The Space “Just Above” BQP

Scott Aaronson
Massachusetts Institute of
Technology
Cambridge, MA, USA
aaronson@csail.mit.edu

Adam Bouland
Massachusetts Institute of
Technology
Cambridge, MA, USA
adam@csail.mit.edu

Joseph Fitzsimons
Singapore University of
Technology and Design
Centre for Quantum
Technologies, National
University of Singapore
Singapore
joe.fitzsimons@nus.edu.sg

Mitchell Lee
Massachusetts Institute of
Technology
Cambridge, MA, USA
mitchlee@mit.edu

ABSTRACT

We explore the space “just above” BQP by defining a complexity class **naCQP** (non-adaptive Collapse-free Quantum Polynomial time) which is larger than BQP but does not contain NP relative to an oracle. The class is defined by imagining that quantum computers can perform (non-adaptive) measurements that do not collapse the wavefunction. This non-physical model of computation can efficiently solve problems such as Graph Isomorphism and Approximate Shortest Vector which are believed to be intractable for quantum computers. Furthermore, it can search an unstructured N -element list in $\tilde{O}(N^{1/3})$ time, but no faster than $\Omega(N^{1/4})$, and hence cannot solve NP-hard problems in a black box manner. In short, this model of computation is more powerful than standard quantum computation, but only slightly so. This is surprising as most modifications of BQP increase the power of quantum computation to NP or beyond.

1. INTRODUCTION

Quantum computers are believed to be strictly more powerful than classical computers, but not so much more powerful that they can solve NP-hard problems efficiently. In particular, it is known that BQP, the class of languages recognizable in polynomial time by a quantum algorithm [11], does not contain NP “relative to an oracle.” This means that there is some “black box” problem $\mathcal{O}$ for which $\text{BQP}^{\mathcal{O}} \not\supseteq \text{NP}^{\mathcal{O}}$. (For more information about the terminology, see [8].) On the other hand, many seemingly innocuous modifications of quantum mechanics—for example, allowing nonlinear transformations, non-unitary transformations, postselection,

or measurement statistics based on the p th power of the amplitudes for $p \neq 2$ —increase the power of quantum computation drastically enough that they can solve NP-hard problems (and even $\#P$ -hard problems) efficiently [6][3]. As a result, it is difficult to find natural complexity classes which are bigger than BQP but which don’t contain NP. Quantum mechanics appears to be an “island in theoryspace” in terms of its complexity-theoretic properties [3].

In this work, we explore a natural modification of quantum mechanics to obtain a complexity class which is only “slightly more powerful” than BQP. In quantum mechanics, when a system is measured, the state of the system “collapses” to its observed value; one cannot observe a quantum system without perturbing it. Here we consider the power of quantum computers which can also make “non-collapsing measurements,” which are identical to usual quantum measurements except that they do not perturb the state. We call the class of problems decidable in polynomial time in this model CQP, which stands for “Collapse-free Quantum Polynomial time.” We additionally consider a weaker version of this model, **naCQP** (non-adaptive CQP) in which the quantum operations performed must be independent of the non-collapsing measurement outcomes.

We show that quantum computers equipped with this power (even in the non-adaptive case) can solve the Graph Isomorphism problem in polynomial time, and in fact can solve any problem in SZK in a black-box manner. Since standard quantum computers cannot solve SZK-hard problems in a black-box manner [1], this implies that there is an oracle $\mathcal{O}$ for which $\text{BQP}^{\mathcal{O}} \neq \text{naCQP}^{\mathcal{O}}$. This is evidence that quantum computation with non-collapsing measurements is more powerful than standard quantum computation. Furthermore, we upper bound the power of both CQP and naCQP by showing that $\text{naCQP} \subseteq \text{CQP} \subseteq \text{BPP}^{\text{PP}}$, so both naCQP and CQP are in the counting hierarchy. In comparison the best known classical upper bound for BQP is AWPP which is contained in PP [14][7].

We also demonstrate that if (even non-adaptive) non-collapsing measurements are possible, then there is a quantum algorithm that searches an unstructured list of N elements in $\tilde{O}(N^{1/3})$ time. Furthermore any such algorithm

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 the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

ITCS’16, January 14–16, 2016, Cambridge, MA, USA.

Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM 978-1-4503-4057-1/16/01 ...\$15.00.

DOI: <http://dx.doi.org/10.1145/2840728.2840739>.

takes at least $\Omega(N^{1/4})$ time in the non-adaptive case. While the upper bound is simple, the proof of the lower bound uses a hybrid argument [10] and properties of Markov chains. We conclude that **naCQP** does not contain **NP** relative to an oracle. To our knowledge this represents the only known complexity class larger than **BQP** which provably does not admit polynomial time black-box algorithms for **NP**-hard problems. This is what we mean when we say **naCQP** is only “slightly more powerful” than **BQP**. Proving the analogous lower bound for **CQP** remains open, so it is possible that **CQP** could be more powerful than **naCQP**.

Note that introducing non-collapsing measurements into quantum mechanics allows for many strange phenomena. In particular, it allows for faster-than-light communication, it allows for quantum cloning¹, and it renders quantum query complexity and quantum communication complexity meaningless. We describe these strange consequences of non-collapsing measurements in detail in Appendix A. For this reason, we are not suggesting that “non-collapsing measurements” should be considered seriously as an amendment to quantum theory. Rather we are simply showing that non-collapsing measurements have interesting complexity-theoretic properties - namely, that they can be used to define an complexity class which is “just above” **BQP**.

2. RELATION TO PRIOR WORK

Our work is inspired by previous work on quantum computing with hidden variables by Aaronson [2]. Aaronson defines a class **DQP** (“Dynamical Quantum Polynomial Time”) by imagining a hidden variable theory is true, and that an experimenter can view the evolution of the hidden variables in real time. Additionally, he requires that the quantum operations are non-adaptive to the hidden variable values (similar to our class **naCQP**). He shows that with this power one can search in $\tilde{O}(N^{1/3})$ time and solve any problem in **SZK** in polynomial time. He additionally claims one cannot search in faster than $\Omega(N^{1/3})$ time in this model. Unfortunately, there is an error which invalidates his proof of the lower bound for search. For the interested reader, we describe this error in and correct it for a modified version of the computational model in the full version of this paper [5]. Proving the lower bound for search under Aaronson’s original computational model is challenging because we have few examples of working hidden variable theories, and therefore have little understanding of how hidden variable values could correlate over time. Note, however, that an $\Omega(N^{1/3})$ lower bound for search might hold even for Aaronson’s original model.

The classes **CQP** and **naCQP**, which we define by imagining one can perform (non-adaptive) non-collapsing measurements, seem incomparable to **DQP** - we do not know if either $\mathbf{CQP} \subseteq \mathbf{DQP}$, nor if $\mathbf{DQP} \subseteq \mathbf{CQP}$. However, we suspect that **naCQP** is a weaker class than **DQP** for several reasons. First, we can prove a polynomial lower bound for search in **naCQP**, which we don’t know how to do in **DQP**. Second, we can prove an upper bound that $\mathbf{naCQP} \subseteq \mathbf{BPP}^{\mathbf{PP}} \subseteq \mathbf{PSPACE}$.

¹Note, however, that this only arises when the quantum operations can depend on the non-collapsing measurement results. Our definition of **naCQP** does not allow cloning due to the non-adaptivity restriction. In contrast the class **CQP** does admit cloning, so might be a more powerful computational model. We discuss this issue in detail in Appendix A, and describe a related open problem in Section 8.

In contrast the best known upper bound for **DQP** is **EXP**. Table 1 summarizes the relationship between **BQP**, **naCQP**, **CQP** and **DQP**.

3. DEFINITION OF CQP AND NACQP

We assume the reader is familiar with the standard definition of **BQP** and the basics of quantum computing; for an introduction to this topic see [16]. We now give a formal definition of our model of quantum computing with non-collapsing measurements.

