Limitations of Quantum Advice and One-Way Communication

Scott Aaronson*
University of California, Berkeley
aaronson@cs.berkeley.edu

Abstract

Although a quantum state requires exponentially many classical bits to describe, the laws of quantum mechanics impose severe restrictions on how that state can be accessed. This paper shows in three settings that quantum messages have only limited advantages over classical ones.

First, we show that $\mathsf{BQP/qpoly} \subseteq \mathsf{PP/poly}$*, where* $\mathsf{BQP/qpoly}$ *is the class of problems solvable in quantum polynomial time, given a polynomial-size "quantum advice state" that depends only on the input length. This resolves a question of Buhrman, and means that we should not hope for an unrelativized separation between quantum and classical advice. Underlying our complexity result is a general new relation between deterministic and quantum one-way communication complexities, which applies to partial as well as total functions.*

Second, we construct an oracle relative to which $\mathsf{NP} \not\subset \mathsf{BQP/qpoly}$*. To do so, we use the polynomial method to give the first correct proof of a direct product theorem for quantum search. This theorem has many other applications; for example, it can be used to fix and even improve on a flawed result of Klauck about quantum time-space tradeoffs for sorting.*

Third, we introduce a new trace distance method for proving lower bounds on quantum one-way communication complexity. Using this method, we obtain optimal quantum lower bounds for two problems of Ambainis, for which no nontrivial lower bounds were previously known even for classical randomized protocols.

1. Introduction

How many classical bits can "really" be encoded into n qubits? Is it n, because of Holevo's Theorem [17]; $2n$, because of dense quantum coding [10] and quantum teleportation [8]; exponentially many, because of quantum fingerprinting [11]; or infinitely many, because amplitudes are continuous? The best general answer to this question is probably *mu*, the Zen word that "unasks" a question.[1]

To a computer scientist, however, it is natural to formalize the question in terms of *quantum one-way communication complexity* [19, 5, 11, 32]. The setting is as follows: Alice has an n-bit string x, Bob has an m-bit string y, and together they wish to evaluate $f(x, y)$ where $f : \{0,1\}^n \times \{0,1\}^m \to \{0,1\}$ is a Boolean function. After examining her input $x = x_1 \ldots x_n$, Alice can send a single quantum message ρ_x to Bob, whereupon Bob, after examining his input $y = y_1 \ldots y_m$, can choose some basis in which to measure ρ_x. He must then output a claimed value for $f(x, y)$. We are interested in how long Alice's message needs to be, for Bob to succeed with high probability on any x, y pair. Ideally the length will be much smaller than if Alice had to send a classical message.

Communication complexity questions have been intensively studied in theoretical computer science (see the book of Kushilevitz and Nisan [22] for example). In both the classical and quantum cases, though, most attention has focused on *two-way* communication, meaning that Alice and Bob get to send messages back and forth. We believe that the study of one-way quantum communication presents two main advantages. First, many open problems about two-way communication look gruesomely difficult—for example, are the randomized and quantum communication complexities of every total Boolean function polynomially related? We might gain insight into these problems by tackling their one-way analogues first. And second, because of its greater simplicity, the one-way model more directly addresses our opening question: how much "useful stuff" can be packed into a quantum state? Thus, results

* Supported by an NSF Graduate Fellowship, by NSF ITR Grant CCR-0121555, and by the Defense Advanced Research Projects Agency (DARPA).

1 Another *mu*-worthy question is, "Where does the power of quantum computing come from? Superposition? Interference? Entanglement? The large size of Hilbert space?"

on one-way communication fall into the quantum information theory tradition initiated by Holevo [17] and others, as much as the communication complexity tradition initiated by Yao [33].

Related to quantum one-way communication is the notion of *quantum advice*. As pointed out by Nielsen and Chuang [24, p.203], there is no compelling physical reason to assume that the starting state of a quantum computer is a computational basis state:[2]

> [W]e know that many systems in Nature 'prefer' to sit in highly entangled states of many systems; might it be possible to exploit this preference to obtain extra computational power? It might be that having access to certain states allows particular computations to be done much more easily than if we are constrained to start in the computational basis.

One way to interpret Nielsen and Chuang's provocative question is as follows. Suppose we could request the *best possible* starting state for a quantum computer, knowing the language to be decided and the input length n but not knowing the input itself.[3] In analogy to the class $\mathsf{P/poly}$ defined by Karp and Lipton [18], denote the class of languages that we could then decide by $\mathsf{BQP/qpoly}$—meaning quantum polynomial time, given an arbitrarily-entangled but polynomial-size quantum advice state.[4] How powerful is this class? Were $\mathsf{BQP/qpoly}$ vastly larger than BQP, we would need to rethink our most basic assumptions about the power of quantum computing. We will see later that quantum advice is closely related to quantum one-way communication, since we can think of an advice state as a one-way message sent to an algorithm by a benevolent "advisor."

This paper is about the *limitations* of quantum advice and one-way communication. It presents three contributions which are basically independent of one another.

First, in Section 3, we show that $D^1(f) = O\left(mQ_2^1(f)\log Q_2^1(f)\right)$ for any Boolean function f, partial or total. Here $D^1(f)$ is deterministic one-way communication complexity, $Q_2^1(f)$ is bounded-error one-way quantum communication complexity, and m is the length of Bob's input. Intuitively, whenever the set of Bob's possible inputs is not too large, Alice can send him a short classical message that lets him learn the outcome of any measurement he would have wanted to make on the quantum message ρ_x. It is interesting that a slightly tighter bound for total functions—$D^1(f) = O\left(mQ_2^1(f)\right)$—follows easily from a result of Klauck [19] together with a lemma of Sauer [30] about VC-dimension. However, the proof of the latter bound is highly nonconstructive, and seems to fail for partial f.

Using our communication complexity result, in Section 3.1 we show that $\mathsf{BQP/qpoly} \subseteq \mathsf{PP/poly}$, resolving a question raised by Buhrman (personal communication). A corollary is that we cannot hope to show an unrelativized separation between quantum and classical advice (that is, that $\mathsf{BQP/poly} \neq \mathsf{BQP/qpoly}$), without also showing that PP does not have polynomial-size circuits.

What makes this result surprising is that, in the minds of many computer scientists, a quantum state is basically an exponentially long vector. Indeed, this belief seems to fuel skepticism of quantum computing (see Goldreich [15] for example). But given an exponentially long advice string, even a classical computer could decide any language whatsoever. So one might imagine naïvely that $\mathsf{BQP/qpoly} \not\subset \mathsf{r.e./poly}$—that is, that quantum advice would let us solve problems that are not even recursively enumerable given classical advice of a similar size! The failure of this naïve intuition supports the view that, although a quantum state takes exponentially many classical bits to specify, it is much more akin to a classical probability distribution than to an exponentially long random-access tape.

Our second contribution, in Section 4, is an oracle relative to which NP is not contained in $\mathsf{BQP/qpoly}$. Underlying this oracle separation is the first correct proof of a *direct product theorem* for quantum search. Given an N-item database with K marked items, the direct product theorem says that if a quantum algorithm makes $o\left(\sqrt{N}\right)$ queries, then the probability that the algorithm finds all K of the marked items decreases exponentially in K.[5] Notice that such a result does not follow from any existing quantum lower bound. Earlier Klauck [20] had claimed a weaker direct product theorem, based on the hybrid method of Bennett et al. [7]. Unfortunately, Klauck's proof is incorrect. Our proof uses the polynomial method of Beals et al. [6]

2 One might object that the starting state is itself the outcome of some computational process, which began no earlier than the Big Bang. However, (1) for all we know highly entangled states were created in the Big Bang, and (2) 14 billion years is a long time.

3 If we knew the input, we would simply request a starting state that contains the right answer!

4 There are two key differences between $\mathsf{BQP/qpoly}$ and the better-known class QMA (Quantum Merlin-Arthur): first, advice can be trusted while proofs cannot; second, proofs can be tailored to a particular input while advice cannot.

5 The actual result is stronger than this, and was recently improved even further by Klauck, Špalek, and de Wolf [21].

in a novel way that we expect will find other applications.

Our final contribution, in Section 5, is a new *trace distance* method for proving lower bounds on quantum one-way communication complexity. Previously there was only one basic lower bound technique: the VC-dimension method of Klauck [19], which relied on lower bounds for quantum random access codes due to Ambainis et al. [4] and Nayak [23]. Using VC-dimension one can show, for example, that $Q_2^1(\mathrm{DISJ}) = \Omega(n)$, where the *disjointness function* $\mathrm{DISJ} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is defined by $\mathrm{DISJ}(x, y) = 1$ if and only if $x_i y_i = 0$ for all $i \in \{1, \ldots, n\}$.

For some problems, however, the VC-dimension method yields no nontrivial quantum lower bound. Seeking to make this point vividly, Ambainis posed the following problem. Alice is given two elements x, y of a finite field $\mathbb{F}_p$ (where p is prime); Bob is given another two elements $a, b \in \mathbb{F}_p$. Bob's goal is to output 1 if $y \equiv ax + b \pmod{p}$ and 0 otherwise. For this problem, the VC-dimension method yields no randomized *or* quantum lower bound better than constant. On the other hand, the well-known fingerprinting protocol for the equality function seems to fail for Ambainis' problem, because of the interplay between addition and multiplication. So it is natural to conjecture that the randomized and even quantum one-way complexities are $\Theta(\log p)$—that is, that no nontrivial protocol exists for this problem.

Ambainis posed a second problem in the same spirit. Here Alice is given $x \in \{1, \ldots, N\}$, Bob is given $y \in \{1, \ldots, N\}$, and both players know a subset $S \subset \{1, \ldots, N\}$. Bob's goal is to decide whether $x - y \in S$ where subtraction is modulo N. The conjecture is that if S is chosen uniformly at random with $|S|$ about $\sqrt{N}$, then with high probability the randomized and quantum one-way complexities are both $\Theta(\log N)$.

