Quantum Versus Classical Proofs and Advice

Scott Aaronson[*] Greg Kuperberg[†]

Abstract

This paper studies whether quantum proofs are more powerful than classical proofs, or in complexity terms, whether $\mathsf{QMA} = \mathsf{QCMA}$. We prove three results about this question. First, we give a "quantum oracle separation" between QMA and QCMA. More concretely, we show that any quantum algorithm needs $\Omega\left(\sqrt{\frac{2^n}{m+1}}\right)$ queries to find an n-qubit "marked state" $|\psi\rangle$, even if given an m-bit classical description of $|\psi\rangle$ together with a quantum black box that recognizes $|\psi\rangle$. Second, we give an explicit QCMA protocol that nearly achieves this lower bound. Third, we show that, in the one previously-known case where quantum proofs seemed to provide an exponential advantage, *classical* proofs are basically just as powerful. In particular, Watrous gave a QMA protocol for verifying non-membership in finite groups. Under plausible group-theoretic assumptions, we give a QCMA protocol for the same problem. Even with no assumptions, our protocol makes only polynomially many queries to the group oracle. We end with some conjectures about quantum versus classical oracles, and about the possibility of a *classical* oracle separation between QMA and QCMA.

1 Introduction

If someone hands you a quantum state, is that more "useful" than being handed a classical string with a comparable number of bits? In particular, are there truths that you can efficiently verify, and are there problems that you can efficiently solve, using the quantum state but not using the string? These are the questions that this paper addresses, and that it answers in several contexts.

Recall that QMA, or Quantum Merlin-Arthur, is the class of decision problems for which a "yes" answer can be verified in quantum polynomial time, with help from a polynomial-size quantum witness state $|\psi\rangle$. Many results are known about QMA: for example, it has natural complete promise problems [18], allows amplification of success probabilities [21], and is contained in PP [21]. Raz and Shpilka [26] have also studied communication complexity variants of QMA.

Yet as Aharonov and Naveh [3] pointed out in 2002, the very definition of QMA raises a fundamental question. Namely: is it really essential that the witness be quantum, or does it suffice for the algorithm *verifying* the witness to be quantum? To address this question, Aharonov and Naveh defined the class QCMA, or "Quantum Classical Merlin-Arthur," to be the same as QMA except that now the witness is classical.[1] We can then ask whether $\mathsf{QMA} = \mathsf{QCMA}$. Not surprisingly, the answer is that we don't know.

If we can't decide whether two complexity classes are equal, the usual next step is to construct a relativized world that separates them. This would provide at least some evidence that the classes are different. But in the case of QMA versus QCMA, even this limited goal has remained elusive.

Closely related to the question of quantum versus classical proofs is that of quantum versus classical *advice*. Compared to a proof, advice has the advantage that it can be trusted, but the disadvantage that it can't be tailored to a particular input. More formally, let $\mathsf{BQP/qpoly}$ be the class of problems solvable in quantum polynomial time, with help from a polynomial-size "quantum advice state" $|\psi_n\rangle$ that depends only on the input length n. Then the question is whether $\mathsf{BQP/qpoly} = \mathsf{BQP/poly}$, where $\mathsf{BQP/poly}$ is the class of problems solvable in quantum polynomial time with help from polynomial-size *classical* advice. Aaronson [2] showed that $\mathsf{BQP/qpoly} \subseteq \mathsf{PP/poly}$, which at least

[*]University of Waterloo. Email: scott@scottaaronson.com.
[†]UC Davis. Email: greg@math.ucdavis.edu.

[1]Some say that this class would more accurately be called CMQA, for "Classical Merlin Quantum Arthur." But QCMA has stuck.

tells us that quantum advice is not "infinitely" more powerful than classical advice. But, like the QMA versus QCMA question, the $\mathsf{BQP/qpoly}$ versus $\mathsf{BQP/poly}$ question has remained open, with not even an oracle separation known.

1.1 Our Results

This paper introduces new tools with which to attack QMA versus QCMA and related questions.

First, we achieve an oracle separation between QMA and QCMA, but only by broadening the definition of "oracle." In particular, we introduce the notion of a *quantum oracle*, which is just an infinite sequence of unitaries $U = \{U_n\}_{n\geq 1}$ that a quantum algorithm can apply in a black-box fashion. Just as a classical oracle models a subroutine to which an algorithm has black-box access, so a quantum oracle models a quantum subroutine, which can take quantum input and produce quantum output. We are able to give a quantum oracle that separates QMA from QCMA:

Theorem 1.1 *There exists a quantum oracle U such that* $\mathsf{QMA}^U \neq \mathsf{QCMA}^U$.

Similarly, there exists a quantum oracle V such that $\mathsf{BQP}^V/\mathsf{qpoly} \neq \mathsf{BQP}^V/\mathsf{poly}$.

Theorem 1.1 implies that if $\mathsf{QMA} = \mathsf{QCMA}$, then any proof of this fact will require "quantumly nonrelativizing techniques": techniques that are sensitive to the presence of quantum oracles. Currently, we do not know of *any* quantumly nonrelativizing techniques that are not also classically nonrelativizing. For this reason, we believe that quantum oracle separations merit the same informal interpretation as classical oracle separations: almost any argument that one might advance against the former, is also an argument against the latter! The difference is that quantum oracle results are sometimes much easier to prove than classical ones. To our knowledge, this paper provides the first example of this phenomenon, but other examples have since emerged [1, 23].

It might be objected that, even if quantum oracle separations are no less trustworthy than classical ones, they certainly aren't *more* trustworthy, and complexity theorists have known since the celebrated $\mathsf{IP} = \mathsf{PSPACE}$ theorem [27] that oracle results sometimes "point in the wrong direction." We wish to stress two points in response. First, oracle results provide at least *some* understanding, thereby opening the way to further progress. This is particularly true in quantum computing, where even the oracle results tend to be much less intuitively obvious than they are in the classical world. Second, complexity theorists do not currently have any nonrelativizing technique for "non-interactive" classes such as QMA and QCMA even remotely analogous to the arithmetization technique that Shamir [27] used to show $\mathsf{IP} = \mathsf{PSPACE}$. We hope such a technique will someday be discovered.