Let $\mathcal{Q}$ be an oracle that takes as input a quantum circuit $C = (U_1, M_1, U_2, M_2, \dots, U_T, M_T)$ and an integer $\ell \geq 0$. Here each U_i is a unitary operator on ℓ qubits composed of gates from some finite universal gate set $\mathcal{U}$, and each M_i is a standard (collapsing) measurement of zero or more qubits in the computational basis. Define a (random) sequence $\{|\psi_t\rangle\}_{t=0}^T$ of quantum states by $|\psi_0\rangle = |0\rangle^{\otimes \ell}$ and for $t > 0$, $|\psi_t\rangle$ is the resulting (random) pure state obtained when measurement M_t is applied to $U_t|\psi_{t-1}\rangle$. Note that we imagine the state of the system $|\psi_t\rangle$ is a (random) pure state for $0 \leq t \leq T$. The oracle $\mathcal{Q}$ samples the sequence $\{|\psi_t\rangle\}_{t=0}^T$ (note that the random variables $|\psi_t\rangle$ are not independent), measures $|\psi_t\rangle$ in the computational basis for every t independently, and outputs the $T+1$ measurement results, which we label $v_0, v_1, \dots, v_T$, respectively. The output of $\mathcal{Q}$ is an element of $(\{0, 1\}^\ell)^{T+1}$. Note that once the $|\psi_t\rangle$ are fixed, the $T+1$ measurement results are independent, however since the $|\psi_t\rangle$ are correlated, the measurement outcomes may be correlated.

naCQP (non-adaptive Collapse-free Quantum Polynomial-time) is then defined as the class of all languages that can be recognized in polynomial time by a deterministic Turing machine with one query to $\mathcal{Q}$, with error probability at most $\frac{1}{3}$. Note that because the base machine is polynomially bounded, the circuit C with which it queries $\mathcal{Q}$ must be polynomially sized. Furthermore, since the base machine can use the oracle to output coin flips, it makes no difference if we define the base machine to be deterministic or randomized. This class contains **BQP**, because one can always query the oracle $\mathcal{Q}$ with a **BQP** circuit, and then ignore all output except the final measurement outcome. The constant $\frac{1}{3}$ is arbitrary: we can decrease the error probability arbitrarily close to 0 by repetition, which can be accomplished by packing multiple copies of a quantum circuit into a single call to $\mathcal{Q}$. Furthermore, it turns out that the definition of **naCQP** is not affected by the choice of universal gate set $\mathcal{U}$; this is a consequence of the Solovay-Kitaev Theorem. The proof is omitted here due to lack of space, but can be found in the full version of this paper [5].

We can think of the $T+1$ measurement samples from $\mathcal{Q}$ as the results of *non-collapsing* measurements on the state vector, which give information about the state without changing it. For instance, let $|\psi_1\rangle = U_1|0\rangle^{\otimes \ell}$, let M_1, M_2 and M_3 be empty measurements, and let U_2, U_3 be the identity. Then the oracle $\mathcal{Q}$ will output the result of three independent non-collapsing measurements of $|\psi_1\rangle$ in the computational basis. The key point is that the oracle’s samples do not disturb the state of the system; only the unitary operators U_i and collapsing measurements M_i do. The oracle $\mathcal{Q}$ gives us information about the intermediate stages of the quantum computation without collapsing the state; this is what gives **naCQP** additional power over **BQP**.

Table 1: Comparison between BQP, naCQP, CQP and DQP

Property	BQP	naCQP	CQP	DQP
Contains SZK	Unknown	Yes	Yes	Yes
Contains SZK relative to all oracles	No	Yes	Yes	Yes
Upper Bound for Search	$O(N^{1/2})$	$\tilde{O}(N^{1/3})$	$\tilde{O}(N^{1/3})$	$\tilde{O}(N^{1/3})$
Lower Bound for Search	$\Omega(N^{1/2})$	$\Omega(N^{1/4})$	$\Omega(1)$	$\Omega(1)$
Upper Bound	AWPP	BPP ^{PP}	BPP ^{PP}	EXP

Note that by requiring the quantum circuit C to be specified up front, we have enforced the condition that the circuit is non-adaptive to the non-collapsing measurement outcomes (hence the name **naCQP**). To define **CQP**, we consider the case where the base machine can query the oracle $\mathcal{Q}$ adaptively. That is, the base machine can first specify $U_1, M_1, \dots, U_t, M_t$ and receive samples $v_1 \dots v_t$, then based on those samples select $U_{t+1}, M_{t+1} \dots U_{t'}, M_{t'}$ and receive samples $v_{t+1} \dots v_{t'}$, etc. **CQP** is then defined analogously to be the class of languages which can be decided in polynomial time with adaptive queries to $\mathcal{Q}$, with error probability at most $1/3$. This class captures the power of generic computations with non-collapsing measurements.

Note that we explicitly allow for intermediate (collapsing) measurements in our model. In the definition of **BQP**, the principle of deferred measurement tells us that this is not necessary; the power of standard quantum computers is unchanged by the inclusion of intermediate collapsing measurements. However, in our model this makes a crucial difference. Indeed, suppose that we did not allow for intermediate collapsing measurements; then this model would be simulable in **BQP** with a polynomial amount of overhead. If there are no intermediate measurements M_i , then $|\psi_t\rangle = U_t U_{t-1} \dots U_1 |0\rangle^{\otimes \ell}$ are no longer random variables but are deterministic pure states, each preparable with a polynomially sized quantum circuit. So a **BQP** machine could simply prepare $|\psi_1\rangle$ and measure it, then prepare $|\psi_2\rangle$ from scratch and measure it, etc. to obtain the samples $v_0, \dots, v_T$. This would incur at most quadratic overhead.

When we add intermediate measurements into our model, this simulation strategy no longer works. Indeed, suppose that we performed measurement M_1 to obtain a random state $|\psi_1\rangle$. If we wanted to reproduce this state with a **BQP** machine, we could try applying M_1 to $U_1 |0\rangle^{\otimes \ell}$. However, it might be that the probability of obtaining the same outcome for M_1 is exponentially small, and hence the **BQP** machine could not prepare another copy of $|\psi_1\rangle$ in polynomial time.

In short, the power of this model comes from the fact that we can perform intermediate measurements which collapse the wave function, and afterwards we can examine the resulting pure state $|\psi_i\rangle$ (which might not be efficiently preparable with a **BQP** machine) using multiple non-collapsing measurements. In the next section we will show how to leverage these properties to solve any problem in **SZK** in polynomial time.

4. SZK IS CONTAINED IN NACQP

We will now describe how to use the peculiarities of non-collapsing measurements to solve any problem in **SZK** in polynomial time. The proof uses essentially the ideas of Aaronson [2], with minor simplifications.

SZK was originally defined as the class of languages admitting statistical zero-knowledge proofs. The precise definition of a statistical zero-knowledge proof can be found in [18], but it is not important here. **SZK** includes important problems such as Graph Isomorphism and Approximate Shortest Vector. It has been a long-standing open problem whether or not these problems can be solved in quantum polynomial time. Ettinger, Høyer and Knill showed that Graph Isomorphism (and indeed any hidden subgroup problem) can be solved in a black box manner with a polynomial number of queries to the black box, but with exponential post-processing time [13]. On the other hand, Aaronson [1] showed that **BQP** does not admit a black-box algorithm for the collision problem, and hence there is an oracle relative to which **SZK** is not in **BQP**.

In contrast, we show that quantum computers with non-collapsing measurements can solve any problem in **SZK** efficiently, i.e. $\text{SZK} \subseteq \text{naCQP}$. It is enough to prove that Statistical Difference, a problem shown in [18] to be **SZK**-complete, is in **naCQP**. The statistical difference problem is to determine, for two functions $P_0, P_1 : \{0, 1\}^n \rightarrow \{0, 1\}^m$ specified by classical circuits, whether the distributions of $P_0(X), P_1(X)$ for uniformly random X are close or far. Here, two distributions are “close” if their total variation distance is less than $\frac{1}{3}$ and they are “far” if their total variation distance is more than $\frac{2}{3}$.

We now show how to solve this efficiently if we have access to non-collapsing measurements.

THEOREM 4.1. *The Statistical Difference problem can be solved in polynomial time in naCQP, and therefore $\text{SZK} \subseteq \text{naCQP}$.*

PROOF. By the Polarization Lemma of Sahai and Vadhan [18, Lemma 3.3], we can assume that the distributions $P_0(X)$ and $P_1(X)$ have total variation distance less than 2^{-n^c} or more than $1 - 2^{-n^c}$, for any constant c . For now, assume that the distributions have total variation distance equal to either 1 or 0.