Using our trace distance method, we are able to show optimal quantum lower bounds for both of Ambainis' problems. Previously, no nontrivial lower bounds were known even for randomized protocols. The key idea is to consider two probability distributions over Alice's quantum message ρ_x. The first distribution corresponds to x chosen uniformly at random; the second corresponds to x chosen uniformly conditioned on $f(x, y) = 1$. These distributions give rise to two mixed states ρ and ρ_y, which Bob must be able to distinguish with non-negligible bias assuming he can evaluate $f(x, y)$. We then show an upper bound on the trace distance $\|\rho - \rho_y\|_{\mathrm{tr}}$, which implies that Bob cannot distinguish the distributions.

Theorem 12 gives a very general condition under which our trace distance method works; Corollaries 13 and 14 then show that the condition is satisfied for Ambainis' two problems. Besides showing that the VC-dimension method is not optimal, we hope our method is a significant step towards proving that $R_2^1(f) = O\left(Q_2^1(f)\right)$ for all total Boolean functions f, where $R_2^1(f)$ is randomized one-way complexity.

We conclude in Section 6 with some open problems.

2. Preliminaries

This section reviews basic definitions and results about quantum one-way communication (in Section 2.1) and quantum advice (in Section 2.2); then Section 2.3 proves a quantum information lemma that will be used throughout the paper.

2.1. Quantum One-Way Communication

Following standard conventions, we denote by $D^1(f)$ the deterministic one-way complexity of f, or the minimum number of bits that Alice must send if her message is a function of x. Also, $R_2^1(f)$, the bounded-error randomized one-way complexity, is the minimum k such that for every x, y, if Alice sends Bob a k-bit message drawn from some distribution $\mathcal{D}_x$, then Bob can output a bit a such that $a = f(x, y)$ with probability at least $2/3$. (The subscript 2 means that the error is two-sided.) The zero-error randomized complexity $R_0^1(f)$ is similar, except that Bob's answer can never be wrong: he must output $f(x, y)$ with probability at least $1/2$ and otherwise declare failure.

The bounded-error quantum one-way complexity $Q_2^1(f)$ is the minimum k such that, if Alice sends Bob a mixed state ρ_x of k qubits, there exists a joint measurement of ρ_x and y enabling Bob to output an a such that $a = f(x, y)$ with probability at least $2/3$. The zero-error and exact complexities $Q_0^1(f)$ and $Q_E^1(f)$ are defined analogously.[6] See Klauck [19] for more detailed definitions of classical and quantum one-way complexity measures.

It is immediate that $D^1(f) \geq R_0^1(f) \geq R_2^1(f) \geq Q_2^1(f)$, that $R_0^1(f) \geq Q_0^1(f) \geq Q_2^1(f)$, and that $D^1(f) \geq Q_E^1(f)$. Also, for total f, Duriš et al. [12] showed that $R_0^1(f) = \Theta\left(D^1(f)\right)$, while Klauck [19] showed that $Q_E^1(f) = D^1(f)$ and that $Q_0^1(f) =$

6 Requiring Alice's message to be a pure state would increase these complexities by at most a factor of 2, since by Kraus' Theorem, every k-qubit mixed state can be realized as half of a $2k$-qubit pure state. Winter [31] has shown that this factor of 2 is tight.

$\Theta\left(D^1(f)\right)$. In other words, randomized and quantum messages yield no improvement for total functions if we are unwilling to tolerate a bounded probability of error. This remains true even if Alice and Bob share arbitrarily many EPR pairs [19]. As is often the case, the situation is dramatically different for partial functions: there it is easy to see that $R_0^1(f)$ can be constant even though $D^1(f) = \Omega(n)$: let $f(x,y) = 1$ if $x_1y_1 + \cdots + x_{n/2}y_{n/2} \geq n/4$ and $x_{n/2+1}y_{n/2+1} + \cdots + x_ny_n = 0$ and $f(x,y) = 0$ if $x_1y_1 + \cdots + x_{n/2}y_{n/2} = 0$ and $x_{n/2+1}y_{n/2+1} + \cdots + x_ny_n \geq n/4$, promised that one of these is the case.

Moreover, Bar-Yossef, Jayram, and Kerenidis [5] have *almost* shown that $Q_E^1(f)$ can be exponentially smaller than $R_2^1(f)$. In particular, they proved that separation for a *relation*, meaning a problem for which Bob has many possible valid outputs. For a partial function f based on their relation, they also showed that $Q_E^1(f) = \Theta(\log n)$ whereas $R_0^1(f) = \Theta(\sqrt{n})$; and they conjectured (but did not prove) that $R_2^1(f) = \Theta(\sqrt{n})$.

2.2. Quantum Advice

Informally, BQP/qpoly is the class of languages decidable in polynomial time on a quantum computer, given a polynomial-size quantum advice state that depends only on the input length. We now make the definition more formal.

Definition 1 *A language L is in* BQP/qpoly *if there exists a polynomial-size quantum circuit family $\{C_n\}_{n\geq1}$, and a polynomial-size family of quantum states $\{|\psi_n\rangle\}_{n\geq1}$, such that for all $x \in \{0,1\}^n$,*

(i) If $x \in L$ then $q(x) \geq 2/3$, where $q(x)$ is the probability that the first qubit is measured to be $|1\rangle$, after C_n is applied to the starting state $|x\rangle \otimes |0\cdots0\rangle \otimes |\psi_n\rangle$.

(ii) If $x \notin L$ then $q(x) \leq 1/3$.[7]

Nishimura and Yamakami [26] showed that EESPACE $\not\subset$ BQP/qpoly; besides that, essentially nothing was known about BQP/qpoly before the present work.

2.3. The Almost As Good As New Lemma

The following simple lemma, which was implicit in [4], is used three times in this paper—in Theorems 6, 7, and 11. It says that, if the outcome of measuring a quantum state ρ could be predicted with near-certainty given knowledge of ρ, then measuring ρ will damage it only slightly. Recall that the trace distance $\|\rho - \sigma\|_{\text{tr}}$ between two mixed states ρ and σ equals $\frac{1}{2}\sum_i |\lambda_i|$, where $\lambda_1, \ldots, \lambda_N$ are the eigenvalues of $\rho - \sigma$.

Lemma 2 *Suppose a 2-outcome measurement of a mixed state ρ yields outcome 0 with probability $1 - \varepsilon$. Then after the measurement, we can recover a state $\widetilde{\rho}$ such that $\|\widetilde{\rho} - \rho\|_{\text{tr}} \leq \sqrt{\varepsilon}$. This is true even if the measurement is a POVM (that is, involves arbitrarily many ancilla qubits).*

Proof. Let $|\psi\rangle$ be a purification of the entire system (ρ plus ancilla). We can represent any measurement as a unitary U applied to $|\psi\rangle$, followed by a 1-qubit measurement. Let $|\varphi_0\rangle$ and $|\varphi_1\rangle$ be the two possible pure states after the measurement; then $\langle\varphi_0|\varphi_1\rangle = 0$ and $U|\psi\rangle = \alpha|\varphi_0\rangle + \beta|\varphi_1\rangle$ for some α, β such that $|\alpha|^2 = 1 - \varepsilon$ and $|\beta|^2 = \varepsilon$. Writing the measurement result as $\sigma = (1-\varepsilon)|\varphi_0\rangle\langle\varphi_0| + \varepsilon|\varphi_1\rangle\langle\varphi_1|$, it is easy to show that

$$\left\|\sigma - U|\psi\rangle\langle\psi|U^{-1}\right\|_{\text{tr}} = \sqrt{\varepsilon(1-\varepsilon)}.$$

So applying U^{-1} to σ,

$$\left\|U^{-1}\sigma U - |\psi\rangle\langle\psi|\right\|_{\text{tr}} = \sqrt{\varepsilon(1-\varepsilon)}.$$

Let $\widetilde{\rho}$ be the restriction of $U^{-1}\sigma U$ to the original qubits of ρ. Theorem 9.2 of Nielsen and Chuang [24] shows that tracing out a subsystem never increases trace distance, so $\|\widetilde{\rho} - \rho\|_{\text{tr}} \leq \sqrt{\varepsilon(1-\varepsilon)} \leq \sqrt{\varepsilon}$. ■

3. Simulating Quantum Messages

Let $f : \{0,1\}^n \times \{0,1\}^m \to \{0,1\}$ be a Boolean function. In this section we first combine existing results to obtain the relation $D^1(f) = O\left(mQ_2^1(f)\right)$ for total f, and then prove using a new method that $D^1(f) = O\left(mQ_2^1(f)\log Q_2^1(f)\right)$ for all f (partial or total).

Define the *communication matrix* M_f to be a $2^n \times 2^m$ matrix with $f(x,y)$ in the x^{th} row and y^{th} column. Then letting $\text{rows}(f)$ be the number of distinct rows in M_f, the following is immediate.

Proposition 3 *For total f,*

$$D^1(f) = \Theta(\log \text{rows}(f)),$$
$$Q_2^1(f) = \Omega(\log\log \text{rows}(f)).$$

Also, let the VC-dimension $\text{VC}(f)$ equal the maximum k for which there exists a $2^n \times k$ submatrix M_g of M_f with $\text{rows}(g) = 2^k$. Then Klauck [19] observed

7 If the starting state is $|x\rangle \otimes |0\cdots0\rangle \otimes |\varphi\rangle$ for some $|\varphi\rangle \neq |\psi_n\rangle$, then we do not require the acceptance probability to lie in $[0,1/3] \cup [2/3,1]$. Therefore, what we call BQP/qpoly corresponds to what Nishimura and Yamakami [26] call BQP/*Qpoly. Also, it does not matter whether the circuit family $\{C_n\}_{n\geq1}$ is uniform, since we are giving it advice anyway.

the following, based on a lower bound for quantum random access codes due to Nayak [23].