Underlying Theorem 1.1 is the following lower bound. Suppose a unitary oracle U_n acts on n qubits, and suppose that either (i) U_n is the identity, or (ii) there exists a secret n-qubit "marked state" $|\psi_n\rangle$ such that $U_n|\psi_n\rangle = -|\psi_n\rangle$, but $U_n|\varphi\rangle = |\varphi\rangle$ whenever $|\varphi\rangle$ is orthogonal to $|\psi_n\rangle$. Then even if a quantum algorithm is given m bits of classical advice about $|\psi_n\rangle$, the algorithm still needs $\Omega\left(\sqrt{\frac{2^n}{m+1}}\right)$ queries to U_n to distinguish these cases. Note that when $m = 0$, we recover the usual $\Omega\left(\sqrt{2^n}\right)$ lower bound for Grover search as a special case. At the other extreme, if $m \approx 2^n$ then our bound gives nothing—not surprisingly, since the classical advice might contain explicit instructions for preparing $|\psi_n\rangle$. The point is that, if m is *not* exponentially large, then exponentially many queries are needed.

Since $|\psi_n\rangle$ is an arbitrary 2^n-dimensional unit vector, it might be thought obvious that $2^{\Omega(n)}$ bits are needed to describe that vector. The key point, however, is that the QCMA verifier is given not only a classical description of $|\psi_n\rangle$, but also oracle access to U_n. So the question is whether some *combination* of these resources might be exponentially more powerful than either one alone. We prove that the answer is no, using the hybrid argument of Bennett et al. [10] together with geometric results about partitionings of the unit sphere.

In Section 4, we show that our lower bound is basically tight, by giving an algorithm that finds $|\psi_n\rangle$ using $O\left(\sqrt{2^n/m}\right)$ queries when $m \geq 2n$. This algorithm has the drawback of being *computationally* inefficient. To fix this, we give another algorithm that finds $|\psi_n\rangle$ using $O\left(n\sqrt{2^n/m}\right)$ queries together with $O\left(n^2\sqrt{2^n/m} + \operatorname{poly}(m)\right)$ computational steps.

Having separated QMA from QCMA by a quantum oracle, we next revisit the question of whether these classes can be separated by a *classical* oracle. Right now, we know of only one candidate problem for such a separation in the literature: the Group Non-Membership (GNM) problem, which Watrous [30] placed in QMA even though Babai [6] showed it is not in MA as an oracle problem.[2] In Group Non-Membership, Arthur is given black-box access to a

[2] Interestingly, the classes MA and AM were originally defined by Babai in connection with GNM [4].

finite group G, together with a subgroup $H \leq G$ specified by its generators and an element $x \in G$. Arthur's goal is to verify that $x \notin H$, using a number of group operations polynomial in $\log|G|$. (Note that the group *membership* problem is in NP by a result of Babai and Szemerédi [8].) In Watrous's protocol, the quantum witness is simply an equal superposition $|H\rangle$ over the elements of H. Given such a witness, Arthur can check non-membership by comparing the states $|H\rangle$ and $|xH\rangle$, and can similarly check the veracity of $|H\rangle$ by comparing it to $|hH\rangle$, where h is an almost-uniformly random element of H.

Evidently a classical proof of non-membership would have to be completely different. Nevertheless, in Section 5 we show the following:

Theorem 1.2 GNM *has polynomially-bounded* QCMA *query complexity.*

Theorem 1.2 implies that it is pointless to try to prove a classical oracle separation between QMA and QCMA by proving a lower bound on the quantum query complexity of Group Non-Membership. If such a separation is possible, then a new approach will be needed.

The idea of the proof of Theorem 1.2 is that Merlin can "pull the group out of the black box." In other words, he can claim an embedding of a model group Γ into G. This claim is entirely classical, but verifying it requires solving the Normal Hidden Subgroup Problem (NHSP) in Γ. This problem has low query complexity by a result of Ettinger, Høyer, and Knill [14], but is not known to be in BQP. In addition, analyzing the description of Γ is not known to be computationally efficient. Nonetheless, in Section 5.1 we discuss evidence that NHSP is in BQP and that non-membership for Γ is in NP. Based on this evidence, we conjecture the following:

Conjecture 1.3 GNM *is in* QCMA*.*

Given our results in Section 5, the question remains of whether there is some other way to prove a classical oracle separation between QMA and QCMA. In Section 6, we conjecture that the answer is yes:

Conjecture 1.4 *There exists a classical oracle A such that* $\mathsf{QMA}^A \neq \mathsf{QCMA}^A$*. Furthermore, this can be proven by exhibiting an oracle problem with polynomial* QMA *query complexity but exponential* QCMA *query complexity.*

The reason we believe Conjecture 1.4 is that it seems possible, for many purposes, to "encode" a quantum oracle into a classical one. In Section 6 we explain more concretely what we mean by that, and present some preliminary results. For example, we show that there exists a BQP algorithm that maps an oracle string A to an n-qubit pure state $|\psi_A\rangle$, such that if A is uniformly random, then $|\psi_A\rangle$ is (under a suitable metric) close to uniformly random under the Haar measure. We also study the question of applying a random $N \times N$ unitary matrix using a random classical oracle in the same way. We do not know how to do this, but we show that one quantum query will not suffice for this purpose. To prove this, we show that a quantum algorithm that uses just one query can apply at most 4^N different $N \times N$ unitaries, whereas the number of unitaries required to approximate the uniform distribution grows like $2^{\Theta(N^2)}$.

We end in Section 7 with some open problems.

2 Preliminaries

Throughout this paper, we refer to the set of N-dimensional pure states as $\mathbb{CP}^{N-1}$ (that is, complex projective space with $N-1$ dimensions). We use $\Pr$ to denote probability, and E to denote expectation.

We assume familiarity with standard complexity classes such as BQP and MA. For completeness, we now define QMA, QCMA, BQP/qpoly, and BQP/poly.

Definition 2.1 QMA *is the class of languages $L \subseteq \{0,1\}^n$ for which there exists a polynomial-time quantum verifier $\mathcal{Q}$ and a polynomial p such that, for all $x \in \{0,1\}^n$:*

(i) If $x \in L$ then there exists a $p(n)$-qubit quantum proof $|\varphi\rangle$ such that $\mathcal{Q}$ accepts with probability at least $2/3$ given $|x\rangle|\varphi\rangle$ as input.