Our algorithm for the statistical difference problem is as follows. Prepare the state

$$\frac{1}{2^{(n+1)/2}} \sum_{b \in \{0,1\}, x \in \{0,1\}^n} |b\rangle |x\rangle |P_b(x)\rangle.$$

Now, measure the third register with a collapsing measurement to obtain a state $|\phi\rangle$ on the first two registers. If the distributions P_0, P_1 have total variation distance 1, then $|\phi\rangle$ will be of the form $|b\rangle |\psi\rangle$ for some b and $|\psi\rangle$. On the other hand, if they have total variation distance 0, then $|\phi\rangle$ will be an equal superposition $\frac{1}{\sqrt{2}}(|0\rangle |\psi_0\rangle + |1\rangle |\psi_1\rangle)$ where $|\psi_0\rangle$ and $|\psi_1\rangle$ have unit norm. We can distinguish the two cases by now repeatedly performing non-collapsing measurements

and examining the value of the first register. If P_0, P_1 have total variation distance 1, then all of these measurements will give the same value b ; if P_0 and P_1 have total variation distance 0, then each of these measurements will independently give 0 with probability $\frac{1}{2}$ and 1 with probability $\frac{1}{2}$. We can distinguish the two cases with probability $3/4$ by performing three non-collapsing measurements and looking at whether or not they yielded identical values of the first register.

Furthermore, the fact that the total variation distances are merely exponentially close to 0 or 1, rather than actually being equal to 0 or 1, makes little difference. One can show that the probability of seeing the same measurement outcome three times is at most $\frac{1}{4} + O(2^{-n^c})$ if P_0 and P_1 are exponentially close and at least $1 - O(2^{-n^c})$ if P_0 and P_1 are exponentially far apart. We provide a detailed proof of this fact in Appendix B. Therefore our algorithm will have error probability at most $1/3$.

□

Hence SZK is in naCQP, and furthermore we can solve SZK problems in naCQP in a black box manner, i.e. relative to any oracle. Since [1] has the result that $\text{SZK} \not\subseteq \text{BQP}$ relative to an oracle, we have the immediate corollary²:

COROLLARY 4.1. *There exists an oracle $\mathcal{O}$ such that $\text{naCQP}^{\mathcal{O}} \neq \text{BQP}^{\mathcal{O}}$.*

5. SEARCH IN $O(N^{1/3})$ TIME

Suppose that we are given query access to a function $f : \{0, 1\}^n \rightarrow \{0, 1\}$ such that the preimage $f^{-1}(1)$ contains exactly one element, x . In the classical randomized computational model, we can find x in $O(N)$ time, where $N = 2^n$, but no faster. In the quantum computational model, on the other hand, we can find x in $O(N^{1/2})$ time using Grover's search algorithm [15], but no faster [10].

Here we show that quantum computers equipped with (non-adaptive) non-collapsing measurements can search in $\tilde{O}(N^{1/3})$ time, where the tilde hides factors in $\log N$. The basic idea is to run $N^{1/3}$ Grover iterations, and then make $N^{1/3}$ non-collapsing measurements of the resulting state. Then with high probability the marked item will be seen. This is a simplification of the proof given in [2, Theorem 10] for DQP. We now formalize this idea below:

THEOREM 5.1. *Suppose, in the definition of naCQP, that the unitary operators $U_1, \dots, U_T$ are now allowed to query f . That is, we are given access to the n -qubit gate U_f defined by $U_f |y\rangle = (-1)^{f(y)} |y\rangle$ for all $y \in \{0, 1\}^n$, as well as controlled- U_f . Then there is a naCQP algorithm to find the value of x that uses $O(N^{1/3})$ queries and $\tilde{O}(N^{1/3})$ time.*

PROOF. Prepare the uniform superposition of all basis states, apply $i = N^{1/3}$ Grover iterations [15], then query the oracle to record whether or not each basis state is marked in an ancilla. We obtain the state

$$\sin((2i+1)\theta) |x\rangle |1\rangle + \cos((2i+1)\theta) \sum_{y \in \{0,1\}^n, y \neq x} 2^{-\frac{N-1}{2}} |y\rangle |0\rangle$$

²Note that when we say $\text{naCQP}^{\mathcal{O}}$, we mean that circuits given in the input to $\mathcal{Q}$ in the definition of naCQP can contain quantum calls to the oracle.

where $\sin(\theta) = 2^{-n/2}$ and $i = 2^{n/3}$. For small x we have $\sin(x) \approx x$, so for large n we have $\theta = \Theta(2^{-n/2})$, so $\sin((2i+1)\theta) = \Theta(2^{-n/6})$.

Now make $O(N^{1/3} \log N)$ non-collapsing measurements. We claim that with high probability, the marked item x will appear at least once. Indeed, the marked item x appears with probability at least $\Omega(N^{-1/3})$ in each non-collapsing measurement outcome, so it occurs at least once with probability more than $1 - (\log N + 1)e^{-\log N} = 1 - o(1)$. □

Note that if we are willing to use an enormous amount of time, we can search in the naCQP model using only one query: just query the oracle in superposition and then perform $O(N)$ non-collapsing measurements. Indeed as we note in the introduction, any function f has query complexity 1 in this model, although this approach requires exponentially many non-collapsing measurements. Therefore in this model of computation, the relevant measure of complexity of an algorithm is the number of queries Q plus the number of non-collapsing measurements T used by the algorithm. Our above algorithm uses $Q + T = \tilde{O}(N^{1/3})$ of each, with $\tilde{O}(N^{1/3})$ post-processing time, so we say it “runs in time $\tilde{O}(N^{1/3})$ ”.

6. LOWER BOUNDS FOR SEARCH

We now show that our search algorithm in section 5 cannot be improved by much; in particular there is no way to solve search in faster than $N^{1/4}$ time, even with non-adaptive non-collapsing measurements. Proving the analogous lower bound for adaptive non-collapsing measurements (i.e. for the class CQP) remains open.

THEOREM 6.1. *Suppose, in the definition of naCQP, that the unitary operators $U_1, \dots, U_T$ are now allowed to query f . Let Q be the number of queries to f made by a naCQP algorithm, and T be the number of non-collapsing measurements. Then any naCQP algorithm to find the value of x obeys $Q + T = \Omega(N^{1/4})$, and hence search requires $\Omega(N^{1/4})$ time.*

In other words, there is no “black box” polynomial-time algorithm for NP-hard problems, even when given access to non-collapsing measurements. This is evidence that the class naCQP does not contain NP. The following corollary follows immediately from the well-known “diagonalization method” of Baker, Gill, and Solovay [9]:

COROLLARY 6.1. *There exists an oracle $\mathcal{O}$ such that $\text{NP}^{\mathcal{O}} \not\subseteq \text{naCQP}^{\mathcal{O}}$.*

We now outline the proof of Theorem 6.1. The following lemma is essential: it bounds the total variation distance between two Markov distributions.

LEMMA 6.1. *Suppose that $T \geq 1$, and that $v = (v_0, \dots, v_T)$ is a random variable governed by a Markov distribution. That is, for all $1 \leq i \leq T$, v_i is independent of $v_0, \dots, v_{i-2}$ conditioned on a particular value of v_{i-1} . Let $w = (w_0, \dots, w_T)$ be another random variable governed by a Markov distribution. If $d_{TV}(\cdot, \cdot)$ denotes the total variation distance between random variables, then*

$$d_{TV}(v, w) \leq 2 \sum_{i=1}^T d_{TV}((v_{i-1}, v_i), (w_{i-1}, w_i)).$$

PROOF. We proceed by induction on T . The base case $T = 1$ is trivial. For $T > 1$, since w_T depends only on w_{T-1} (by the Markov property), it is equal to $A(w_{T-1})$ for some randomized process A ; let $w'_T := A(v_{T-1})$ be a variable that depends on v_{T-1} in exactly the same way that w_T depends on w_{T-1} . Then, define the random variable $v' = (v_0, \dots, v_{T-1}, w'_T)$. By the triangle inequality,

$$d_{TV}(v, w) \leq d_{TV}(v, v') + d_{TV}(v', w). \quad (1)$$

Applying the same randomized process to two random variables cannot increase their total variation distance [18]. We can generate random variables identically distributed to v and v' by applying a suitable randomized process to (v_{T-1}, v_T) and (v_{T-1}, w'_T) . We can also generate random variables identically distributed to v' and w by applying a suitable randomized process to $(v_0, \dots, v_{T-1})$ and $(w_0, \dots, w_{T-1})$. Therefore, the right hand side of (1) is bounded above by

$$\begin{aligned} & d_{TV}((v_{T-1}, v_T), (v_{T-1}, w'_T)) \\ & + d_{TV}((v_0, \dots, v_{T-1}), (w_0, \dots, w_{T-1})). \end{aligned}$$