Proposition 4 *$Q_2^1(f) = \Omega(\mathrm{VC}(f))$ for total f.*

Now let $\mathrm{cols}(f)$ be the number of distinct columns in M_f. Then Proposition 4 yields the following general lower bound:

Corollary 5 *$D^1(f) = O\left(mQ_2^1(f)\right)$ for total f, where m is the size of Bob's input.*

Proof. It follows from a lemma of Sauer [30] that

$$\mathrm{rows}(f) \leq \sum_{i=0}^{\mathrm{VC}(f)} \binom{\mathrm{cols}(f)}{i} \leq \mathrm{cols}(f)^{\mathrm{VC}(f)+1}.$$

Hence $\mathrm{VC}(f) \geq \log_{\mathrm{cols}(f)} \mathrm{rows}(f) - 1$, so

$$Q_2^1(f) = \Omega(\mathrm{VC}(f)) = \Omega\left(\frac{\log \mathrm{rows}(f)}{\log \mathrm{cols}(f)}\right) = \Omega\left(\frac{D^1(f)}{m}\right).$$

■

In particular, $D^1(f)$ and $Q_2^1(f)$ are polynomially related for total f, provided that $m = O(n^c)$ for some $c < 1$ and Alice's input is not "padded" (that is, all rows of M_f are distinct).

We now give a new method for replacing quantum messages by classical ones when Bob's input is small. Although the best bound we know how to obtain with this method—$D^1(f) = O\left(mQ_2^1(f)\log Q_2^1(f)\right)$—is slightly weaker than the $D^1(f) = O\left(mQ_2^1(f)\right)$ of Corollary 5, our method works for *partial* Boolean functions as well as total ones. It also yields a (relatively) efficient procedure by which Bob can reconstruct Alice's quantum message, a fact we will exploit in Section 3.1 to show $\mathsf{BQP/qpoly} \subseteq \mathsf{PP/poly}$. By contrast, the method based on Sauer's Lemma seems to be nonconstructive.

Theorem 6 *$D^1(f) = O\left(mQ_2^1(f)\log Q_2^1(f)\right)$ for all f (partial or total).*

Proof. Let $f : \mathcal{D} \to \{0,1\}$ be a partial Boolean function with $\mathcal{D} \subseteq \{0,1\}^n \times \{0,1\}^m$, and for all $x \in \{0,1\}^n$, let $\mathcal{D}_x = \{y \in \{0,1\}^m : (x,y) \in \mathcal{D}\}$. Suppose Alice can send Bob a mixed state with $Q_2^1(f)$ qubits, that enables him to compute $f(x,y)$ for any $y \in \mathcal{D}_x$ with error probability at most $1/3$. Then she can also send him a boosted state ρ_x with $K = O\left(Q_2^1(f)\log Q_2^1(f)\right)$ qubits, such that for all $y \in \mathcal{D}_x$,

$$|P_y(\rho_x) - f(x,y)| \leq 1/Q_2^1(f)^{10},$$

where $P_y(\rho_x)$ is the probability that some measurement $\Lambda[y]$ yields a '1' outcome when applied to ρ_x. Furthermore, let $\mathcal{Y}$ be any subset of $\mathcal{D}_x$ satisfying $|\mathcal{Y}| \leq Q_2^1(f)^2$. Then starting with ρ_x, Bob can measure $\Lambda[y]$ for each $y \in \mathcal{Y}$ in lexicographic order, reusing the same message state again and again but uncomputing whatever garbage he generates while measuring. Since his probability of error on each y is at most $1/Q_2^1(f)^{10}$, by Lemma 2, after measuring ρ_x he obtains a new state $\widetilde{\rho}_x$ such that $\|\widetilde{\rho}_x - \rho_x\|_{\mathrm{tr}} \leq 1/Q_2^1(f)^5$. So even if he received a fresh copy of ρ_x after each query, the probability that he could tell the difference is at most $\sum_{i=1}^{Q_2^1(f)^2} i/Q_2^1(f)^5 = O\left(1/Q_2^1(f)\right)$, since trace distance satisfies the triangle inequality. So Bob will output $f(x,y)$ for every $y \in \mathcal{Y}$ simultaneously with probability at least (say) 0.9.

Now imagine that the communication channel is blocked, so Bob has to guess what message Alice wants to send him. He does this by using the K-qubit maximally mixed state I in place of ρ_x. We can write I as $\sum_{j=1}^{2^K} \sigma_j/2^K$, where $\sigma_1, \ldots, \sigma_{2^K}$ are orthonormal vectors such that $\sigma_1 = \rho_x$. So if Bob uses the same procedure as above except with I instead of ρ_x, then for any $\mathcal{Y} \subseteq \mathcal{D}_x$ with $|\mathcal{Y}| \leq Q_2^1(f)^2$, he will output $f(x,y)$ for every $y \in \mathcal{Y}$ simultaneously with probability at least $0.9/2^K$.

We now give the classical simulation of the quantum protocol. Alice's message to Bob consists of $T \leq K$ inputs $y_1, \ldots, y_T \in \mathcal{D}_x$, together with $f(x,y_1), \ldots, f(x,y_T)$.[8] Thus the message length is $mT + T = O\left(mQ_2^1(f)\log Q_2^1(f)\right)$. Here are the semantics of Alice's message: *"Bob, suppose you looped over all $y \in \mathcal{D}_x$ in lexicographic order; and for each one, guessed that $f(x,y) = \mathrm{round}(P_y(I))$, where $\mathrm{round}(p)$ is 1 if $p \geq 1/2$ and 0 if $p < 1/2$. Then y_1 is the first y for which you would guess the wrong value of $f(x,y)$. In general, denote by I_t the state obtained by starting from I and then measuring $\Lambda[y_1], \ldots, \Lambda[y_t]$ in that order, given that the outcomes of the measurements are $f(x,y_1), \ldots, f(x,y_t)$ respectively. If you looped over all $y \in \mathcal{D}_x$ in lexicographic order beginning from y_t, then y_{t+1} is the first y you would encounter for which $\mathrm{round}(P_y(I_t)) \neq f(x,y)$."*

Given the sequence of y_t's as defined above, it is obvious that Bob can compute $f(x,y)$ for any $y \in \mathcal{D}_x$. Let t^* be the largest t for which $y_t \leq y$ lexicographically. Then Bob simply needs to prepare the state I_{t^*}—which he can do since he knows $y_1, \ldots, y_{t^*}$ and $f(x,y_1), \ldots, f(x,y_{t^*})$—and then output $\mathrm{round}(P_y(I_{t^*}))$ as his claimed value of $f(x,y)$. Notice that, although Alice uses her knowledge of $\mathcal{D}_x$

8 Strictly speaking, Bob will be able to compute $f(x,y_1), \ldots, f(x,y_T)$ for himself given $y_1, \ldots, y_T$; he does not need Alice to tell him the f values.

to prepare her message, Bob does not need to know $\mathcal{D}_x$ in order to interpret the message. That is why the simulation works for partial as well as total functions.

But why we can assume that the sequence of y_t's stops at y_T for some $T \leq K$? Suppose $T > K$; we will derive a contradiction. Let $\mathcal{Y} = \{y_1, \ldots, y_{K+1}\}$. Then $|\mathcal{Y}| = K + 1 \leq Q_2^1(f)^2$, so we know from previous reasoning that if Bob starts with I and then measures $\Lambda[y_1], \ldots, \Lambda[y_{K+1}]$ in that order, he will observe $f(x, y_1), \ldots, f(x, y_{K+1})$ simultaneously with probability at least $0.9/2^K$. But by the definition of y_t, the probability that $\Lambda[y_t]$ yields the correct outcome is at most $1/2$, conditioned on $\Lambda[y_1], \ldots, \Lambda[y_{t-1}]$ having yielded the correct outcomes. Therefore $f(x, y_1), \ldots, f(x, y_{K+1})$ are observed simultaneously with probability at most $1/2^{K+1} < 0.9/2^K$, contradiction. ■

3.1. Simulating Quantum Advice

We now apply the techniques of Theorem 6 to upper-bound the power of quantum advice.

Theorem 7 $\mathsf{BQP/qpoly} \subseteq \mathsf{PP/poly}$.

Proof. Because of the close connection between advice and one-way communication, this theorem follows almost immediately from Theorem 6. Suppose the quantum advice states for a language $L \in \mathsf{BQP/qpoly}$ have $p(n)$ qubits for some polynomial p. Then by using a boosted advice state on $K = O(p(n) \log p(n))$ qubits, a polynomial-size quantum circuit C_n could compute the function $L_n(x)$, which is 1 if $x \in \{0,1\}^n$ is in L and 0 otherwise, with error probability at most $1/p(n)^{10}$. Now the classical advice to the PP simulation algorithm consists of $T \leq K$ inputs $x_1, \ldots, x_T \in \{0,1\}^n$, together with $L_n(x_1), \ldots, L_n(x_T)$. Let I be the maximally mixed state on K qubits, and let I_t be the state obtained by starting with I as the advice and then running C_n on $x_1, \ldots, x_t$ in that order (uncomputing garbage along the way), given that C_n correctly computed $L_n(x_1), \ldots, L_n(x_t)$. Then x_{t+1} is the lexicographically first input x after x_t for which $|P_x(I_t) - L_n(x)| \geq 1/2$, where $P_x(I_t)$ is the probability that C_n outputs '1' on input x given I_t as its advice.