(ii) If $x \notin L$ then $\mathcal{Q}$ accepts with probability at most $1/3$ given $|x\rangle|\varphi\rangle$ as input, for all purported proofs $|\varphi\rangle$.

The class QCMA *is defined similarly, except that* $|\varphi\rangle$ *is replaced by a classical string* $z \in \{0,1\}^{p(n)}$.

Definition 2.2 BQP/qpoly *is the class of languages* $L \subseteq \{0,1\}^n$ *for which there exists a polynomial-time quantum algorithm* $\mathcal{Q}$*, together with a set of states* $\{|\psi_n\rangle\}_{n\geq 1}$ *(where* $|\psi_n\rangle$ *has size* $p(n)$ *for some polynomial* p*), such that for all* $x \in \{0,1\}^n$*:*

(i) If $x \in L$ *then* $\mathcal{Q}$ *accepts with probability at least* $2/3$ *given* $|x\rangle|\psi_n\rangle$ *as input.*

(ii) If $x \notin L$ *then* $\mathcal{Q}$ *accepts with probability at most* $1/3$ *given* $|x\rangle|\psi_n\rangle$ *as input.*

The class BQP/poly *is defined similarly, except that* $|\psi_n\rangle$ *is replaced by a classical string* $a_n \in \{0,1\}^{p(n)}$.

Let us now explain what we mean by a "quantum oracle." For us, a quantum oracle is simply an infinite sequence of unitary transformations, $U = \{U_n\}_{n\geq 1}$. We assume that each U_n acts on $p(n)$ qubits for some known polynomial p. We also assume that given an n-bit string as input, a quantum algorithm calls only U_n, not U_m for any $m \neq n$. This assumption is only made for simplicity; our results would go through without it.[3] When there is no danger of confusion, we will refer to U_n simply as U.

Formally, the oracle access mechanism is as follows. Assume a quantum computer's state has the form

$$|\Phi\rangle = \sum_z \alpha_z |z\rangle |\phi_z\rangle ,$$

where $|z\rangle$ is a workspace register and $|\phi_{b,z}\rangle$ is a $p(n)$-qubit answer register. Then to "query U_n" means to apply the $p(n)$-qubit unitary transformation that maps $|\Phi\rangle$ to

$$|\Phi'\rangle = \sum_z \alpha_z |z\rangle U_n |\phi_z\rangle .$$

Let $\mathcal{C}$ be a quantum complexity class, and let $U = \{U_n\}_{n\geq 1}$ be a quantum oracle. Then by $\mathcal{C}^U$, we will mean the class of problems solvable by a $\mathcal{C}$ machine that, given an input of length n, can query U_n at unit cost as many times as it likes.

In defining the notion of quantum oracle, at least two choices present themselves that have no counterpart for classical oracles:

(1) If we can apply a quantum oracle U, then can we also apply controlled-U (that is, U conditioned on a control qubit $|b\rangle$)?

(2) If we can apply U, then can we also apply U^{-1}?

At least for the present paper, the answers to these questions will not matter, for the following reasons. First, all of the quantum oracles U that we consider will be self-inverse (that is, $U = U^{-1}$). Second, while our algorithms *will* need to apply controlled-U, that is only for the technical reason that we will define U so that $U|\psi\rangle = -|\psi\rangle$ if $|\psi\rangle$ is the marked state, and $U|\varphi\rangle = |\varphi\rangle$ whenever $\langle\varphi|\psi\rangle = 0$. If we stipulated instead that $U|\psi\rangle|b\rangle = |\psi\rangle|b \oplus 1\rangle$ and $U|\varphi\rangle|b\rangle = |\varphi\rangle|b\rangle$ whenever $\langle\varphi|\psi\rangle = 0$, then U alone would suffice.

Yet even though these choices will not matter for our results, it still seems worthwhile to discuss them a bit, since they might arise in future work involving quantum oracles.

One could argue that (i) the purpose of an oracle is to model a *subroutine* that an algorithm can call without understanding its internal structure, and that (ii) given a quantum circuit for applying some unitary operation U, one can easily produce a circuit for applying controlled-U or U^{-1}, without understanding anything about the original circuit's structure. In particular, to produce a circuit for controlled-U, one simply conditions each gate on the control qubit; while to produce a circuit for U^{-1}, one simply inverts all the gates and reverses their order. These considerations suggest that the answers to questions (1) and (2) should both be 'yes.' On the other hand, it would still be interesting to know whether disallowing controlled-U or U^{-1} would let us prove more quantum oracle separations. (Note that if we disallow these operations, then the set of inequivalent quantum oracles becomes *larger*.)

[3] If one made the analogous assumption in *classical* complexity—that given an input of length n, an algorithm can query the oracle only on strings of length n—one could simplify a great many oracle results without any loss of conceptual content.

3 Quantum Oracle Separations

The aim of this section is to prove Theorem 1.1: that there exists a quantum oracle U such that $\mathsf{QMA}^U \neq \mathsf{QCMA}^U$. The same ideas will also yield a quantum oracle V such that $\mathsf{BQP}^V/\mathsf{qpoly} \neq \mathsf{BQP}^V/\mathsf{poly}$.

To prove these oracle separations, we first need a geometric lemma about probability measures on quantum states. Let μ be the uniform probability measure over N-dimensional pure states (that is, over $\mathbb{CP}^{N-1}$). The following notion will play a key role in our argument.