By the triangle inequality, $d_{TV}((v_{T-1}, v_T), (v_{T-1}, w'_T))$ is upper bounded by

$$\begin{aligned} & d_{TV}((v_{T-1}, v_T), (w_{T-1}, w_T)) \\ & + d_{TV}((w_{T-1}, w_T), (v_{T-1}, w'_T)) \\ & = d_{TV}((v_{T-1}, v_T), (w_{T-1}, w_T)) \\ & + d_{TV}(v_{T-1}, w_{T-1}) \\ & \leq 2d_{TV}((v_{T-1}, v_T), (w_{T-1}, w_T)). \end{aligned}$$

Putting all of this together, we get that $d_{TV}(v, w)$ is upper bounded by

$$\begin{aligned} & 2d_{TV}((v_{T-1}, v_T), (w_{T-1}, w_T)) \\ & + d_{TV}((v_0, \dots, v_{T-1}), (w_0, \dots, w_{T-1})). \end{aligned}$$

The result follows from induction. $\square$

LEMMA 6.2. *The trace distance between two pure states $|\psi\rangle$ and $|\phi\rangle$ is less than or equal to the 2-norm $\| |\psi\rangle - |\phi\rangle \|_2$.*

PROOF. The trace distance between $|\psi\rangle$ and $|\phi\rangle$ is equal to $\sqrt{1 - |\langle \psi | \phi \rangle|^2}$ [16], and the 2-norm $\| |\psi\rangle - |\phi\rangle \|_2$ is $\sqrt{2 - 2\text{Re}(\langle \psi | \phi \rangle)}$. The inequality follows from $|\langle \psi | \phi \rangle| \leq 1$. $\square$

From the hybrid argument of [10], we have the following:

LEMMA 6.3. *For all t , if there are no measurements made before time t , we have $\sum_{x=0}^{N-1} \| |\psi_t\rangle - |\psi_t(x)\rangle \|_2^2 \leq 4Q^2$.*

With these facts, we can now prove Theorem 6.1. We provide an outline of the proof here, and the full proof can be found in Appendix C. The basic idea is to realize that the non-collapsing measurement outcomes form a Markov chain, because the distribution of any non-collapsing measurement is independent once the results of the previous intermediate collapsing measurements are fixed. So, letting v and $v(x)$ be the distributions on non-collapsing measurement outcomes when the marked item is absent or present at x , by applying Lemma 6.1, we have that

$$\frac{1}{3} \leq d_{TV}(v, v(x)) \leq 2 \sum_{i=1}^T d_{TV}((v_{i-1}, v_i), (v_{i-1}(x), v_i(x))).$$

Here the lower bound on d_{TV} comes from the fact that our algorithm can distinguish whether or not a marked item is present with probability $2/3$, and hence these distributions must be $1/3$ -far apart for all x .

Lemma 6.3 tells us that there is some x for which the marginal distributions v_i and $v_i(x)$ are close. However, this isn't sufficient to upper bound the quantity on the right of this inequality, because the correlations between the distributions at steps $i-1$ and i (which are induced by the intermediate collapsing measurements) might make the distributions easier to distinguish³. Hence in order to prove this lower bound, we have to substantially strengthen the hybrid argument [10] to show that the correlations induced by the collapsing measurement outcomes do not allow $d_{TV}((v_{i-1}, v_i), (v_{i-1}(x), v_i(x)))$ to be large. By carefully keeping track of these induced correlations, we show in Appendix C that there is some x for which

$$d_{TV}(v, v(x)) \leq 2 \sum_{i=1}^T d_{TV}((v_{i-1}, v_i), (v_{i-1}(x), v_i(x))) \leq \frac{20TQ}{\sqrt{N}}.$$

where Q is the number of queries made by the algorithm and T is the number of non-collapsing measurements. Combining this with the fact that $d_{TV}(v, v(x)) \geq \frac{1}{3}$ for all x , this implies

$$\frac{20TQ}{\sqrt{N}} \geq \frac{1}{3},$$

and hence the running time of the algorithm is at least $T + Q = \Omega(N^{1/4})$.

7. AN UPPER BOUND ON CQP

We now show that CQP is contained in the class BPP^{PP} . Since $\text{naCQP} \subseteq \text{CQP}$, this places both classes in the second level of the counting hierarchy. By comparison, the best known upper bound for BQP is AWPP which is contained in PP [14][7].

THEOREM 7.1. $\text{CQP} \subseteq \text{BPP}^{\text{PP}}$.

PROOF. First note that $\text{BPP}^{\text{PP}} = \text{BPP}^{\#P}$, because one can always use a PP oracle to count with only polynomial overhead. Therefore it suffices to show $\text{CQP} \subseteq \text{BPP}^{\#P}$. We now show how to simulate the sampling oracle $\mathcal{Q}$ in $\text{BPP}^{\#P}$. Our algorithm will work for adaptive queries as well. Since $\text{CQP} = \text{BPP}^{\mathcal{Q},1}$, this implies the claim.

Suppose we wish to simulate a sample from the oracle $\mathcal{Q}$ with input circuit $C = (U_1, M_1, \dots, U_T, M_T)$ on n qubits. Since the choice of gate set does not matter (for a proof of this, please refer to the full version of this paper [5]), without loss of generality we can assume our circuit is composed of only Toffoli and Hadamard gates, which are universal by a result of Shi [19].

We first simulate the result of the measurement M_1 . Suppose without loss of generality that M_1 measures the first k qubits and gets outcome $x_1 \dots x_k \in \{0, 1\}^k$. Following the techniques of Adleman, DeMarrais, and Huang [7], we can write the probability that x_1 is 0 or 1 as an exponential sum

³To see how this could happen in general, consider the following two Markov distributions on two bits: D_1 outputs 00 or 11 with equal probability, and D_2 outputs 01 or 10 with equal probability. These have identical marginals on each bit, but are perfectly distinguishable due to the correlations between their bits.

of poly-time-computable terms (since U_1 is specified by a poly-sized circuit). Since we chose Hadamard and Toffoli as our gate set, all terms in the sum are of the form $\frac{\pm 1}{2^k}$, where k is the number of Hadamard gates in U_1 . Hence using the $\#P$ oracle, we can compute $\Pr[x_1 = 1]$ exactly in binary, and then flip a coin with bias p using the base BPP machine to obtain outcome $x_1 \in \{0, 1\}$ with this probability.

We've now sampled the value of x_1 . To sample the value of x_2 , note that we can also express $\Pr[x_2 = 1 | x_1 = 0]$ as a sum of exponentially many terms, each of which is poly-time computable and takes values in $\frac{\pm 1}{2^k}$. Therefore using the $\#P$ oracle, we can exactly compute the *conditional* probability that $x_2 = 1$ given our sampled value of x_1 ; in other words the $\#P$ oracle can compute the probabilities of measurement outcomes under post-selection. In this way we can sample x_2 , then x_3 , etc. obtain a sample $x_1 \dots x_k \in \{0, 1\}^k$ as desired.

Now suppose we wish to sample the variable $v_1 \in \{0, 1\}^n$ which is the result of a non-collapsing measurement on the state remaining after measurement M_1 yields value $x_1 \dots x_k$. As noted above, using the $\#P$ oracle, we can compute the marginal probability that any qubit is 1, postselected on a particular measurement outcome. Hence using the $\#P$ oracle, we can draw the sample v_1 using n queries to the oracle. We can continue this process to simulate M_2 , then sample v_2 , etc. Therefore we can draw a sample from $\mathcal{Q}$ using $O(nT)$ queries to the $\#P$ oracle. Note that this simulation works when the U_i and the M_i are chosen adaptively, since for each t the base BPP machine receives the non-collapsing measurement samples $v_0 \dots v_t$ before proposing the next unitary U_{t+1} and measurement M_{t+1} . Hence this shows $\text{CQP} \subseteq \text{BPP}^{\text{PP}}$.