The algorithm that exploits the classical advice to compute L_n, and the proof of its correctness, are the same as in Theorem 6. All we need to show is that the algorithm can be implemented in PP. This follows easily from the techniques used by Adleman, DeMarrais, and Huang [3] to show that $\mathsf{BQP} \subseteq \mathsf{PP}$. Let α_z be the amplitude of basis state $|z\rangle$. We simply simulate running C_n on $x_1, \ldots, x_t$ and then on x, and accept if and only if $S_1 > S_0$, where S_i is the sum of $|\alpha_z|^2$ over all $|z\rangle$ corresponding to the output sequence $L_n(x_1), \ldots, L_n(x_t), i$. A technicality is that, using finite-precision arithmetic, it might not be possible to compute S_1 and S_0 exactly. This causes no problems so long as the PP algorithm and the advice agree on a convention for when to declare $S_1 > S_0$. ■

We make four remarks about Theorem 7. First, for the same reason that Theorem 6 works for partial as well as total functions, we actually obtain the stronger result that $\mathsf{PromiseBQP/qpoly} \subseteq \mathsf{PP/poly}$.

Second, as pointed out to us by Fortnow, a corollary of Theorem 7 is that we cannot hope to show an unrelativized separation between $\mathsf{BQP/poly}$ and $\mathsf{BQP/qpoly}$, without also showing that PP does not have polynomial-size circuits. For $\mathsf{BQP/poly} \neq \mathsf{BQP/qpoly}$ clearly implies that $\mathsf{P/poly} \neq \mathsf{PP/poly}$. But the latter then implies that $\mathsf{PP} \not\subset \mathsf{P/poly}$, since assuming $\mathsf{PP} \subset \mathsf{P/poly}$ we could also obtain polynomial-size circuits for a language $L \in \mathsf{PP/poly}$ by defining a new language $L' \in \mathsf{PP}$, consisting of all (x, a) pairs such that the PP machine would accept x given advice string a. The reason this works is that PP is a syntactically defined class.

Third, an earlier version of this paper showed that $\mathsf{BQP/qpoly} \subseteq \mathsf{EXP/poly}$, by using a simulation in which an EXP algorithm keeps track of a subspace H of the advice Hilbert space to which the 'true' advice state must be close. In that simulation, the classical advice specifies inputs $x_1, \ldots, x_T$ for which $\dim(H)$ is at least halved; the observation that $\dim(H)$ must be at least 1 by the end then implies that $T \leq K = O(p(n) \log p(n))$, meaning that the advice is of polynomial size. The huge improvement from EXP to PP came solely from working with *measurement outcomes* and their *probabilities* instead of with *subspaces* and their *dimensions*. We can compute the former using the same 'Feynman path integral' that Adleman et al. [3] used to show $\mathsf{BQP} \subseteq \mathsf{PP}$, but could not see any way to compute the latter without explicitly storing and diagonalizing exponentially large matrices.

Fourth, assuming $\mathsf{BQP/poly} \neq \mathsf{BQP/qpoly}$, Theorem 7 is *almost* the best result of its kind that one could hope for, since the only classes known to lie between BQP and PP and not known to equal either are obscure ones such as AWPP [14]. Initially the theorem seemed to us to prove something stronger, namely that $\mathsf{BQP/qpoly} \subseteq \mathsf{PostBQP/poly}$. Here $\mathsf{PostBQP}$ is the class of languages decidable by polynomial-size quantum circuits with *postselection*—meaning the ability to measure a qubit that has a nonzero probability of being $|1\rangle$, and then *assume* that the measurement outcome will be $|1\rangle$. Clearly $\mathsf{PostBQP}$ lies somewhere between BQP and PP; one can think of it as a quantum analogue of the classical complexity class $\mathsf{BPP_{path}}$ [16]. It turns out, however, that PostBQP = PP [2].

4. Oracle Limitations

Can quantum computers solve NP-complete problems in polynomial time? In 1996 Bennett et al. [7] gave an oracle relative to which $\mathsf{NP} \not\subset \mathsf{BQP}$, providing what is still the best evidence we have that the answer is no. It is easy to extend Bennett et al.'s result to give an oracle relative to which $\mathsf{NP} \not\subset \mathsf{BQP/poly}$; that is, NP is hard even for nonuniform quantum algorithms. But when we try to show $\mathsf{NP} \not\subset \mathsf{BQP/qpoly}$ relative to an oracle, a new difficulty arises: even if the oracle encodes 2^n exponentially hard search problems for each input length n, the quantum advice, being an "exponentially large object" itself, might somehow encode information about all 2^n problems. We need to argue that even if so, only a miniscule fraction of that information can be extracted by measuring the advice.

How does one prove such a statement? As we observe in Theorem 11 below, it suffices to prove a *direct product theorem* for quantum search. This is a theorem that in its weakest form says the following: given N items, K of which are marked, if we lack enough time to find even *one* marked item, then the probability of finding all K items decreases exponentially in K. In other words, there are no devious correlations by which success in finding one marked item leads to success in finding the others. Although this statement is intuitively obvious, the intuition that reassures us of its obviousness is uncomfortably close to that which tells us Bell inequalities cannot be violated. Therefore proof is required.

A few years ago Klauck [20] claimed to prove a direct product theorem using the hybrid method of Bennett et al. [7]. However, Klauck's proof is incorrect.[9] In this section we give the first correct proof, based on the polynomial method of Beals et al. [6]. Besides giving a relativized separation of NP from BQP/qpoly, our result can be used to recover and even improve upon the conclusions in [20] about the hardness of quantum sorting (see [21] for details). We expect the result to have other applications as well.

We first need a lemma about the behavior of functions under repeated differentiation.

Lemma 8 *Let $f : \mathbb{R} \to \mathbb{R}$ be a smooth function such that for some positive integer K, we have $f(i) = 0$ for all $i \in \{0, \ldots, K-1\}$ and $f(K) = \delta$. Then for all $m \in \{1, \ldots, K\}$, there exists an $x \in [K-m, K]$ such that $f^{(m)}(x) \geq \delta/m!$, where $f^{(m)}(x)$ is the m^{th} derivative of f evaluated at x.*

Proof. We claim, by induction on m, that there exist $K-m+1$ points $0 \leq x_0^{(m)} < \cdots < x_{K-m}^{(m)} \leq K$, with $i \leq x_i^{(m)}$ for all i, such that $f^{(m)}\left(x_i^{(m)}\right) = 0$ for all $i \leq K-m-1$ and $f^{(m)}\left(x_{K-m}^{(m)}\right) \geq \delta/m!$. If we define $x_i^{(0)} = i$, then the base case $m = 0$ is immediate from the conditions of the lemma. Suppose the claim is true for m; then by elementary calculus, for all $i \leq K-m-2$ there exists a point $x_i^{(m+1)} \in \left(x_i^{(m)}, x_{i+1}^{(m)}\right)$ such that $f^{(m+1)}\left(x_i^{(m+1)}\right) = 0$. Notice that $x_i^{(m+1)} \geq x_i^{(m)} \geq \cdots \geq x_i^{(0)} = i$. So there is also a point $x_{K-m-1}^{(m+1)} \in \left(x_{K-m-1}^{(m)}, x_{K-m}^{(m)}\right)$ such that

$$\begin{aligned} f^{(m+1)}\left(x_{K-m-1}^{(m+1)}\right) &\geq \frac{f^{(m)}\left(x_{K-m}^{(m)}\right) - f^{(m)}\left(x_{K-m-1}^{(m)}\right)}{x_{K-m}^{(m)} - x_{K-m-1}^{(m)}} \\ &\geq \frac{\delta/m! - 0}{K - (K-m-1)} \\ &= \frac{\delta}{(m+1)!}. \end{aligned}$$

■

Using Lemma 8, we can sometimes lower-bound the degree of a real polynomial even if its derivative is small throughout the region of interest.

Lemma 9 *Let p be a real polynomial such that*

(i) $p(x) \in [0,1]$ at all integer points $x \in \{0, \ldots, N\}$, and

(ii) for some positive integer $K \leq N$, we have $p(i) = 0$ for all $i \in \{0, \ldots, K-1\}$ and $p(K) = \delta$.

Then $\deg(p) = \Omega\left(\sqrt{N\delta^{1/K}}\right)$, and $\deg(p) = \Omega\left(\frac{\sqrt{NK}}{\log^{3/2}(1/\delta)}\right)$ provided $\delta \geq 1/2^K$.

Proof. We first prove the $\Omega\left(\sqrt{N\delta^{1/K}}\right)$ bound. Let $r^{(0)} = \max_{0 \leq x,y \leq N}(p(x) - p(y))$ be the *range* of p on the interval $[0, N]$. Also, for $m \geq 1$, let $p^{(m)}$ be the m^{th} derivative of p and let $r^{(m)} = \max_{0 \leq x \leq N}\left|p^{(m)}(x)\right|$. Then as discussed by Rivlin [28], an inequality due to V. A. Markov, the younger brother of A. A. Markov, states that for all $m \in \{1, \ldots, \deg(p)\}$,

$$\begin{aligned} r^{(m)} &\leq \left(\frac{r^{(0)}}{N}\right)^m T_{\deg(p)}^{(m)}(1) \\ &\leq \left(\frac{r^{(0)}}{N}\right)^m \frac{2^{m-1}(m-1)!}{(2m-1)!} \deg(p)^{2m}. \end{aligned}$$

9 Specifically, the last sentence in the proof of Lemma 5 in [20] ("Clearly this probability is at least $q_x(p_x - \alpha)$") is not justified by what precedes it.