Definition 3.1 *For all $p \in [0,1]$, a probability measure σ over $\mathbb{CP}^{N-1}$ is called p-uniform if $p\sigma \leq \mu$. Equivalently, σ is p-uniform if it can be obtained by starting from μ, and then conditioning on an event that occurs with probability at least p.*

So for example, we obtain a p-uniform measure if we start from μ and then condition on $\log_2 1/p$ bits of classical information about $|\psi\rangle$. Our geometric lemma says that if $|\psi\rangle$ is drawn from a p-uniform measure, then for every mixed state ρ, the squared fidelity between $|\psi\rangle$ and ρ has small expectation. More precisely:

Lemma 3.2 *Let σ be a p-uniform probability measure over $\mathbb{CP}^{N-1}$. Then for all ρ,*

$$\operatorname*{E}_{|\psi\rangle \in \sigma}\left[\langle\psi|\rho|\psi\rangle\right] = O\left(\frac{1+\log 1/p}{N}\right).$$

The proof of Lemma 3.2 is deferred to Section 3.1. In this section we assume the lemma, and show how to use it to prove our main result. In particular, we show that any quantum algorithm needs $\Omega\left(\sqrt{\frac{2^n}{m+1}}\right)$ queries to find an n-qubit marked state $|\psi\rangle$, even if given m bits of classical advice about $|\psi\rangle$.

Theorem 3.3 *Suppose we are given oracle access to an n-qubit unitary U, and want to decide which of the following holds:*

(i) There exists an n-qubit "quantum marked state" $|\psi\rangle$ such that $U|\psi\rangle = -|\psi\rangle$, but $U|\phi\rangle = |\phi\rangle$ whenever $\langle\phi|\psi\rangle = 0$; or

(ii) $U = I$ is the identity operator.

Then even if we have an m-bit classical witness w in support of case (i), we still need $\Omega\left(\sqrt{\frac{2^n}{m+1}}\right)$ queries to verify the witness, with bounded probability of error.

Proof. If $m = \Omega(2^n)$ then the theorem is certainly true, so suppose $m = o(2^n)$. Let A be a quantum algorithm that queries U. Also, let U_ψ be an n-qubit unitary such that $U_\psi|\psi\rangle = -|\psi\rangle$, but $U_\psi|\phi\rangle = |\phi\rangle$ whenever $\langle\phi|\psi\rangle = 0$. Then A's goal is to accept if and only if $U = U_\psi$ for some $|\psi\rangle$.

For each n-qubit pure state $|\psi\rangle$, let us fix a classical witness $w \in \{0,1\}^m$ that maximizes the probability that A accepts, given U_ψ as oracle. Let $S(w)$ be the set of $|\psi\rangle$'s associated with a given witness w. Since the $S(w)$'s form a partition of $\mathbb{CP}^{2^n-1}$, clearly there exists a witness, call it w^*, such that

$$\Pr_{|\psi\rangle \in \mu}\left[|\psi\rangle \in S(w^*)\right] \geq \frac{1}{2^m}.$$

Fix that w^* (or in other words, hardwire w^* into A). Then to prove the theorem, it suffices to establish the following claim: A cannot distinguish the case $U = U_\psi$ from the case $U = I$ by making $o\left(\sqrt{\frac{2^n}{m+1}}\right)$ queries to U, with high probability if $|\psi\rangle$ is chosen uniformly at random from $S(w^*)$.

To prove the claim, we use a generalization of the hybrid argument of Bennett et al. [10]. Suppose that A makes T queries to U. (Technically speaking, we should also allow queries to controlled-U, but this will make no difference in our analysis.) Then for all $0 \leq t \leq T$, let $|\Phi_t\rangle$ be the final state of A, assuming that $U = I$ for the first t queries, and $U = U_\psi$ for the remaining $T-t$ queries. Thus $|\Phi_0\rangle$ is the final state in case (i), while $|\Phi_T\rangle$ is the final state in case (ii). We will argue that $|\Phi_t\rangle$ cannot be very far from $|\Phi_{t-1}\rangle$, with high probability over the choice of marked state $|\psi\rangle$. Intuitively, this is because the computations of $|\Phi_t\rangle$ and $|\Phi_{t-1}\rangle$ differ in only a single query, and with high probability that query cannot have much overlap with $|\psi\rangle$. We will then conclude, by the triangle inequality, that $|\Phi_0\rangle$ cannot be far from $|\Phi_T\rangle$ unless T is large.

More formally, let ρ_t be the marginal state of the query register just before the t^{th} query, assuming the "control case" $U = I$. Also, let $\rho_t = \sum p_i |\varphi_i\rangle\langle\varphi_i|$ be an arbitrary decomposition of ρ_t into pure states. Then for every i, the component of $|\varphi_i\rangle$ orthogonal to $|\psi\rangle$ is unaffected by the t^{th} query. Therefore

$$\begin{aligned} \| |\Phi_t\rangle - |\Phi_{t-1}\rangle \|_2 &\le \sum_i p_i \cdot 2 |\langle\varphi_i|\psi\rangle| \\ &= 2\sum_i p_i \sqrt{\langle\psi|\varphi_i\rangle\langle\varphi_i|\psi\rangle} \\ &\le 2\sqrt{\sum_i p_i \langle\psi|\varphi_i\rangle\langle\varphi_i|\psi\rangle} \\ &= 2\sqrt{\langle\psi|\rho_t|\psi\rangle}, \end{aligned}$$

where the third line uses the Cauchy-Schwarz inequality (the average of the square root is at most the square root of the average). Now let σ be the uniform probability measure over $S(w^*)$, and observe that σ is 2^{-m}-uniform. So by Lemma 3.2,

$$\begin{aligned} \operatorname*{E}_{|\psi\rangle\in\sigma}\left[\| |\Phi_t\rangle - |\Phi_{t-1}\rangle \|_2\right] &\le 2 \operatorname*{E}_{|\psi\rangle\in\sigma}\left[\sqrt{\langle\psi|\rho_t|\psi\rangle}\right] \\ &\le 2\sqrt{\operatorname*{E}_{|\psi\rangle\in\sigma}\left[\langle\psi|\rho_t|\psi\rangle\right]} \\ &\le 2\sqrt{\frac{1+\ln(1/2^{-m})}{2^n}} \\ &= O\left(\sqrt{\frac{m+1}{2^n}}\right), \end{aligned}$$

where the second line again uses the Cauchy-Schwarz inequality. Finally,

$$\operatorname*{E}_{|\psi\rangle\in S(w^*)}\left[\| |\Phi_T\rangle - |\Phi_0\rangle \|_2\right] \le \sum_{t=1}^{T} \operatorname*{E}_{|\psi\rangle\in S(w^*)}\left[\| |\Phi_t\rangle - |\Phi_{t-1}\rangle \|_2\right] = O\left(T\sqrt{\frac{m+1}{2^n}}\right)$$

by the triangle inequality. This implies that, for $|\Phi_T\rangle$ and $|\Phi_0\rangle$ to be distinguishable with $\Omega(1)$ bias, we must have $T = \Omega\left(\sqrt{\frac{2^n}{m+1}}\right)$. ■

Using Theorem 3.3, we can straightforwardly show a quantum oracle separation between QMA and QCMA.