□

An open question is whether or not we can improve this upper bound to show $\text{naCQP} \subseteq \text{PP}$. This seems difficult because $\text{SZK} \subseteq \text{naCQP}$, and it is open whether or not $\text{SZK} \subseteq \text{PP}$. In fact, Aaronson [4] showed that there is an oracle separation between SZK and a weaker class A_0PP , and left open the problem of finding an oracle relative to which SZK is not contained in PP . If such an oracle exists, it would imply one could not show $\text{naCQP} \subseteq \text{PP}$ with a relativizing proof.

One natural approach to showing $\text{naCQP} \subseteq \text{PP}$ is to use the fact that $\text{PP} = \text{PostBQP}$ [3], and design a post-selected quantum circuit to simulate the oracle $\mathcal{Q}$. However, the most naive way of trying to do this fails. Suppose that one tried the following: to simulate the oracle's output under $C = (U_1, M_1, \dots, U_T, M_T)$ on n qubits, create a post-selected circuit C' on nT qubits which runs $U_1 M_1$ on the first n qubits, $U_1 M_1 U_2 M_2$ on the second n qubits, etc, and post-selects on them receiving the same outcomes for the intermediate measurements. While this superficially looks like what the oracle $\mathcal{Q}$ performs, this approach does not sample from the correct distribution on outputs. Suppose the probability that the outcome of M_1 is 1 is p . Then the probability one sees $M_1 = 1$ in the final output of C' will be $\frac{p^T}{p^T + (1-p)^T}$, while the quantum oracle $\mathcal{Q}$ will sample $M_1 = 1$ with probability p . For this reason it seems difficult to generate a sample from $\mathcal{Q}$ with a post-selected circuit, and hence difficult to place naCQP in PP .

8. OPEN QUESTIONS

We leave many questions about the complexity classes DQP , CQP and naCQP unanswered.

1. We demonstrated a $\tilde{O}(N^{1/3})$ -time algorithm for the search problem in the naCQP model, as well as the result that any search algorithm takes $\Omega(N^{1/4})$ time. Is it possible to close the gap between these two bounds? If we disallow intermediate collapsing measurements, then we can prove an $N^{1/3}$ lower bound for search (a proof is included in Appendix D). However proving an $N^{1/3}$ lower bound when there are intermediate measurements remains open.
2. Can we demonstrate a lower bound, superpolynomial in $\log N$, for the running time of a search algorithm in the DQP model? The proof given in [2] of an $\Omega(N^{1/3})$ lower bound is flawed (as discussed in the full version of this paper [5]).
3. Is there a hierarchy of computational models for which the k th allows searching in $\tilde{O}(N^{1/k})$ time?
4. Can we improve the upper bound $\text{naCQP} \subseteq \text{BPP}^{\text{PP}}$ to $\text{naCQP} \subseteq \text{P}^{\text{PP}}$ or $\text{naCQP} \subseteq \text{PP}$? One possible way to approach this problem is to use the alternative formulation of PP as PostBQP [3], however a straightforward application of this result does not seem to work.
5. How powerful is the class CQP ? In particular, can one prove a polynomial lower bound for search in CQP ?
6. More generally, what is the power of quantum computers which have the ability to clone quantum states? Such devices could clearly simulate computations in naCQP and CQP - to simulate a non-collapsing measurement, simply clone the state and measure in the computational basis - but may be more powerful than either class. For a further discussion see Appendix A.

9. ACKNOWLEDGEMENTS

S.A. was supported in part by an Alan T. Waterman Award. A.B. was supported in part by the National Science Foundation Graduate Research Fellowship under Grant No. 1122374 and by the Center for Science of Information (CSOI), an NSF Science and Technology Center, under grant agreement CCF-0939370. J.F. was supported in part by the Singapore National Research Foundation under NRF Award No. NRF-NRFF2013-01. M.L. was supported by the MIT SPUR program.

10. REFERENCES

- [1] Scott Aaronson. Quantum lower bound for the collision problem. In *Proceedings of the thirty-fourth annual ACM symposium on Theory of computing*, STOC '02, pages 635–642, New York, NY, USA, 2002. ACM.
- [2] Scott Aaronson. Quantum computing and hidden variables. *Phys. Rev. A*, 71:032325, Mar 2005.
- [3] Scott Aaronson. Quantum computing, postselection, and probabilistic polynomial-time. In *Proceedings of the Royal Society A*, page 0412187, 2005.

- [4] Scott Aaronson. Impossibility of Succinct Quantum Proofs for Collision-Freeness. *Quantum Information and Computation*, 12:21-28, 2012.
- [5] Scott Aaronson, Adam Bouland, Joseph Fitzsimons, and Mitchell Lee. The space "just above" BQP. Technical report, arXiv:1412.6507, 2014.
- [6] Daniel S. Abrams and Seth Lloyd. Nonlinear quantum mechanics implies polynomial-time solution for NP-complete and $\#P$ problems. *Phys. Rev. Lett.*, 81, 3992-3995, 1998.
- [7] Leonard M. Adleman, Jonathan Demarrais, Ming-deh, and A. Huang. Quantum computability. *SIAM Journal of Computation*, pages 1524-1540, 1997.
- [8] Sanjeev Arora and Boaz Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, New York, NY, USA, 1st edition, 2009.
- [9] T. Baker, J. Gill, and R. Solovay. Relativizations of the $\mathcal{P} = ? \mathcal{NP}$ question. *SIAM Journal on Computing*, 4(4):431-442, 1975.
- [10] Charles H. Bennett, Ethan Bernstein, Gilles Brassard, and Umesh Vazirani. Strengths and weaknesses of quantum computing. *SIAM J. Comput.*, 26(5):1510-1523, October 1997.
- [11] Ethan Bernstein and Umesh Vazirani. Quantum complexity theory. In *in Proc. 25th Annual ACM Symposium on Theory of Computing, ACM*, pages 11-20, 1993.
- [12] Christopher M. Dawson and Michael A. Nielsen. The Solovay-Kitaev algorithm. *Quantum Info. Comput.*, 6(1):81-95, January 2006.
- [13] Mark Ettinger, Peter Høyer, and Emanuel Knill. The quantum query complexity of the hidden subgroup problem is polynomial. *Information Processing Letters*, 91(1):43-48, July 2004.
- [14] Lance Fortnow and John Rogers. Complexity limitations on quantum computation. In *Proc. IEEE CCC'98*, p. 202-209, 1998.
- [15] Lov K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, STOC '96, pages 212-219, New York, NY, USA, 1996. ACM.
- [16] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information (Cambridge Series on Information and the Natural Sciences)*. Cambridge University Press, 1 edition, January 2004.
- [17] A Peres. Two simple proofs of the Kochen-Specker theorem. *Journal of Physics A: Mathematical and General*, 24(4):L175, 1991.
- [18] A. Sahai and S.P. Vadhan. A complete promise problem for statistical zero-knowledge. In *Foundations of Computer Science, 1997. Proceedings., 38th Annual Symposium on*, pages 448-457, 1997.
- [19] Y. Shi. Both Toffoli and controlled-NOT need little help to do universal quantum computing. *Quantum Information & Computation*, 3:1 pp. 84-92 (2003).

APPENDIX

Appendix

A. STRANGE PROPERTIES OF NONCOLLAPSING MEASUREMENTS

Here we show why allowing non-collapsing measurements in quantum mechanics allows for faster than light communication, renders quantum query complexity and quantum communication complexity meaningless, and allows for quantum cloning. We also discuss the relationship between naCQP and quantum computers which have the ability to clone.

To see that non-collapsing measurements allow for faster-than-light communication: suppose two players Alice and Bob share n EPR pairs. Then Alice can send a bit of information to Bob with probability $1 - 2^{-n}$. To see this, suppose Alice makes collapsing measurements in the 0/1 basis on her share of the EPR pairs to send a 0, and measures in the $\pm$ -basis to send a 1. Now Bob makes two non-collapsing measurements in the 0/1 basis on his half of the EPR pairs. If Alice had sent a 0, then Bob will see the same outcome each time. If Alice had sent a 1, then Bob will see a random string each time, so with probability $1 - 2^{-n}$ Bob will see two different outcomes. Thus Bob can tell which basis Alice measured in with high probability, and Alice and Bob can communicate faster than light.