Here T_d is the d^{th} Chebyshev polynomial. Rearranging,

$$\deg(p) \geq \sqrt{\frac{N}{r^{(0)}}\left(\frac{(2m-1)!r^{(m)}}{2^{m-1}(m-1)!}\right)^{1/m}}$$

for all $m \geq 1$ (if $m > \deg(p)$ then $r^{(m)} = 0$ so the bound is trivial). By an observation of Nisan and Szegedy [25] (see also [13, 29]), condition (i) implies that $r^{(0)} \leq 1 + r^{(1)}$, and hence that $\deg(p) \geq \sqrt{N\left(r^{(0)}-1\right)/r^{(0)}}$ by a well-known inequality of A. A. Markov. If $r^{(0)} > 2$ we are done, so assume $r^{(0)} \leq 2$. Then $r^{(m)} \geq \delta/m!$ for all $m \leq K$ by Lemma 8, so taking $m = K$ yields

$$\begin{aligned}\deg(p) &\geq \sqrt{\frac{N}{2}\left(\frac{(2K-1)!}{2^{K-1}(K-1)!}\cdot\frac{\delta}{K!}\right)^{1/K}}\\ &= \Omega\left(\sqrt{N\delta^{1/K}}\right).\end{aligned}$$

We now prove the $\Omega\left(\sqrt{NK/\log^{3/2}(1/\delta)}\right)$ bound. Let

$$c^{(m)} = \max_{mw \leq x \leq N-mw}\left|p^{(m)}(x)\right|$$

where we set $w = K/\log(1/\delta)$ with foresight. Then Bernstein's inequality (see Rivlin [28]) states that

$$\begin{aligned}\deg(p) &\geq \max_{0\leq x\leq N}\left[\sqrt{x(N-x)}\frac{p^{(1)}(x)}{c^{(0)}}\right]\\ &\geq \sqrt{w(N-w)}\frac{c^{(1)}}{c^{(0)}}.\end{aligned}$$

More generally, since $p^{(m)}$ has range at most $2c^{(m)}$,

$$\deg(p) \geq \max\left\{\begin{array}{c}\sqrt{w((N-2w)-w)}\frac{c^{(2)}}{2c^{(1)}},\ldots,\\ \sqrt{w((N-2dw)-w)}\frac{c^{(d+1)}}{2c^{(d)}}\end{array}\right\},$$

where $d = K/(5w)$ (we omit floor and ceiling signs for convenience) and we have used the fact that $\deg(p) \geq \deg\left(p^{(m)}\right)$ for all m. Assume $\deg(p) \leq \sqrt{Nw}$ (otherwise we are done); then a result of Paturi [27] together with condition (i) implies that $c^{(1)}$ is at most some constant B. Since

$$\sqrt{w((N-mw)-w)} \geq \sqrt{\frac{Nw}{2}}$$

for all $m \in \{1,\ldots,d\}$, we have

$$c^{(m)} \leq \sqrt{\frac{8}{Nw}}\deg(p)\,c^{(m-1)} \leq B\left(\sqrt{\frac{8}{Nw}}\deg(p)\right)^m$$

for such m. Notice that since $mw+w < K-m$, Lemma 8 yields $c^{(m)} \geq \delta/m! \geq \delta/m^m$. Combining and setting $m = d$,

$$\begin{aligned}\delta &\leq Bd^d\left(\sqrt{\frac{8}{Nw}}\deg(p)\right)^d\\ &\leq B\left(\sqrt{\frac{8}{N}}\frac{K}{5w^{3/2}}\deg(p)\right)^{K/(5w)}\end{aligned}$$

and hence

$$\begin{aligned}\deg(p) &= \Omega\left(\left(\frac{\delta}{B}\right)^{5w/K}\frac{\sqrt{N}w^{3/2}}{K}\right)\\ &= \Omega\left(\frac{\sqrt{NK}}{\log^{3/2}(1/\delta)}\right)\end{aligned}$$

provided $\delta \geq 1/2^K$. ∎

In an earlier version of this paper, Lemma 9 applied A. A. Markov's inequality inductively to show the lower bounds $\deg(p) = \Omega\left(\sqrt{N\delta^{1/K}/K}\right)$, and $\deg(p) = \Omega\left(\sqrt{N/\log(1/\delta)}\right)$ provided $\delta \geq 1/2^K$. Both of these bounds suffice for an oracle separation, but they are weaker than the bounds obtained from either V. A. Markov's or Bernstein's inequalities, and they yield correspondingly weaker direct product theorems. (We thank Klauck for alerting us to V. A. Markov's inequality.)

Also, after a version of this paper was circulated, Klauck, Špalek, and de Wolf [21] improved our bound to the essentially tight $\deg(p) = \Omega\left(\sqrt{NK\delta^{1/K}}\right)$, which implies in particular that δ decreases exponentially in K whenever $\deg(p) = o\left(\sqrt{NK}\right)$. They did this by *factoring* p instead of differentiating it as in Lemma 8.

We can now prove the direct product theorem.

Theorem 10 (Direct Product Theorem) *Suppose a quantum algorithm makes T queries to an oracle string $X \in \{0,1\}^N$ with Hamming weight $|X| = K$. Let δ be the probability that the algorithm finds all K of the '1' bits in the worst case. Then $\delta \leq \left(cT^2/N\right)^K$ for some constant c, and $\delta \leq \exp\left(-\Omega\left(\sqrt[3]{NK/T^2}\right)\right)$ provided $T \geq \sqrt{N}/K$.*

Proof. Have the algorithm accept if it finds K or more '1' bits and reject otherwise. Let $p(i)$ be the expected probability of acceptance if X is chosen uniformly at random subject to $|X| = i$. Then we know the following about p:

(i) $p(i) \in [0,1]$ at all integer points $i \in \{0,\ldots,N\}$.
(ii) $p(i) = 0$ for all $i \in \{0,\ldots,K-1\}$.
(iii) $p(K) \geq \delta$.

Furthermore, by Beals et al. [6], p is a polynomial in i satisfying $\deg(p) \leq 2T$. The theorem now follows from Lemma 9. ■

The desired oracle separation is an easy corollary of Theorem 10.

Theorem 11 *There exists an oracle relative to which* $\mathsf{NP} \not\subset \mathsf{BQP/qpoly}$.

Proof. Given an oracle $A : \{0,1\}^* \rightarrow \{0,1\}$, define the language L_A by $(y,z) \in L_A$ if and only if $y \leq z$ lexicographically and there exists an x such that $y \leq x \leq z$ and $A(x) = 1$. Clearly $L_A \in \mathsf{NP}^A$ for all A. We argue that for some A, no $\mathsf{BQP/qpoly}$ machine M with oracle access to A can decide L_A. Without loss of generality we assume M is fixed, so that only the advice states $\{|\psi_n\rangle\}_{n\geq 1}$ depend on A. We also assume the advice is boosted, so that M's error probability on any input (y,z) is $2^{-\Omega(n^2)}$.

Choose a set $S \subset \{0,1\}^n$ uniformly at random subject to $|S| = 2^{n/10}$; then for all $x \in \{0,1\}^n$, set $A(x) = 1$ if and only if $x \in S$. We claim that by using M, an algorithm could find all $2^{n/10}$ elements of S with high probability after only $2^{n/10}\operatorname{poly}(n)$ queries to A. Here is how: first use binary search (repeatedly halving the distance between y and z) to find the lexicographically first element of S. By Lemma 2, the boosted advice state $|\psi_n\rangle$ is good for $2^{\Omega(n^2)}$ uses, so this takes only $\operatorname{poly}(n)$ queries. Then use binary search to find the lexicographically second element, and so on until all elements have been found.

Now replace $|\psi_n\rangle$ by the maximally mixed state as in Theorem 6. This yields an algorithm that uses no advice, makes $2^{n/10}\operatorname{poly}(n)$ queries, and finds all $2^{n/10}$ elements of S with probability $2^{-O(\operatorname{poly}(n))}$. But taking $N = 2^n$, $K = 2^{n/10}$, $\delta = 2^{-O(\operatorname{poly}(n))}$, and $T = 2^{n/10}\operatorname{poly}(n)$, such an algorithm violates the lower bound of Theorem 10. ■

Indeed one can show $\mathsf{NP} \not\subset \mathsf{BQP/qpoly}$ relative a random oracle with probability 1.[10]

5. The Trace Distance Method

This section introduces a new method for proving lower bounds on quantum one-way communication complexity. Unlike in Section 3, here we do not try to simulate quantum protocols using classical ones. Instead we prove lower bounds for quantum protocols directly, by reasoning about the trace distance between two possible distributions over Alice's quantum message (that is, between two mixed states). The result is a method that works even if Alice's and Bob's inputs are the same size.

We first state our method as a general theorem; then, in Section 5.1 and Appendix 8, we apply the theorem to prove lower bounds for two problems of Ambainis. Let $\|\mathcal{D} - \mathcal{E}\|$ denote the variation distance between probability distributions $\mathcal{D}$ and $\mathcal{E}$.

Theorem 12 *Let* $f : \{0,1\}^n \times \{0,1\}^m \rightarrow \{0,1\}$ *be a total Boolean function. For each* $y \in \{0,1\}^m$*, let* $\mathcal{A}_y$ *be a distribution over* $x \in \{0,1\}^n$ *such that* $f(x,y) = 1$*. Let* $\mathcal{B}$ *be a distribution over* $y \in \{0,1\}^m$*, and let* $\mathcal{D}_k$ *be the distribution over* $(\{0,1\}^n)^k$ *formed by first choosing* $y \in \mathcal{B}$ *and then choosing* k *samples independently from* $\mathcal{A}_y$*. Suppose that* $\Pr_{x\in\mathcal{D}_1, y\in\mathcal{B}}[f(x,y) = 0] = \Omega(1)$ *and that* $\|\mathcal{D}_2 - \mathcal{D}_1^2\| \leq \delta$*. Then* $Q_2^1(f) = \Omega(\log 1/\delta)$.