Proof of Theorem 1.1. Let L be a unary language chosen uniformly at random. The oracle $U = \{U_n\}_{n\ge1}$ is as follows: if $0^n \in L$, then $U_n|\psi_n\rangle = -|\psi_n\rangle$ for some n-qubit marked state $|\psi_n\rangle$ chosen uniformly at random, while $U_n|\varphi\rangle = |\varphi\rangle$ whenever $\langle\varphi|\psi_n\rangle = 0$. Otherwise, if $0^n \notin L$, then U_n is the n-qubit identity operation.

Almost by definition, $L \in \mathsf{QMA}^U$. For given a quantum witness $|\varphi\rangle$, the QMA verifier first prepares the state $\frac{1}{\sqrt{2}}(|0\rangle|\varphi\rangle + |1\rangle|\varphi\rangle)$, then applies U_n to the second register conditioned on the first register being $|1\rangle$. Next the verifier applies a Hadamard gate to the first register, measures it, and accepts if and only if $|1\rangle$ is observed. If $0^n \in L$, then there exists a witness—namely $|\varphi\rangle = |\psi_n\rangle$—that causes the verifier to accept with probability 1. On the other hand, if $0^n \notin L$, then *no* witness causes the verifier to accept with nonzero probability.

On the other hand, we claim that $L \notin \mathsf{QCMA}^U$ with probability 1 over the choice of L and U. This can be seen as follows. Fix a QCMA machine M, and let $S_M(n)$ be the event that M^U *succeeds* on 0^n: that is, either $0^n \in L$ and there exists a string w such that M^U accepts $|0^n\rangle|w\rangle$ with probability at least $2/3$, or $0^n \notin L$ and M^U accepts

$|0^n\rangle |w\rangle$ with probability at most $1/3$ for all w. Then Theorem 3.3 readily implies that there exists a positive integer N such that for all $n \geq N$,

$$\Pr_{L,U}\left[S_M(n) \mid S_M(1), \ldots, S_M(n-1)\right] \leq \frac{2}{3}.$$

Hence

$$\Pr_{L,U}\left[S_M(1) \wedge S_M(2) \wedge \cdots\right] = 0.$$

Now, because of the Solovay-Kitaev Theorem [20], the number of possible QCMA machines is only countably infinite. So by the union bound,

$$\Pr_{L,U}\left[\exists M : S_M(1) \wedge S_M(2) \wedge \cdots\right] = 0$$

as well. ■

We can similarly show a quantum oracle separation between $\mathsf{BQP}/\mathsf{qpoly}$ and $\mathsf{BQP}/\mathsf{poly}$.

Theorem 3.4 *There exists a quantum oracle U such that $\mathsf{BQP}^U/\mathsf{qpoly} \neq \mathsf{BQP}^U/\mathsf{poly}$.*

Proof. In this case U_n will act on $2n$ qubits. Let L be a binary language chosen uniformly at random, and let $L(x) = 1$ if $x \in L$ and $L(x) = 0$ otherwise. Also, for all n, let $|\psi_n\rangle$ be an n-qubit state chosen uniformly at random. Then U_n acts as follows: for all $x \in \{0,1\}^n$,

$$U_n(|\psi_n\rangle |x\rangle) = (-1)^{L(x)} |\psi_n\rangle |x\rangle,$$

but $U_n(|\phi\rangle |x\rangle) = |\phi\rangle |x\rangle$ whenever $\langle\phi|\psi_n\rangle = 0$. Clearly $L \in \mathsf{BQP}^U/\mathsf{qpoly}$; we just take $|\psi_n\rangle$ as the advice. On the other hand, by essentially the same argument as for Theorem 1.1, one can show that $L \notin \mathsf{BQP}^U/\mathsf{poly}$ with probability 1 over L and U. ■

3.1 Proof of Geometric Lemma

In this section we fill in the proof of Lemma 3.2, thereby completing the oracle separation theorems.

In proving Lemma 3.2, the first step is to ask the following question: among all p-uniform probability measures σ, which is the one that maximizes $\mathrm{E}_{|\psi\rangle\in\sigma}\left[|\langle\psi|0\rangle|^2\right]$? We can think of the set of quantum states $\mathbb{CP}^{N-1}$ as a container, which contains a fluid σ that is gravitationally attracted to the state $|0\rangle$. Then intuitively, the answer is clear: the way to maximize $\mathrm{E}_{|\psi\rangle\in\sigma}\left[|\langle\psi|0\rangle|^2\right]$ is to "fill the container from the bottom," subject to the density constraint $p\sigma \leq \mu$. In other words, the optimal σ should be the uniform measure over the region $\mathcal{R}(p)$ given by $|\langle\psi|0\rangle| \geq h(p)$, where $h(p)$ is chosen so that the volume of $\mathcal{R}(p)$ is a p fraction of the total volume of $\mathbb{CP}^{N-1}$. The following lemma makes this intuition rigorous.

Lemma 3.5 *Among all p-uniform probability measures σ over $\mathbb{CP}^{N-1}$, the one that maximizes $\mathrm{E}_{|\psi\rangle\in\sigma}\left[|\langle\psi|0\rangle|^2\right]$ is $\tau(p)$, the uniform measure over the region $\mathcal{R}(p)$ defined above.*

Proof. Since $|\langle\psi|0\rangle|^2$ is nonnegative, we can write

$$\mathop{\mathrm{E}}_{|\psi\rangle\in\sigma}\left[|\langle\psi|0\rangle|^2\right] = \int_0^\infty \Pr_{|\psi\rangle\in\sigma}\left[|\langle\psi|0\rangle|^2 \geq y\right] dy.$$

We claim that setting $\sigma := \tau(p)$ maximizes the integrand for every value of y. Certainly, then, setting $\sigma := \tau(p)$ maximizes the integral itself as well.