We now explain why with non-collapsing measurements, the quantum query complexity and quantum communication complexity of any function is 1. Suppose one wishes to evaluate $f(x)$ where $x = x_1 \dots x_N$. Then one can prepare the superposition $\sum_i |i\rangle |x_i\rangle$ with one query to the oracle, and make $O(N \log N)$ non-collapsing measurements of this state to observe the value of each x_i and compute the function. Similarly, in the context of communication complexity, one player can simply encode their input $x \in \{0, 1\}^n$ into the state $\cos \theta_x |0\rangle + \sin \theta_x |1\rangle$ where $\theta_x = \frac{x}{2^n} \frac{\pi}{2}$. By performing roughly 2^n non-collapsing measurements, the other player can learn θ_x and hence x , with only one quantum bit of communication. Note that although these example algorithms use only one query or one qubit of communication, respectively, they use a large number of non-collapsing measurements. For this reason, when we prove lower bounds for naCQP, we lower bound the number of queries *plus* the number of non-collapsing measurements required, rather than the number of queries alone.

To see that non-collapsing measurements allow for cloning: given a quantum state ψ on n qubits, one could perform $2^{O(n)}$ non-collapsing measurements to characterize the state using tomography, and then (approximately) reproduce the state. This "approximate cloning" operation takes exponential time for generic states, and is non-unitary.

Note, that the class of computations considered in naCQP cannot clone, even for states of $O(\log(n))$ qubits, since the naCQP machine cannot perform further quantum computations after receiving the non-collapsing measurement results (i.e. because of the non-adaptivity restriction). In contrast, if the circuit could depend on the non-collapsing measurement results (as in CQP), then one could clone states of $O(\log(n))$ qubits to polynomial accuracy, which is a non-unitary operation. Hence following the result of Abrams and Lloyd [6], the power of CQP class might include NP or even $\#P$, though we do not know if this is the case. A broader related open problem is: what is the power of quantum com-

puters which are given the ability to clone? Such devices could clearly simulate **naCQP** and **CQP** computations - to simulate a non-collapsing measurement, simply clone and measure in the computational basis. However, it's unclear how powerful such quantum devices would be.

B. A DETAILED PROOF THAT SZK IS IN NACQP

Here we provide a detailed analysis of the probability of error in our **naCQP** algorithm for solving the Statistical Difference problem, which is **SZK**-complete.

Let's briefly recap the algorithm. Suppose we're given an instance of Statistical Difference, and apply the Polarization Lemma of Sahai and Vadhan [18] to obtain two circuits P_0 and P_1 which encode probability distributions D_0 and D_1 which satisfy either $d_{TV}(D_0, D_1) \leq \epsilon$ or $d_{TV}(D_0, D_1) \geq 1 - \epsilon$, where $\epsilon = 2^{-O(n^c)}$ for some constant c . We now prepare the state

$$\frac{1}{2^{(n+1)/2}} \sum_{b \in \{0,1\}, x \in \{0,1\}^n} |b\rangle |x\rangle |P_b(x)\rangle$$

Now, measure the third register with a collapsing measurement to obtain some outcome y , and then perform three non-collapsing measurements on the b register to obtain outcomes b_1, b_2, b_3 . If $b_1 = b_2 = b_3$ then output the distributions were $(1 - \epsilon)$ -far in total variation distance, otherwise output they were ϵ -close in total variation distance.

We will now compute the probability this algorithm makes an error when the input distributions are ϵ -close in total variation distance.

Let $D_b(x)$ denote the probability that distribution D_b outputs string x . The probability of seeing outcome y in the $P_b(x)$ register under our collapsing measurement is

$$\frac{1}{2} (D_0(y) + D_1(y)).$$

Conditioned on seeing outcome y , one can easily compute that the probability of obtaining outcome $b_1 = b_2 = b_3$ (which causes the algorithm to err) is $\frac{D_0(y)^3 + D_1(y)^3}{(D_0(y) + D_1(y))^3}$.

Hence the total probability of error in this case is given by

$$\Pr[\text{error}] = \sum_y \frac{D_0(y) + D_1(y)}{2} \frac{D_0(y)^3 + D_1(y)^3}{(D_0(y) + D_1(y))^3}$$

Let $\delta(y) = D_1(y) - D_0(y)$. So $\sum_y \delta(y) = 0$ and $\sum_y |\delta(y)| \leq 2\epsilon$ by our promise on the total variation distance between D_0 and D_1 . Hence we have by direct calculation that

$$\begin{aligned} \Pr[\text{error}] &= \frac{1}{2} \sum_y \frac{D_0(y)^3 + D_1(y)^3}{(D_0(y) + D_1(y))^2} \\ &= \frac{1}{2} \sum_y \frac{D_0(y)^3 + (D_0(y) + \delta(y))^3}{(2D_0(y) + \delta(y))^2} \\ &= \frac{1}{2} \sum_y \frac{2D_0(y)^3 + 3D_0(y)^2\delta(y) + 3D_0(y)\delta(y)^2 + \delta(y)^3}{4D_0(y)^2 + 4D_0(y)\delta(y) + \delta(y)^2} \\ &= \frac{1}{2} \sum_y \frac{D_0(y)}{2} + \frac{D_0(y)^2\delta(y) + \frac{5}{2}D_0(y)\delta(y)^2 + \delta(y)^3}{4D_0(y)^2 + 4D_0(y)\delta(y) + \delta(y)^2} \\ &= \frac{1}{4} + \frac{1}{2} \sum_y \delta(y) \frac{D_0(y)^2 + \frac{5}{2}D_0(y)\delta(y) + \delta(y)^2}{4D_0(y)^2 + 4D_0(y)\delta(y) + \delta(y)^2} \end{aligned}$$

$$\begin{aligned} &\leq \frac{1}{4} + \frac{1}{2} \sum_y |\delta(y)| \frac{D_0(y)^2 + \frac{5}{2}D_0(y)|\delta(y)| + |\delta(y)|^2}{4D_0(y)^2 + 4D_0(y)\delta(y) + \delta(y)^2} \\ &\leq \frac{1}{4} + \frac{1}{2} \sum_y |\delta(y)| \frac{4D_0(y)^2 + 4D_0(y)|\delta(y)| + |\delta(y)|^2}{4D_0(y)^2 + 4D_0(y)\delta(y) + \delta(y)^2} \\ &= \frac{1}{4} + \frac{1}{2} \sum_y |\delta(y)| \leq \frac{1}{4} + \epsilon \end{aligned}$$

Where the last two lines follow from the fact that all terms in the sum are non-negative, and the fact that $\sum_y |\delta(y)| \leq 2\epsilon$. Hence the probability of error in the case is upper bounded by $\frac{1}{4} + \epsilon = \frac{1}{4} + O(2^{-n^c})$, so the algorithm has probability of error $< \frac{1}{3}$ for sufficiently large n as desired.

We now bound the probability of error in the case that the distributions are far apart. In this case, the probability of getting an outcome where b_1, b_2, b_3 are not all at the same, conditioned on measuring y , is given by

$$\frac{3D_0(y)^2D_1(y) + 3D_0(y)D_1(y)^2}{(D_0(y) + D_1(y))^3}.$$

Hence by direct calculation we have that the probability of error is

$$\begin{aligned} &\sum_y \frac{D_0(y) + D_1(y)}{2} \frac{3D_0(y)^2D_1(y) + 3D_0(y)D_1(y)^2}{(D_0(y) + D_1(y))^3} \\ &= \frac{3}{2} \sum_y \frac{D_0(y)^2D_1(y)}{(D_0(y) + D_1(y))^2} + \frac{D_0(y)D_1(y)^2}{(D_0(y) + D_1(y))^2} \end{aligned}$$

Let us upper bound the first of these terms; the upper bound on the second term follows analogously by switching D_0 and D_1 .

Since D_0 and D_1 are $1 - \epsilon$ -far in total variation distance, there must exist some set S of y 's, and its complement $\bar{S}$, such that $\sum_{y \in S} D_0(y) \geq 1 - \epsilon$ and $\sum_{y \in S} D_1(y) \leq \epsilon$, which implies that $\sum_{y \in \bar{S}} D_0(y) \leq \epsilon$ and $\sum_{y \in \bar{S}} D_1(y) \geq 1 - \epsilon$. Hence we have that

$$\begin{aligned} \sum_y \frac{D_0(y)^2D_1(y)}{(D_0(y) + D_1(y))^2} &= \sum_{y \in S} \frac{D_0(y)^2D_1(y)}{(D_0(y) + D_1(y))^2} \\ &\quad + \sum_{y \in \bar{S}} \frac{D_0(y)^2D_1(y)}{(D_0(y) + D_1(y))^2} \\ &\leq \sum_{y \in S} \frac{D_0(y)^2D_1(y)}{(D_0(y))^2} \\ &\quad + \sum_{y \in \bar{S}} \frac{D_0(y)^2D_1(y)}{(D_1(y))^2} \\ &\leq \sum_{y \in S} D_1(y) + \sum_{y \in \bar{S}} D_0(y) \\ &\leq 2\epsilon \end{aligned}$$

By applying an analogous bound to the second term, we have that $\Pr[\text{error}] = O(\epsilon) = O(2^{-n^c})$ as desired, so the probability of error in this case is vanishingly small.