Proof. Suppose that if Alice's input is x, then she sends Bob the l-qubit mixed state ρ_x. Suppose also that for every $x \in \{0,1\}^n, y \in \{0,1\}^m$, Bob outputs $f(x,y)$ with probability $1/2 + \Omega(1)$. Then by amplifying a constant number of times, Bob's success probability can be made $1 - \varepsilon$ for any constant $\varepsilon > 0$. So with $L = O(l)$ qubits of communication, Bob can distinguish the following two cases with constant bias:

Case I. y was drawn from $\mathcal{B}$ and x from $\mathcal{D}_1$.

Case II. y was drawn from $\mathcal{B}$ and x from $\mathcal{A}_y$.

For in Case I, we assumed that $f(x,y) = 0$ with constant probability, whereas in Case II, $f(x,y) = 1$ always. An equivalent way to say this is that with constant probability over y, Bob can distinguish the mixed states $\rho = \mathrm{EX}_{x\in\mathcal{D}_1}[\rho_x]$ and $\rho_y = \mathrm{EX}_{x\in\mathcal{A}_y}[\rho_x]$ with constant bias. The maximum probability of distinguishing ρ from ρ_y is given by the trace distance $\|\rho - \rho_y\|_{\mathrm{tr}} = \frac{1}{2}\sum_{i=1}^{2^L}|\lambda_i|$, where $\lambda_1, \ldots, \lambda_{2^L}$ are the eigenvalues of the Hermitian matrix $\rho - \rho_y$. By the Cauchy-Schwarz inequality, this is at most

$$\frac{1}{2}\sqrt{2^L \sum_{i=1}^{2^L} \lambda_i^2} = 2^{L/2-1}\sqrt{\sum_{i,j=1}^{2^L}\left|(\rho)_{ij} - (\rho_y)_{ij}\right|^2}$$

where $(\rho)_{ij}$ is the (i,j) entry of ρ. We claim that

$$\mathop{\mathrm{EX}}_{y\in\mathcal{B}}\left[\sum_{i,j=1}^{2^L}\left|(\rho)_{ij} - (\rho_y)_{ij}\right|^2\right] \leq 2\delta.$$

Since

$$\mathop{\mathrm{EX}}_{y\in\mathcal{B}}\left[\|\rho - \rho_y\|_{\mathrm{tr}}^2\right] \leq \mathop{\mathrm{EX}}_{y\in\mathcal{B}}\left[2^{L-2}\sum_{i,j=1}^{2^L}\left|(\rho)_{ij} - (\rho_y)_{ij}\right|^2\right],$$

10 First group the oracle bits into polynomial-size blocks as Bennett and Gill [9] did, then use the techniques of Aaronson [1] to show that the acceptance probability is a low-degree univariate polynomial in the number of all-0 blocks. The rest of the proof follows Theorem 11.

it follows by another application of Cauchy-Schwarz that $\mathrm{EX}_{y\in\mathcal{B}}\left[\|\rho-\rho_y\|_{\mathrm{tr}}\right] \leq \sqrt{2^{L-1}\delta}$ and hence $L = \Omega(\log 1/\delta)$ is needed for Bob to distinguish Case I from Case II with constant bias.

Let us now prove the claim. We have

$$\begin{aligned}
&\underset{y\in\mathcal{B}}{\mathrm{EX}}\left[\sum_{i,j=1}^{2^L}\left|(\rho)_{ij}-(\rho_y)_{ij}\right|^2\right] \\
&= \sum_{i,j=1}^{2^L}\left(\begin{array}{c}\left|(\rho)_{ij}\right|^2 - 2\,\mathrm{Re}\left((\rho)_{ij}^*\,\mathrm{EX}_{y\in\mathcal{B}}\left[(\rho_y)_{ij}\right]\right)\\ +\,\mathrm{EX}_{y\in\mathcal{B}}\left[\left|(\rho_y)_{ij}\right|^2\right]\end{array}\right) \\
&= \sum_{i,j=1}^{2^L}\left(\underset{y\in\mathcal{B}}{\mathrm{EX}}\left[\left|(\rho_y)_{ij}\right|^2\right]-\left|(\rho)_{ij}\right|^2\right),
\end{aligned}$$

since $\mathrm{EX}_{y\in\mathcal{B}}\left[(\rho_y)_{ij}\right]=(\rho)_{ij}$. For a given (i,j) pair,

$$\begin{aligned}
&\underset{y\in\mathcal{B}}{\mathrm{EX}}\left[\left|(\rho_y)_{ij}\right|^2\right]-\left|(\rho)_{ij}\right|^2 \\
&= \underset{y\in\mathcal{B}}{\mathrm{EX}}\left[\left|\underset{x\in\mathcal{A}_y}{\mathrm{EX}}\left[(\rho_x)_{ij}\right]\right|^2\right]-\left|\underset{x\in\mathcal{D}_1}{\mathrm{EX}}\left[(\rho_x)_{ij}\right]\right|^2 \\
&= \underset{y\in\mathcal{B},x,z\in\mathcal{A}_y}{\mathrm{EX}}\left[(\rho_x)_{ij}^*(\rho_z)_{ij}\right]-\underset{x,z\in\mathcal{D}_1}{\mathrm{EX}}\left[(\rho_x)_{ij}^*(\rho_z)_{ij}\right] \\
&= \sum_{x,z}\left(\Pr_{\mathcal{D}_2}[x,z]-\Pr_{\mathcal{D}_1^2}[x,z]\right)(\rho_x)_{ij}^*(\rho_z)_{ij}.
\end{aligned}$$

Now for all x,z,

$$\left|\sum_{i,j=1}^{2^L}(\rho_x)_{ij}^*(\rho_z)_{ij}\right| \leq \sum_{i,j=1}^{2^L}\left|(\rho_x)_{ij}\right|^2 \leq 1.$$

Hence

$$\begin{aligned}
&\sum_{x,z}\left(\Pr_{\mathcal{D}_2}[x,z]-\Pr_{\mathcal{D}_1^2}[x,z]\right)\sum_{i,j=1}^{2^L}(\rho_x)_{ij}^*(\rho_z)_{ij} \\
&\leq \sum_{x,z}\left(\Pr_{\mathcal{D}_2}[x,z]-\Pr_{\mathcal{D}_1^2}[x,z]\right) = 2\left\|\mathcal{D}_2-\mathcal{D}_1^2\right\| \leq 2\delta,
\end{aligned}$$

and we are done. ∎

The difficulty in extending Theorem 12 to partial functions is that the distribution $\mathcal{D}_1$ might not make sense, since it might assign a nonzero probability to some x for which $f(x,y)$ is undefined.

5.1. Applications

In this subsection we apply Theorem 12 to prove lower bounds for two problems of Ambainis. To facilitate further research and to investigate the scope of our method, we state the problems in a more general way than Ambainis did. Given a group G, the *coset problem* $\mathrm{Coset}(G)$ is defined as follows. Alice is given a coset C of a subgroup in G, and Bob is given an element $y\in G$. Bob must output 1 if $y\in C$ and 0 otherwise. By restricting the group G, we obtain many interesting and natural problems. For example, if p is prime then $\mathrm{Coset}(\mathbb{Z}_p)$ is just the equality problem, so $Q_2^1(\mathrm{Coset}(\mathbb{Z}_p)) = \Theta(\log\log p)$.

Corollary 13 $Q_2^1\left(\mathrm{Coset}\left(\mathbb{Z}_p^2\right)\right)=\Theta(\log p)$.

Proof. The upper bound is obvious. For the lower bound, it suffices to consider a function f_p defined as follows. Alice is given $\langle x,y\rangle\in\mathbb{F}_p^2$ and Bob is given $\langle a,b\rangle\in\mathbb{F}_p^2$; then $f_p(x,y,a,b)=1$ if $y\equiv ax+b\,(\mathrm{mod}\,p)$ and $f_p(x,y,a,b)=0$ otherwise. Let $\mathcal{B}$ be the uniform distribution over $\langle a,b\rangle\in\mathbb{F}_p^2$, and let $\mathcal{A}_{a,b}$ be the uniform distribution over $\langle x,y\rangle$ such that $y\equiv ax+b\,(\mathrm{mod}\,p)$. Thus $\mathcal{D}_1$ is the uniform distribution over $\langle x,y\rangle\in\mathbb{F}_p^2$; note that

$$\Pr_{\langle x,y\rangle\in\mathcal{D}_1,\langle a,b\rangle\in\mathcal{B}}\left[f_p(x,y,a,b)=0\right]=1-\frac{1}{p}.$$

Given $\langle x,y\rangle,\langle z,w\rangle\in\mathbb{F}_p^2$, there are three cases regarding $\Pr_{\mathcal{D}_2}[\langle x,y\rangle,\langle z,w\rangle]$:

(1) $\langle x,y\rangle=\langle z,w\rangle$ (p^2 input pairs). In this case the probability is proportional to $1/p$.

(2) $x\neq z$ (p^4-p^3 input pairs). In this case the probability is proportional to $1/p^2$ by pairwise independence.

(3) $x=z$ but $y\neq w$ (p^3-p^2 input pairs). In this case the probability is 0.