To prove the claim, we consider two cases. First, if $y \leq h(p)^2$, then

$$\Pr_{|\psi\rangle\in\tau(p)}\left[|\langle\psi|0\rangle|^2 \geq y\right] = 1,$$

which is certainly maximal. Second, if $y > h(p)^2$, then

$$\Pr_{|\psi\rangle\in\tau(p)}\left[|\langle\psi|0\rangle|^2 \geq y\right] = \frac{1}{p}\cdot\Pr_{|\psi\rangle\in\mu}\left[|\langle\psi|0\rangle|^2 \geq y\right].$$

This is maximal as well, since

$$\Pr_{|\psi\rangle\in\sigma}\left[|\langle\psi|0\rangle|^2 \geq y\right] \leq \frac{1}{p}\cdot\Pr_{|\psi\rangle\in\mu}\left[|\langle\psi|0\rangle|^2 \geq y\right].$$

for all p-uniform probability measures σ. ■

Lemma 3.5 completely describes the probability measure that maximizes $\mathrm{E}_{|\psi\rangle\in\sigma}\left[|\langle\psi|0\rangle|^2\right]$, except for one detail: the value of $h(p)$ (or equivalently, the radius of $\mathcal{R}(p)$). The next lemma completes the picture.

Lemma 3.6 *For all* p,

$$h(p) = \sqrt{1-p^{1/(N-1)}} = \Theta\left(\sqrt{\frac{\log 1/p}{N}}\right).$$

Proof. We will show that for all h,

$$\Pr_{|\psi\rangle\in\mu}\left[|\langle\psi|0\rangle| \geq h\right] = \left(1-h^2\right)^{N-1},$$

where μ is the uniform probability measure over $\mathbb{CP}^{N-1}$. Setting $p := \Pr_{|\psi\rangle\in\mu}\left[|\langle\psi|0\rangle| \geq h\right]$ and solving for h then yields the lemma.

Let $\overrightarrow{z} = (z_0,\ldots,z_{N-1})$ be a complex vector; then let $\overrightarrow{r} = (r_0,\ldots,r_{N-1})$ and $\overrightarrow{\theta} = (\theta_0,\ldots,\theta_{N-1})$ be real vectors such that $z_k = r_k e^{i\theta_k}$ for each coordinate k. Also, let $\mathcal{D}$ be a Gaussian probability measure on $\mathbb{C}^N$, with density function

$$P(\overrightarrow{z}) = P(\overrightarrow{r}) = \frac{1}{\pi^N}e^{-\|\overrightarrow{r}\|_2^2}.$$

Let $d\overrightarrow{r}$ be shorthand for $dr_0\cdots dr_{N-1}$. Then we can express the probability that $|\langle\psi|0\rangle| \geq h$ as

$$\begin{aligned}
\Pr_{|\psi\rangle\in\mu}\left[|\langle\psi|0\rangle| \geq h\right] &= \Pr_{\overrightarrow{z}\in\mathcal{D}}\left[|z_0| \geq h\,\|\overrightarrow{z}\|_2\right] \\
&= \Pr_{\overrightarrow{r},\overrightarrow{\theta}}\left[r_0 \geq h\,\|\overrightarrow{r}\|_2\right] \\
&= \int_{\overrightarrow{r},\overrightarrow{\theta}\,:\,r_0\geq h\|\overrightarrow{r}\|_2} P(\overrightarrow{r})\; r_0\cdots r_{N-1}\,d\overrightarrow{r}\,d\overrightarrow{\theta} \\
&= (2\pi)^N \int_{\overrightarrow{r}\,:\,r_0\geq h\|\overrightarrow{r}\|_2} \frac{1}{\pi^N}e^{-\|\overrightarrow{r}\|_2^2}\, r_0\cdots r_{N-1}\,d\overrightarrow{r} \\
&= \int_{r_1,\ldots,r_{N-1}=0}^{\infty}\left(\int_{r_0=h\sqrt{\frac{r_1^2+\cdots+r_{N-1}^2}{1-h^2}}}^{\infty} 2e^{-r_0^2}r_0\,dr_0\right)2^{N-1}e^{-r_1^2-\cdots-r_{N-1}^2}\; r_1dr_1\cdots r_{N-1}dr_{N-1} \\
&= \int_{r_1,\ldots,r_{N-1}=0}^{\infty} e^{-\left(r_1^2+\cdots+r_{N-1}^2\right)\cdot h^2/\left(1-h^2\right)}2^{N-1}e^{-r_1^2-\cdots-r_{N-1}^2}\; r_1dr_1\cdots r_{N-1}dr_{N-1} \\
&= \int_{r_1,\ldots,r_{N-1}=0}^{\infty} 2^{N-1}e^{-\left(r_1^2+\cdots+r_{N-1}^2\right)/\left(1-h^2\right)}\; r_1dr_1\cdots r_{N-1}dr_{N-1} \\
&= \left(\int_{r=0}^{\infty} 2e^{-r^2/\left(1-h^2\right)}r\,dr\right)^{N-1} \\
&= \left(1-h^2\right)^{N-1}.
\end{aligned}$$

■

By combining Lemmas 3.5 and 3.6, we can now prove Lemma 3.2: that if σ is p-uniform, then for all mixed states ρ,

$$\operatorname*{E}_{|\psi\rangle\in\sigma}\left[\langle\psi|\rho|\psi\rangle\right]=O\left(\frac{1+\log 1/p}{N}\right).$$