Hence the net probability that the algorithm errs is $\frac{1}{4} + \epsilon$ in the case the distributions are ϵ -close and $O(\epsilon)$ in the case the distributions are $(1 - \epsilon)$ -far.

C. AN $N^{1/4}$ LOWER BOUND FOR SEARCH IN NACQP

Here we show that any naCQP algorithm for search requires at least $N^{1/4}$ time.

PROOF OF THEOREM 6.1. Since it is always possible to copy measured qubits, we can assume that qubits which are measured in an intermediate step of the algorithm are never directly modified again. Now, assume that the algorithm uses ℓ qubits and applies unitary operators $U_1, \dots, U_T$, each of which is either a (controlled) query to the search function f or a gate from the finite universal gate set $\mathcal{U}$. The measurements $M_1 \dots M_T$ (which may or may not be empty) are applied between the operators $U_1 \dots U_T$.

Let $v(x) = (v_0(x), v_1(x), \dots, v_T(x))$ be the non-collapsing measurement results when the marked item is x , so that $v_i(x)$ is sampled immediately before the application of U_{i+1} . Let $v = (v_0, \dots, v_T)$ be the non-collapsing measurement results when there is no marked item. In general, both $v(x)$ and v are random variables. Since the postprocessing step can distinguish the distributions of v and $v(x)$ with success probability $2/3$, $d_{TV}(v, v(x)) \geq \frac{1}{3}$ for all x . On the other hand, each v and $v(x)$ is a Markov process. Therefore, by Lemma 6.1,

$$d_{TV}(v, v(x)) \leq 2 \sum_{i=1}^T d_{TV}((v_{i-1}, v_i), (v_{i-1}(x), v_i(x))).$$

Now, we bound the term

$$d_{x,i} := d_{TV}((v_{i-1}, v_i), (v_{i-1}(x), v_i(x))).$$

Since it is possible to defer measurements in a quantum circuit to a later stage [16], we can assume that all intermediate measurements that occurred before the application of U_i occurred immediately before the sampling of v_i . Suppose that these measurements were applied to the first k qubits of the state. Let $|\phi\rangle$ and $|\phi(x)\rangle$ be the state vectors immediately before these measurements. Then, we decompose $|\phi\rangle = \sum_{s \in \{0,1\}^k} \alpha_s |s\rangle |\phi_s\rangle$ and $|\phi(x)\rangle = \sum_{s \in \{0,1\}^k} \beta_s |s\rangle |\phi_s(x)\rangle$. Possible values for (v_{i-1}, v_i) and $(v_{i-1}(x), v_i(x))$ can be written in the form (st_1, st_2) , where s is a k -bit string and t_1, t_2 are $(\ell - k)$ -bit strings.

Assume for now that U_i does not contain a query to f . Then, since it does not affect the first k qubits, it can be decomposed into the sum $\sum_{s \in \{0,1\}^k} |s\rangle V_s \langle s|$ for some unitary operators V_s . The transformation U_i can be thought of as applying the unitary V_s to the last $\ell - k$ qubits if the (measured) first k qubits are equal to s . Then, the probability that $(v_{i-1}, v_i) = (st_1, st_2)$ is equal to

$$|\alpha_s|^2 |\langle t_1 | \phi_s \rangle|^2 |\langle t_2 | V_s | \phi_s \rangle|^2,$$

and the probability that $(v_{i-1}(x), v_i(x)) = (st_1, st_2)$ is equal to $|\beta_s|^2 |\langle t_1 | \phi_s(x) \rangle|^2 |\langle t_2 | V_s | \phi_s(x) \rangle|^2$. Therefore, the total variation distance $d_{x,i}$ is by the triangle inequality

$$\begin{aligned} d_{x,i} &= \frac{1}{2} \sum_{s, t_1, t_2} \left| |\alpha_s|^2 |\langle t_1 | \phi_s \rangle|^2 |\langle t_2 | V_s | \phi_s \rangle|^2 \right. \\ &\quad \left. - |\beta_s|^2 |\langle t_1 | \phi_s(x) \rangle|^2 |\langle t_2 | V_s | \phi_s(x) \rangle|^2 \right| \\ &\leq \frac{1}{2} \sum_{s, t_1, t_2} \left(\left| |\alpha_s|^2 |\langle t_1 | \phi_s(x) \rangle|^2 |\langle t_2 | V_s | \phi_s(x) \rangle|^2 \right. \right. \\ &\quad \left. \left. - |\beta_s|^2 |\langle t_1 | \phi_s(x) \rangle|^2 |\langle t_2 | V_s | \phi_s(x) \rangle|^2 \right| \right) \end{aligned}$$

$$\begin{aligned} &+ \frac{1}{2} \sum_{s, t_1, t_2} \left(|\alpha_s|^2 |\langle t_1 | \phi_s \rangle|^2 |\langle t_2 | V_s | \phi_s \rangle|^2 \right. \\ &\quad \left. - |\langle t_1 | \phi_s(x) \rangle|^2 |\langle t_2 | V_s | \phi_s \rangle|^2 \right) \\ &+ \frac{1}{2} \sum_{s, t_1, t_2} \left(|\alpha_s|^2 |\langle t_1 | \phi_s(x) \rangle|^2 |\langle t_2 | V_s | \phi_s \rangle|^2 \right. \\ &\quad \left. - |\langle t_1 | \phi_s(x) \rangle|^2 |\langle t_2 | V_s | \phi_s(x) \rangle|^2 \right) \\ &=: \frac{1}{2} (S_1 + S_2 + S_3) \end{aligned}$$

where S_1, S_2, S_3 are the three sums written above, which range over $s \in \{0,1\}^k$ and $t_1, t_2 \in \{0,1\}^{\ell-k}$. Now, we have:

$$\begin{aligned} S_1 &:= \sum_{s, t_1, t_2} \left(|\alpha_s|^2 |\langle t_1 | \phi_s(x) \rangle|^2 |\langle t_2 | V_s | \phi_s(x) \rangle|^2 \right. \\ &\quad \left. - |\beta_s|^2 |\langle t_1 | \phi_s(x) \rangle|^2 |\langle t_2 | V_s | \phi_s(x) \rangle|^2 \right) \\ &= \sum_s \left(|\alpha_s|^2 - |\beta_s|^2 \right) \left(\sum_{t_1, t_2} |\langle t_1 | \phi_s(x) \rangle|^2 |\langle t_2 | V_s | \phi_s(x) \rangle|^2 \right) \\ &= \sum_s \left(|\alpha_s|^2 - |\beta_s|^2 \right) \\ &\leq \| |\phi\rangle \langle \phi| - |\phi(x)\rangle \langle \phi(x)| \|_{tr} \\ &\leq 2 \| |\phi(x)\rangle - |\phi\rangle \|_2. \end{aligned}$$

Additionally,

$$\begin{aligned} S_2 &:= \sum_{s, t_1, t_2} \left(|\alpha_s|^2 |\langle t_1 | \phi_s \rangle|^2 |\langle t_2 | V_s | \phi_s \rangle|^2 \right. \\ &\quad \left. - |\langle t_1 | \phi_s(x) \rangle|^2 |\langle t_2 | V_s | \phi_s \rangle|^2 \right) \\ &= \sum_{s, t_1} \left(|\alpha_s|^2 |\langle t_1 | \phi_s \rangle|^2 - |\langle t_1 | \phi_s(x) \rangle|^2 \right) \\ &\leq \sum_{s, t_1} \left(\left| |\alpha_s|^2 |\langle t_1 | \phi_s \rangle|^2 - |\beta_s|^2 |\langle t_1 | \phi_s(x) \rangle|^2 \right| \right. \\ &\quad \left. + \sum_{t_2} \left(|\alpha_s|^2 |\langle t_1 | \phi_s(x) \rangle|^2 |\langle t_2 | V_s | \phi_s \rangle|^2 - |\beta_s|^2 |\langle t_1 | \phi_s(x) \rangle|^2 |\langle t_2 | V_s | \phi_s(x) \rangle|^2 \right) \right) \\ &= \sum_{s, t_1} \left(\left| |\alpha_s|^2 |\langle t_1 | \phi_s \rangle|^2 - |\beta_s|^2 |\langle t_1 | \phi_s(x) \rangle|^2 \right| \right. \\ &\quad \left. + \sum_s \left(|\alpha_s|^2 - |\beta_s|^2 \right) \right) \\ &\leq 2 \| |\phi\rangle \langle \phi| - |\phi(x)\rangle \langle \phi(x)| \|_{tr} \\ &\leq 4 \| |\phi(x)\rangle - |\phi\rangle \|_2. \end{aligned}$$