Dividing by a normalizing factor of $p^2/p+\left(p^4-p^3\right)/p^2=p^2$, we obtain

$$\left\|\mathcal{D}_2-\mathcal{D}_1^2\right\|=\frac{1}{2}\left(\begin{array}{c}p^2\left|\frac{1}{p^3}-\frac{1}{p^4}\right|+\\ \left(p^4-p^3\right)\left|\frac{1}{p^4}-\frac{1}{p^4}\right|+\\ \left(p^3-p^2\right)\left|0-\frac{1}{p^4}\right|\end{array}\right)=\frac{1}{p}-\frac{1}{p^2}.$$

Therefore $\log(1/\delta)=\Omega(\log p)$. ∎

We now consider Ambainis' second problem. Given a group G and nonempty set $S\subset G$ with $|S|\leq|G|/2$, the *subset problem* $\mathrm{Subset}(G,S)$ is defined as follows. Alice is given $x\in G$ and Bob is given $y\in G$; then Bob must output 1 if $xy\in S$ and 0 otherwise. Let $\mathcal{S}$ be the distribution over $st^{-1}\in G$ formed by drawing s and t uniformly from S. Then let $\Delta=\|\mathcal{S}-\mathcal{D}_1\|$, where $\mathcal{D}_1$ is the uniform distribution over G. Also, let q be the periodicity of S, defined as the number of distinct sets $gS=\{gs:s\in S\}$ where $g\in G$.

Corollary 14 *For all* G, S,

$$Q_2^1(\text{Subset}(G,S)) = \Omega(\log 1/\Delta + \log\log q).$$

The proof of Corollary 14 is just another calculation, which can be found in Appendix 8. Once we lower-bound $Q_2^1(\text{Subset}(G,S))$ in terms of $1/\Delta$, it remains only to upper-bound the variation distance Δ. The following proposition implies that for all constants $\varepsilon > 0$, if S is chosen uniformly at random subject to $|S| = |G|^{1/2+\varepsilon}$, then $Q_2^1(\text{Subset}(G,S)) = \Omega(\log(|G|))$ with constant probability over S.

Proposition 15 *For all groups* G *and integers* $K \in \{1,\ldots,|G|\}$, *if* $S \subset G$ *is chosen uniformly at random subject to* $|S| = K$, *then* $\Delta = O\left(\sqrt{|G|}/K\right)$ *with* $\Omega(1)$ *probability over* S.

From fingerprinting we also have the following upper bound.

Proposition 16

$$R_2^1(\text{Subset}(G,S)) = O(\log|S| + \log\log q).$$

The proofs of Propositions 15 and 16 are given in Appendix 8 as well.

6. Open Problems

(1) Are $R_2^1(f)$ and $Q_2^1(f)$ polynomially related for every total Boolean function f? Also, can we exhibit *any* asymptotic separation between these measures? The best separation we know of is a factor of 2: for the equality function we have $R_2^1(\text{EQ}) \geq (1-o(1))\log_2 n$, whereas Winter [31] has shown that $Q_2^1(\text{EQ}) \leq (1/2 + o(1))\log_2 n$ using a protocol involving mixed states.[11] This factor-2 savings is tight for equality: a simple counting argument shows that $Q_2^1(\text{EQ}) \geq (1/2 - o(1))\log_2 n$; and although the usual randomized protocol for equality uses $(2+o(1))\log_2 n$ bits, there exist protocols based on error-correcting codes that use only $\log_2(cn) = \log_2 n + O(1)$ bits. All of this holds for any constant error probability $0 < \varepsilon < 1/2$.

(2) As a first step toward answering the above questions, can we lower-bound $Q_2^1(\text{Coset}(G))$ for groups other than $\mathbb{Z}_p^2$ (such as $\mathbb{Z}_2^n$, or nonabelian groups)? Also, can we characterize $Q_2^1(\text{Subset}(G,S))$ for all sets S, closing the gap between the upper and lower bounds?

11 If we restrict ourselves to pure states, then $(1-o(1))\log_2 n$ qubits are needed. Based on that fact, a previous version of this paper claimed incorrectly that $Q_2^1(\text{EQ}) \geq (1-o(1))\log_2 n$.

(3) Is there an oracle relative to which $\mathsf{BQP/poly} \neq \mathsf{BQP/qpoly}$?

(4) Can we give oracles relative to which $\mathsf{NP} \cap \mathsf{coNP}$ and SZK are not contained in $\mathsf{BQP/qpoly}$? Bennett et al. [7] gave an oracle relative to which $\mathsf{NP} \cap \mathsf{coNP} \not\subset \mathsf{BQP}$, while Aaronson [1] gave an oracle relative to which $\mathsf{SZK} \not\subset \mathsf{BQP}$.

(5) Even more ambitiously, can we prove a direct product theorem for quantum query complexity that applies to any partial or total function (not just search)?

(6) For all f (partial or total), is $R_2^1(f) = O(\sqrt{n})$ whenever $Q_2^1(f) = O(\log n)$? In other words, is the separation conjectured by Bar-Yossef et al. [5] the best possible?

(7) Can the result $D^1(f) = O\left(mQ_2^1(f)\log Q_2^1(f)\right)$ for partial f be improved to $D^1(f) = O\left(mQ_2^1(f)\right)$? We do not even know how to rule out $D^1(f) = O\left(m + Q_2^1(f)\right)$.

(8) In the Simultaneous Messages (SM) model, there is no direct communication between Alice and Bob; instead, Alice and Bob both send messages to a third party called the *referee*, who then outputs the function value. The complexity measure is the sum of the two message lengths. Let $R_2^{||}(f)$ and $Q_2^{||}(f)$ be the randomized and quantum bounded-error SM complexities of f respectively, and let $R_2^{||,\text{pub}}(f)$ be the randomized SM complexity if Alice and Bob share an arbitrarily long random string. Building on work by Buhrman et al. [11], Yao [32] showed that $Q_2^{||}(f) = O(\log n)$ whenever $R_2^{||,\text{pub}}(f) = O(1)$. He then asked about the other direction: for some $\varepsilon > 0$, does $R_2^{||,\text{pub}}(f) = O\left(n^{1/2-\varepsilon}\right)$ whenever $Q_2^{||}(f) = O(\log n)$, and does $R_2^{||}(f) = O\left(n^{1-\varepsilon}\right)$ whenever $Q_2^{||}(f) = O(\log n)$? In an earlier version of this paper, we showed that $R_2^{||}(f) = O\left(\sqrt{n}\left(R_2^{||,\text{pub}}(f) + \log n\right)\right)$, which means that a positive answer to Yao's first question would imply a positive answer to the second. Later we learned that Yao independently proved the same result [34].

Here we ask a related question: can $Q_2^{||}(f)$ ever be exponentially smaller than $R_2^{||,\text{pub}}(f)$? (Buhrman et al. [11] showed that $Q_2^{||}(f)$ can be exponentially smaller than $R_2^{||}(f)$.) Kerenidis has pointed out to us that, using the results of Bar-Yossef et al. [5], one can define a *relation* f for which $Q_2^{||}(f)$ is exponentially smaller than $R_2^{||,\text{pub}}(f)$. However, as in the case of $Q_2^1(f)$ versus $R_2^1(f)$, it remains to extend that result to functions.

7. Acknowledgments

I thank Oded Regev for pointing out the connection between advice and one-way communication; Andris Ambainis for posing the coset and subset problems; Harry Buhrman for asking about an upper bound on $\mathsf{BQP/qpoly}$; Lance Fortnow for pointing out that $\mathsf{P/poly} \neq \mathsf{PP/poly}$ implies $\mathsf{PP} \not\subset \mathsf{P/poly}$; Hartmut Klauck and Ronald de Wolf for discussions that led to improvements of Lemma 9; and Iordanis Kerenidis and Andreas Winter for correcting me about the constant factors in $R_2^1(\text{EQ})$ and $Q_2^1(\text{EQ})$.

References

[1] S. Aaronson. Quantum lower bound for the collision problem, *Proc. ACM STOC*, pp. 635–642, 2002. quant-ph/0111102.

[2] S. Aaronson. Is quantum mechanics an island in theoryspace?, *Proceedings of the Växjö Conference "Quantum Theory: Reconsideration of Foundations"* (A. Khrennikov, ed.), 2004. quant-ph/0401062.

[3] L. Adleman, J. DeMarrais, and M.-D. Huang. Quantum computability, *SIAM J. Comp.* 26(5):1524–1540, 1997.

[4] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani. Dense quantum coding and a lower bound for 1-way quantum automata, *Proc. ACM STOC*, pp. 376–383, 1999. quant-ph/9804043.

[5] Z. Bar-Yossef, T. S. Jayram, and I. Kerenidis. Exponential separation of quantum and classical one-way communication complexity, to appear in *Proc. ACM STOC*, 2004.

[6] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. Quantum lower bounds by polynomials, *J. ACM* 48(4):778–797, 2001. quant-ph/9802049.

[7] C. Bennett, E. Bernstein, G. Brassard, and U. Vazirani. Strengths and weaknesses of quantum computing, *SIAM J. Comp.* 26(5):1510–1523, 1997. quant-ph/9701001.

[8] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. Wootters. Teleporting an unknown quantum state by dual classical and EPR channels, *Phys. Rev. Lett.* 70:1895-1898, 1993.

[9] C. H. Bennett and J. Gill. Relative to a random oracle A, $\mathsf{P}^A \neq \mathsf{NP}^A \neq \mathsf{coNP}^A$ with probability 1, *SIAM J. Comp.* 10(1):96–113, 1981.

[10] C. H. Bennett and S. J. Wiesner. Communication via 1- and 2-particle operators on Einstein-Podolsky-Rosen states, *Phys. Rev. Lett.* 69:2881–2884, 1992.

[11] H. Buhrman, R. Cleve, J. Watrous, and R. de Wolf. Quantum fingerprinting, *Phys. Rev. Lett.* 87(16), 2001. quant-ph/0102001.

[12] P. Duriš, J. Hromkovič, J. D. P. Rolim, and G. Schnitger. Las Vegas versus determinism for one-way communication complexity, finite automata, and polynomial-time computations, *Proc. Symposium on Theoretical Aspects of Computer Science (STACS)*, pp. 117–128, 1997.