Proof of Lemma 3.2. If $p\leq e^{-\Omega(N)}$ then the lemma is certainly true, so suppose $p\geq e^{-O(N)}$. Since the concluding inequality is linear in ρ, we can assume without loss of generality that ρ is a pure state. Indeed, by symmetry we can assume that $\rho=|0\rangle\langle 0|$. So our aim is to upper-bound $\operatorname{E}_{|\psi\rangle\in\sigma}\left[|\langle\psi|0\rangle|^2\right]$, where σ is any p-uniform probability measure. By Lemma 3.5, we can assume without loss of generality that $\sigma=\tau(p)$ is the uniform measure over all $|\psi\rangle$ such that $|\langle\psi|0\rangle|\geq h(p)$. Then letting

$$\begin{aligned}|\psi\rangle&=\alpha_0|0\rangle+\cdots+\alpha_{N-1}|N-1\rangle,\\ r&=\sqrt{|\alpha_1|^2+\cdots+|\alpha_{N-1}|^2},\end{aligned}$$

we have

$$\begin{aligned}\operatorname*{E}_{|\psi\rangle\in\tau(p)}\left[|\langle\psi|0\rangle|^2\right]&=\operatorname*{E}_{|\psi\rangle\,:\,|\alpha_0|\geq h(p)}\left[|\alpha_0|^2\right]\\ &=\operatorname*{E}_{|\psi\rangle\,:\,r^2\leq 1-h(p)^2}\left[1-r^2\right]\\ &=\frac{\int_0^{\sqrt{1-h(p)^2}}r^{2N-3}\left(1-r^2\right)dr}{\int_0^{\sqrt{1-h(p)^2}}r^{2N-3}dr}\\ &=\frac{\left[\frac{r^{2N-2}}{2N-2}-\frac{r^{2N}}{2N}\right]_0^{\sqrt{1-h(p)^2}}}{\left[\frac{r^{2N}}{2N}\right]_0^{\sqrt{1-h(p)^2}}}\\ &=\frac{1-\left(1-\frac{1}{N}\right)\left(1-h(p)^2\right)}{\left(1-\frac{1}{N}\right)\left(1-h(p)^2\right)}\\ &=O\left(\frac{1}{N}+h(p)^2\right)\\ &=O\left(\frac{1+\log 1/p}{N}\right),\end{aligned}$$

where the last line follows from Lemma 3.6. ■

4 Upper Bound

In this section we show that the lower bound of Theorem 3.3 is basically tight. In particular, let U be an n-qubit quantum oracle, and suppose we are given an m-bit classical proof that U is not the identity, but instead conceals a marked state $|\psi\rangle$ such that $U|\psi\rangle=-|\psi\rangle$. Then provided $2n\leq m\leq 2^n$, a quantum algorithm can verify the proof by making $O\left(\sqrt{2^n/m}\right)$ oracle calls to U. This matches our lower bound when $m\geq 2n$.[4]

Let $N=2^n$ be the dimension of U's Hilbert space. Then the idea of our algorithm is to use a "mesh" of states $|\phi_1\rangle,\ldots,|\phi_M\rangle\in\mathbb{CP}^{N-1}$, at least one of which has nontrivial overlap with every pure state in $\mathbb{CP}^{N-1}$. A classical proof can then help the algorithm by telling it the $|\phi_i\rangle$ that is closest to $|\psi\rangle$. More formally, define the h-*ball* about

[4] When $m\ll 2n$, the best upper bound we know is the trivial $O\left(\sqrt{2^n}\right)$. However, we conjecture that $O\left(\sqrt{2^n/m}\right)$ is achievable in this case as well.

$|\phi\rangle$ to be the set of $|\varphi\rangle$ such that $|\langle\phi|\varphi\rangle| \geq h$. Then define an *$h$-net for $\mathbb{CP}^{N-1}$ of size M* to be a set of states $|\phi_1\rangle, \ldots, |\phi_M\rangle$ such that every $|\psi\rangle \in \mathbb{CP}^{N-1}$ is contained in the h-ball about $|\phi_i\rangle$ for some i.[5] We will use the following theorem, which follows from Corollary 1.2 of Böröczky and Wintsche [11].

Theorem 4.1 ([11]) *For all $0 < h < 1$, there exists an h-net for $\mathbb{CP}^{N-1}$ of size*

$$O\left(\frac{N^{3/2}\log\left(2+Nh^2\right)}{\left(1-h^2\right)^N}\right).$$

Böröczky and Wintsche do not provide an explicit construction of such an h-net; they only prove that it exists.[6] Later, we will give an explicit construction with only slightly worse parameters than those of Theorem 4.1. But first, let us prove an upper bound on query complexity.

Theorem 4.2 *Suppose we have an n-qubit quantum oracle U such that either (i) $U = U_\psi$ for some $|\psi\rangle$, or (ii) $U = I$ is the identity operator. Then given an m-bit classical witness in support of case (i), where $m \geq 2n$, there exists a quantum algorithm that verifies the witness using $O\left(\sqrt{2^n/m}+1\right)$ queries to U.*

Proof. By Theorem 4.1, there exists an h-net $\mathcal{S}$ for $\mathbb{CP}^{2^n-1}$ of cardinality

$$|\mathcal{S}| = O\left(\frac{2^{3n/2}\log\left(2+2^n h^2\right)}{\left(1-h^2\right)^{2^n}}\right).$$

Setting $|\mathcal{S}| = 2^m$ gives

$$m \leq \frac{3n}{2} + 2^n \log\left(\frac{1}{1-h^2}\right) + O\left(\log n\right).$$

Solving for h, we obtain

$$h \geq \sqrt{\frac{m - 3n/2 - O\left(\log n\right)}{2^n}},$$

which is $\Omega\left(\sqrt{m/2^n}\right)$ provided $m \geq 2n$. So there exists a collection of $M = 2^m$ states, $|\phi_1\rangle, \ldots, |\phi_M\rangle \in \mathbb{CP}^{2^n-1}$, such that for every $|\psi\rangle$, there exists an i such that $|\langle\phi_i|\psi\rangle| \geq h$ where $h = \Omega\left(\sqrt{m/2^n}\right)$.

Given an oracle $U = U_{|\psi\rangle}$, the classical witness $w \in \{0,1\}^m$ will simply encode an index i such that $|\langle\phi_i|\psi\rangle| \geq h$. If we prepare $|\phi_i\rangle$ and feed it to U, then the probability of finding the marked state $|\psi\rangle$ is $|\langle\phi_i|\psi\rangle|^2 \geq h^2$. Furthermore, if we do find $|\psi\rangle$, we will know we did (i.e. a control qubit will be $|1\rangle$ instead of $|0\rangle$). From these facts, it follows immediately from the amplitude amplification theorem of Grover [15] and Brassard et al. [12] that we can find $|\psi\rangle$ with probability $\Omega\left(1\right)$ using

$$O\left(\sqrt{\frac{1}{h^2}}+1\right) = O\left(\sqrt{\frac{2^n}{m}}+1\right)$$

queries to U. ■

Of course, if we care about *computational* complexity as well as query complexity, then it is not enough for an h-net to exist—we also need the states in the h-net to be efficiently preparable. Fortunately, proving an explicit version of Theorem 4.1 turns out to be simpler than one might expect. We will do so with the help of the following inequality.