Finally,

$$\begin{aligned} S_3 &= \sum_{s, t_1, t_2} \left(|\alpha_s|^2 |\langle t_1 | \phi_s(x) \rangle|^2 |\langle t_2 | V_s | \phi_s \rangle|^2 \right. \\ &\quad \left. - |\langle t_1 | \phi_s(x) \rangle|^2 |\langle t_2 | V_s | \phi_s(x) \rangle|^2 \right) \\ &= \sum_{s, t_2} \left(|\alpha_s|^2 |\langle t_2 | V_s | \phi_s \rangle|^2 - |\langle t_2 | V_s | \phi_s(x) \rangle|^2 \right) \\ &\leq \sum_{s, t_2} \left(\left| |\alpha_s|^2 |\langle t_2 | V_s | \phi_s \rangle|^2 - |\beta_s|^2 |\langle t_2 | V_s | \phi_s(x) \rangle|^2 \right| \right. \\ &\quad \left. + \sum_{t_1} \left(|\alpha_s|^2 |\langle t_2 | V_s | \phi_s(x) \rangle|^2 |\langle t_1 | \phi_s \rangle|^2 - |\beta_s|^2 |\langle t_2 | V_s | \phi_s(x) \rangle|^2 |\langle t_1 | \phi_s(x) \rangle|^2 \right) \right) \end{aligned}$$

$$\begin{aligned}
&= \sum_{s,t_2} (|\alpha_s|^2 |\langle t_2 | V_s | \phi_s \rangle|^2 - |\beta_s|^2 |\langle t_2 | V_s | \phi_s(x) \rangle|^2) \\
&\quad + \sum_s (|\alpha_s|^2 - |\beta_s|^2) \\
&\leq 2 \|\phi\| \langle \phi | - \phi(x) \rangle \langle \phi(x) | \|_{tr} \\
&= 4 \|\phi(x) - \phi\|_2
\end{aligned}$$

Therefore,

$$d_{x,i} \leq \frac{1}{2}(S_1 + S_2 + S_3) \leq 5 \|\phi(x) - \phi\|_2.$$

On the other hand, if U_i is a query to f , then it only applies a local phase of -1 to some of the probability amplitudes of $|\phi\rangle$ and $|\phi_x\rangle$. Therefore, the same argument still shows that $d_{x,i} \leq 5 \|\phi(x) - \phi\|_2$.

By the Cauchy-Schwarz inequality and Lemma 6.3,

$$\begin{aligned}
\frac{1}{N} \sum_{x=0}^{N-1} d_{x,i} &\leq 5 \cdot \frac{1}{N} \sum_{x=0}^{N-1} \|\phi(x) - \phi\|_2 \\
&\leq 5 \sqrt{\frac{1}{N} \sum_{x=0}^{N-1} \|\phi(x) - \phi\|_2^2} \\
&\leq \frac{10Q}{\sqrt{N}}
\end{aligned}$$

for all i . Therefore, there is some x for which

$$d_{TV}(v, v(x)) \leq 2 \sum_{i=1}^T d_{x,i} \leq \frac{20TQ}{\sqrt{N}}.$$

On the other hand, $d_{TV}(v, v(x)) \geq \frac{1}{3}$ for all x , so

$$\frac{20TQ}{\sqrt{N}} \geq \frac{1}{3},$$

and the running time of the algorithm is at least $T + Q = \Omega(N^{1/4})$. $\square$

D. AN $N^{1/3}$ LOWER BOUND FOR SEARCH IN NACQP IF THERE ARE NO COLLAPSING MEASUREMENTS

Assume that intermediate measurements are not allowed in our search algorithm. As we discuss in Section 3, this gives a model with only the power of BQP, because then the states $|\psi_t\rangle = U_t U_{t-1} \dots U_1 |0\rangle^{\otimes n}$ can be generated with poly-sized circuits, and hence a BQP machine could prepare and measure them to sample from $\mathcal{Q}$. Trivially one can prove a lower bound of $N^{1/4}$ for search in this model, either by noting that this class can achieve at most quadratic speedups over BQP by the previous comment, or by using the argument put forth in Theorem 6.1. Here we tighten this result to give an $N^{1/3}$ lower bound for search in this class.

Suppose that an algorithm A searches with Q queries and T timesteps, where $Q+T = o(N^{1/3})$. Let ψ_t be the quantum state after t steps with no marked item, and let ψ_t^x be defined likewise when the marked item is at location x .

By the hybrid argument we have that $\forall t$

$$\sum_x \|\psi_t - \psi_t^x\|_2^2 \leq 4Q^2$$

where $\|a\|_2^2$ is the 2-norm squared of a . This implies

$$\sum_t \sum_x \|\psi_t - \psi_t^x\|_2^2 \leq 4TQ^2$$

Hence there must exist x such that

$$\sum_t \|\psi_t - \psi_t^x\|_2^2 \leq \frac{4TQ^2}{N} \quad (2)$$

Since we assumed $Q + T = o(N^{1/3})$, we have that $\frac{4TQ^2}{N} = o(1)$. Therefore for sufficiently large N and for all t we have

$$\|\psi_t - \psi_t^x\|_2^2 \leq 0.01$$

(The choice of constant here is arbitrary, we simply need it to be less than around 0.5.) Now consider the states $\Psi := \bigotimes_t |\psi_t\rangle$ and $\Psi^x := \bigotimes_t |\psi_t^x\rangle$. Let V the distribution on samples with no marked item, and let V^x be defined likewise. Then clearly we have that

$$|V - V^x|_1 \leq \|\Psi - \Psi^x\|$$

where $\|a\|$ denotes the trace norm of a . This is because the output distributions of V and V^x can be obtained by (independent) measurements on the states Ψ and Ψ^x in the computational basis. Note that $|V - V^x|_1$ must be $\Omega(1)$ in order to distinguish the presence of a marked item at x in postprocessing. Therefore we have

$$\Omega(1) \leq |V - V^x|_1 \leq \|\Psi - \Psi^x\| \quad (3)$$

$$= \sqrt{1 - |\langle \Psi | \Psi^x \rangle|^2} \quad (4)$$

$$= \sqrt{1 - |\Pi_t \langle \psi_t | \psi_t^x \rangle|^2} \quad (5)$$

$$\leq \sqrt{1 - \Pi_t e^{-\|\psi_t - \psi_t^x\|_2^2}} \quad (6)$$

$$= \sqrt{1 - e^{-\Sigma_t \|\psi_t - \psi_t^x\|_2^2}} \quad (7)$$

$$\leq \sqrt{1 - e^{-\frac{4TQ^2}{N}}} \quad (8)$$

$$= o(1) \quad (9)$$

Where in line 4 we use the formula for trace distance of pure states, in line 8 we used equation 2, in line 9 we used the fact that $T + Q = o(N^{1/3})$, and in line 6 we use the inequality

$$|\langle \psi_t | \psi_t^x \rangle| \geq \text{Re}(\langle \psi_t | \psi_t^x \rangle) \quad (10)$$

$$= 1 - \frac{\|\psi_t - \psi_t^x\|_2^2}{2} \quad (11)$$

$$\geq e^{-\|\psi_t - \psi_t^x\|_2^2} \quad (12)$$

where we have use the fact that $1 - x \geq e^{-2x}$ for $0 \leq x \leq 0.01$. Therefore we have shown $\Omega(1) = o(1)$, a contradiction. Hence such an algorithm A cannot exist, so searching takes $Q + T = \Omega(N^{1/3})$ time when there are non-collapsing measurements, but no collapsing measurements, in the model.