[13] H. Ehlich and K. Zeller. Schwankung von polynomen zwischen gitterpunkten, *Mathematische Zeitschrift* 86:41–44, 1964.

[14] L. Fortnow and J. Rogers. Complexity limitations on quantum computation, *J. Comp. Sys. Sci.* 59(2):240–252, 1999. cs.CC/9811023.

[15] O. Goldreich. On quantum computing, 2004. http://www.wisdom.weizmann.ac.il/˜oded/on-qc.html.

[16] Y. Han, L. Hemaspaandra, and T. Thierauf. Threshold computation and cryptographic security, *SIAM J. Comp.* 26(1):59–78, 1997.

[17] A. S. Holevo. Some estimates of the information transmitted by quantum communication channels, *Problems of Information Transmission* 9:177–183, 1973 (English translation).

[18] R. M. Karp and R. J. Lipton. Turing machines that take advice, *Enseign. Math.* 28:191–201, 1982.

[19] H. Klauck. Quantum communication complexity, *Proc. International Colloquium on Automata, Languages, and Programming (ICALP)*, 2000. quant-ph/0005032.

[20] H. Klauck. Quantum time-space tradeoffs for sorting, *Proc. ACM STOC*, pp. 69–76, 2003. quant-ph/0211174.

[21] H. Klauck, R. Špalek, and R. de Wolf. Quantum and classical strong direct product theorems and optimal time-space tradeoffs, manuscript, 2004.

[22] E. Kushilevitz and N. Nisan. *Communication Complexity*, Cambridge, 1997.

[23] A. Nayak. Optimal lower bounds for quantum automata and random access codes, *Proc. IEEE FOCS*, pp. 369–377, 1999. quant-ph/9904093.

[24] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*, Cambridge, 2000.

[25] N. Nisan and M. Szegedy. On the degree of Boolean functions as real polynomials, *Comput. Complexity* 4(4):301–313, 1994.

[26] H. Nishimura and T. Yamakami. Polynomial time quantum computation with advice, ECCC TR03-059, 2003. quant-ph/0305100.

[27] R. Paturi. On the degree of polynomials that approximate symmetric Boolean functions, *Proc. ACM STOC*, pp. 468–474, 1992.

[28] T. J. Rivlin. *Chebyshev Polynomials: From Approximation Theory to Algebra and Number Theory*, Wiley, 1990.

[29] T. J. Rivlin and E. W. Cheney. A comparison of uniform approximations on an interval and a finite subset thereof, *SIAM J. Numerical Analysis* 3(2):311–320, 1966.

[30] N. Sauer. On the density of families of sets, *J. Combinatorial Theory Series A*, 13:145–147, 1972.

[31] A. Winter. Quantum and classical message identification via quantum channels, to appear in A. S. Holevo *Festschrift*, Rinton, 2004. quant-ph/0401060.

[32] A. C-C. Yao. On the power of quantum fingerprinting, *Proc. ACM STOC*, pp. 77–81, 2003.

[33] A. C-C. Yao. Some complexity questions related to distributive computing, *Proc. ACM STOC*, pp. 209–213, 1979.

[34] A. C-C. Yao. "Current Problems in Theory" course assignment #3, 2001. http://www.cs.princeton.edu/courses/archive/spr01/cs598a/assignments/hw3.ps.

8. Appendix: Proofs for Subset Problem

Proof of Corollary 14. Let $\mathcal{B}$ be the uniform distribution over $y \in G$, and let $\mathcal{A}_y$ be the uniform distribution over x such that $xy \in S$. Thus $\mathcal{D}_1$ is the uniform distribution over $x \in G$; note that

$$\Pr_{x \in \mathcal{D}_1, y \in \mathcal{B}} [xy \notin S] = 1 - \frac{|S|}{|G|} \geq \frac{1}{2}.$$

We have

$$\begin{aligned}
&\left\| \mathcal{D}_2 - \mathcal{D}_1^2 \right\| \\
&= \frac{1}{2} \sum_{x,z \in G} \left| \frac{|\{y \in G, s,t \in S : xy = s, zy = t\}|}{|G| \, |S|^2} - \frac{1}{|G|^2} \right| \\
&= \frac{1}{2} \sum_{x,z \in G} \left| \frac{\left|\left\{s,t \in S : xz^{-1} = st^{-1}\right\}\right|}{|S|^2} - \frac{1}{|G|^2} \right| \\
&= \frac{1}{2} \sum_{x \in G} \left| \frac{\left|\left\{s,t \in S : x = st^{-1}\right\}\right|}{|S|^2} - \frac{1}{|G|} \right| \\
&= \frac{1}{2} \sum_{x \in G} \left| \Pr_{\mathcal{S}} [x] - \frac{1}{|G|} \right| = \|\mathcal{S} - \mathcal{D}_1\| = \Delta.
\end{aligned}$$

Therefore $\log(1/\delta) = \Omega(\log 1/\Delta)$. The $\log \log q$ part follows easily from Proposition 3. ■

Proof of Proposition 15. We have

$$\begin{aligned}
\Delta = \|\mathcal{S} - \mathcal{D}_1\| &= \frac{1}{2} \sum_{x \in G} \left| \Pr_{\mathcal{S}} [x] - \frac{1}{|G|} \right| \\
&\leq \frac{\sqrt{|G|}}{2} \sqrt{\sum_{x \in G} \left(\Pr_{\mathcal{S}} [x] - \frac{1}{|G|} \right)^2}
\end{aligned}$$

by Cauchy-Schwarz. We claim that

$$\operatorname*{EX}_S \left[\sum_{x \in G} \left(\Pr_{\mathcal{S}} [x] - \frac{1}{|G|} \right)^2 \right] \leq \frac{c}{K^2}$$

for some constant c. It then follows by Markov's inequality that

$$\Pr_S \left[\sum_{x \in G} \left(\Pr_{\mathcal{S}} [x] - \frac{1}{|G|} \right)^2 \geq \frac{2c}{K^2} \right] \leq \frac{1}{2}$$

and hence $\Delta \leq \left(\sqrt{|G|}/2\right) \sqrt{2c/K^2}$ with probability at least 1/2.

Let us now prove the claim. For a particular element $x \in G$,

$$\operatorname*{EX}_S \left[\left(\Pr_{\mathcal{S}} [x] - \frac{1}{|G|} \right)^2 \right] = \operatorname*{EX}_S \left[\left(\Pr_{\mathcal{S}} [x] \right)^2 \right] - \frac{1}{|G|^2},$$

since $\mathrm{EX}_S [\Pr_{\mathcal{S}} [x]] = 1/|G|$ by the fact that G is a group. Letting $S = \{s_1, \ldots, s_K\}$,

$$\begin{aligned}
\left(\Pr_{\mathcal{S}} [x] \right)^2 &= \left(\Pr_{i,j \in \{1,\ldots,K\}} \left[s_i s_j^{-1} = x \right] \right)^2 \\
&= \Pr_{i,j,k,l \in \{1,\ldots,K\}} \left[s_i s_j^{-1} = s_k s_l^{-1} = x \right].
\end{aligned}$$

Then some grunt work shows that

$$\begin{aligned}
&\operatorname*{EX}_S \left[\left(\Pr_{\mathcal{S}} [x] \right)^2 \right] \\
&= \operatorname*{EX}_{i,j,k,l \in \{1,\ldots,K\}} \left[\Pr_S \left[s_i s_j^{-1} = s_k s_l^{-1} = x \right] \right] \\
&= \begin{cases}
\frac{1}{K^4} \left(K^2 + K(K-1) \right) \\
\text{if } x = e \\
\frac{1}{K^4} \left(\frac{K(K-1)}{|G|-1} + \frac{K(K-1)(K-2)(K-3)}{(|G|-1)(|G|-3)} \right) \\
\text{if } x \neq e \text{ but } x^2 = e \\
\frac{1}{K^4} \left(\begin{array}{c} \frac{K(K-1)}{|G|-1} + \frac{2K(K-1)(K-2)}{(|G|-1)(|G|-2)} \\ + \frac{K(K-1)(K-2)(K-3)}{(|G|-1)(|G|-3)} \end{array} \right) \\
\text{if } x^2 \neq e
\end{cases}
\end{aligned}$$

where e is the identity. So

$$\sum_{x \in G} \operatorname*{EX}_S \left[\left(\Pr_{\mathcal{S}} [x] \right)^2 \right] \leq \frac{c}{K^2} + \frac{1}{|G|}$$

for some constant c, and we are done. ■

Proof of Proposition 16. Assume for simplicity that $q = |G|$; otherwise we could reduce to a subgroup $H \leq G$ with $|H| = q$. The protocol is as follows: Alice chooses a prime p uniformly at random from the range $\left[|S|^2 \log^2 |G|, 2|S|^2 \log^2 |G| \right]$; she then sends Bob the pair $(p, x \bmod p)$ where x is interpreted as an integer. This uses $O(\log |S| + \log \log |G|)$ bits. Bob outputs 1 if and only if there exists a $z \in G$ such that $zy \in S$ and $x \equiv z \pmod p$. To see the protocol's correctness, observe that if $x \neq z$ then there at most $\log |G|$ primes p such that $x - z \equiv 0 \pmod p$, whereas the relevant range contains $\Omega\left(\frac{|S|^2 \log^2 |G|}{\log(|S| \log |G|)} \right)$ primes. Therefore if $xy \notin S$, then by the union bound

$$\begin{aligned}
&\Pr_p [\exists z : zy \in S, x \equiv z \pmod p] \\
&= O\left(|S| \log |G| \frac{\log(|S| \log |G|)}{|S|^2 \log^2 |G|} \right) = o(1).
\end{aligned}$$

■