$ circuit for f to $W_1^2[f]$.

More concretely, we first present the following theorem of Wu [Wu20], which shows that a boolean function $f : \{\pm 1\}^N \to \{\pm 1\}$ with $W_1^2[f] \leq t$ has a deviation no more than $O(\epsilon\gamma t)$ between its action on inputs drawn from an eagerly terminated brownian walk in N-dimensions and the uniform distribution on the N-dimensional hypercube. Here, ϵ is the stopping time for the walk and γ the upper bound on the pair-wise covariance of any two one-dimensional walks.

Theorem 4 (Indistinguishability of mixed brownian walk & $\mathcal{U}_N$). *Let $f : \{\pm 1\}^N \to \{\pm 1\}$ be a boolean function such that, for some $t > 0$, for any restriction ρ of f, $W_1^2[f_\rho] \leq t$. Consider N-dimensional brownian motion $\{B_t\}$ with mean 0 and pairwise covariance $\leq \gamma$ mixed till a stopping time τ,*

$$\tau = \min \left\{ \epsilon,\ B_t \textit{ exits } \left[-\frac{1}{2}, \frac{1}{2} \right]^N \right\} .$$

Then, F cannot distinguish B_τ from $\mathcal{U}_{\{\pm 1\}^N}$ on average:

$$|\mathbb{E}[f(B_\tau)] - \mathbb{E}[f(\mathcal{U}_N)]| \leq 2\epsilon\gamma t \, .$$

Theorem 4 is proved using Dynkin's formula (Theorem 2) and is a consequence of the following two facts: (1) the operator A described in Theorem 2 applied to f results in an expression involving f's second-order partial derivatives evaluated at $x = 0$, and (2) if f is a multilinear polynomial, its second-order partial derivatives evaluated at $x = 0$ equal its second-level fourier coefficients. This application of Dynkin's formula allows Wu [Wu20] to simplify the original proof of Raz and Tal [RT19] considerably.

We now combine Theorem 4 with the fooling distribution $\mathcal{D} = B_\tau$ with a result (Theorem 5) shown by [Tal17] bounding the k-th level fourier coefficients of boolean functions computable by $\mathbf{AC}^0$ circuits vis-a-vis their size and depth.

Theorem 5 ($\mathbf{AC}^0$ computability & fourier weight, [Tal17]). *Given $\ell, d > 0$, every boolean function f on N variables computable by an $\mathbf{AC}^0$ circuit with no more than $\ln(N)^\ell$ gates and depth d satisfies*

$$W_1^2[f] \leq (c \cdot \ln^\ell(N))^{2(d-1)} \, .$$

Proof sketch of Theorem 3. Instantiate Theorem 4 by setting t to be the RHS from Theorem 5 and choosing a stopping time $\epsilon = \frac{1}{8\ln(N)}$ and pairwise covariance bound $\gamma = \frac{1}{\sqrt{n}}$ to obtain the desired fooling of $\mathbf{AC}^0$ circuits as expressed in Lemma 1. The above step also uses the fact that $\mathbf{AC}^0$ is closed under restrictions. □

References

[AA15] Scott Aaronson and Andris Ambainis. "Forrelation: A Problem That Optimally Separates Quantum from Classical Computing". In: STOC '15. Portland, Oregon, USA: Association for Computing Machinery, 2015, pp. 307–316. DOI: 10.1145/2746539.2746547.

[Aar10] Scott Aaronson. "BQP and the Polynomial Hierarchy". In: *Proceedings of the Forty-Second ACM Symposium on Theory of Computing*. STOC 2010. Cambridge, Massachusetts, USA: Association for Computing Machinery, 2010, pp. 141–150. DOI: 10.1145/1806689.1806711. arXiv: arXiv:0910.4698 [quant-ph].

[Ben+97] Charles H. Bennett et al. "Strengths and Weaknesses of Quantum Computing". In: *SIAM Journal on Computing* 26.5 (1997), pp. 1510–1523. DOI: 10.1137/S0097539796300933.

[BGS75] Theodore Baker, John Gill, and Robert Solovay. "Relativizations of the P = NP Question". In: *SIAM Journal on Computing* 4.4 (1975), pp. 431–442. DOI: 10.1137/0204037.

[BV93] Ethan Bernstein and Umesh Vazirani. "Quantum Complexity Theory". In: *Proceedings of the Twenty-Fifth Annual ACM Symposium on Theory of Computing*. STOC '93. San Diego, California, USA: Association for Computing Machinery, 1993, pp. 11–20. DOI: 10.1145/167088.167097.

[Fef+12] Bill Fefferman et al. "On Beating the Hybrid Argument". In: *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*. ITCS '12. Cambridge, Massachusetts: Association for Computing Machinery, 2012, pp. 468–483. DOI: 10.1145/2090236.2090273.

[FSS81] Merrick Furst, James B. Saks, and Michael Sipser. "Parity, Circuits, and the Polynomial-Time Hierarchy". In: *Proceedings of the 22nd Annual Symposium on Foundations of Computer Science*. SFCS '81. USA: IEEE Computer Society, 1981, pp. 260–270. DOI: 10.1007/BF01744431.

[Hås86] Johan Håstad. "Almost Optimal Lower Bounds for Small Depth Circuits". In: *Proceedings of the 18th Annual ACM Symposium on Theory of Computing*. STOC '86. Berkeley, California, USA: Association for Computing Machinery, 1986, pp. 6–20. DOI: 10.1145/12130.12132.

[RT19] Ran Raz and Avishay Tal. "Oracle Separation of BQP and PH". In: STOC '19. Phoenix, AZ, USA: Association for Computing Machinery, 2019, pp. 13–23. DOI: 10.1145/3313276.3316315.

[Tal17] Avishay Tal. "Tight bounds on the Fourier spectrum of AC0". In: *32nd Computational Complexity Conference (CCC 2017)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. 2017.

[Wu20] Xinyu Wu. *A stochastic calculus approach to the oracle separation of BQP and PH*. 2020. arXiv: 2007.02431 [cs.CC].

[Yao85] Andrew Chi-Chih Yao. "Separating the polynomial-time hierarchy by oracles". In: *Proceedings of the 26th Annual Symposium on Foundations of Computer Science*. SFCS '85. 1985, pp. 1–10. DOI: 10.1109/SFCS.1985.49.