Lemma 4.3 *Let $x_1 \geq \cdots \geq x_N \geq 0$ be nonnegative real numbers with $x_1^2 + \cdots + x_N^2 = 1$. Then for all $k \in \{1, \ldots, N\}$,*

$$\max_{1 \leq t \leq k}\left[\frac{x_1 + \cdots + x_t}{\sqrt{t}}\right] \geq \sqrt{\frac{k}{N\lceil\log_2 N\rceil}}.$$

[5] These objects are often called ε-nets, with the obvious relation $h = \cos\varepsilon$.

[6] Note that we cannot just start from an explicit construction of a sphere-packing, and then double the radius of the spheres to get a covering. We could do that if we wanted a covering of $\mathbb{CP}^{N-1}$ by *small* balls. But in our case, h is close to zero, which means that the balls already have close to the maximal radius.

Proof. Let $L = \lceil \log_2 N \rceil$. Then for all $i \in \{1, \dots, L\}$, let $s_i = x_{2^{i-1}}^2 + \cdots + x_{2^i-1}^2$, where we adopt the convention that $x_j = 0$ if $j > N$. Then

$$s_1 + \cdots + s_L = x_1^2 + \cdots + x_N^2 = 1,$$

so certainly there exists an $i \in \{1, \dots, L\}$ such that $s_i \geq 1/L$. Fix that i. Then since the x_j's are arranged in nonincreasing order, we have

$$x_{2^{i-1}} \geq \sqrt{\frac{s_i}{2^{i-1}}} \geq \sqrt{\frac{1}{2^{i-1}L}}.$$

There are now two cases. First, if $k \leq 2^{i-1}$ then

$$\max_{1 \leq t \leq k} \left[\frac{x_1 + \cdots + x_t}{\sqrt{t}} \right] \geq \frac{x_1 + \cdots + x_k}{\sqrt{k}} \geq \frac{k}{\sqrt{k}} x_{2^{i-1}} \geq \sqrt{\frac{k}{2^{i-1}L}} \geq \sqrt{\frac{k}{N \lceil \log_2 N \rceil}}.$$

Second, if $2^{i-1} \leq k$ then

$$\max_{1 \leq t \leq k} \left[\frac{x_1 + \cdots + x_t}{\sqrt{t}} \right] \geq \frac{x_1 + \cdots + x_{2^{i-1}}}{\sqrt{2^{i-1}}} \geq \frac{2^{i-1}}{\sqrt{2^{i-1}}} x_{2^{i-1}} \geq \sqrt{\frac{1}{L}} \geq \sqrt{\frac{k}{N \lceil \log_2 N \rceil}}.$$

This completes the proof.[7] ■

We now use Lemma 4.3 to construct an h-net.

Theorem 4.4 *For all $0 < h < 1$, there exists an h-net $|\phi_1\rangle, \dots, |\phi_M\rangle$ for $\mathbb{CP}^{N-1}$ of size $M = 4N \cdot 2^{O\left(h^2 N \log^2 N\right)}$, as well as a quantum algorithm that runs in time polynomial in $\log M$ and that prepares the state $|\phi_i\rangle$ given i as input.*

Proof. Assume without loss of generality that $N = 2^n$ and $M = 2^m$ are both powers of 2, and let $|\psi\rangle$ be an n-qubit target state. Then it suffices to show that a quantum algorithm, using

$$m = \log_2 M = n + 2 + O\left(h^2 2^n n^2\right)$$

bits of classical advice, can prepare a state $|\phi\rangle$ such that $|\langle\phi|\psi\rangle| \geq h$ in time polynomial in m.

Let $k := \left\lfloor \frac{m}{n+2} \right\rfloor$. Also, let us express $|\psi\rangle$ in the computational basis as

$$|\psi\rangle = \sum_{z \in \{1,\dots,N\}} \alpha_z |z\rangle,$$

and let $|z_1\rangle, \dots, |z_N\rangle$ be an ordering of basis states with the property that $|\alpha_{z_1}| \geq \cdots \geq |\alpha_{z_N}|$. Then by Lemma 4.3, there exists an integer $t \in \{1, \dots, k\}$ such that

$$\frac{|\alpha_{z_1}| + \cdots + |\alpha_{z_t}|}{\sqrt{t}} \geq \sqrt{\frac{k}{N \lceil \log_2 N \rceil}} = \sqrt{\frac{k}{Nn}}.$$

Here we can assume that $\alpha_{z_1}, \dots, \alpha_{z_t}$ are all nonzero, since otherwise we simply decrease t. Now let β_z be the element of $\{1, -1, i, -i\}$ that is closest to $\alpha_z / |\alpha_z|$, with ties broken arbitrarily. Then our approximation to $|\psi\rangle$ will be the following:

$$|\phi\rangle := \frac{1}{\sqrt{t}} \sum_{i=1}^{t} \beta_{z_i} |z_i\rangle.$$

[7] One might wonder whether the $\sqrt{1/\lceil \log_2 N \rceil}$ factor can be eliminated. However, a simple example shows that it can be improved by at most a constant factor. Suppose $x_j := \sqrt{\frac{1}{jw}}$, where $w = \sum_{j=1}^{n} \frac{1}{j} \approx \ln N$. Then for all $t \in \{1, \dots, N\}$, we have

$$\frac{x_1 + \cdots + x_t}{\sqrt{t}} \approx \frac{2}{\sqrt{\ln N}}.$$

To specify $|\phi\rangle$, the classical advice just needs to list $z_1, \ldots, z_t$ and $\beta_{z_1}, \ldots, \beta_{z_t}$. Since $t \leq k$, this requires at most $k(n+2) \leq m$ bits. Given the specification, it is clear that $|\phi\rangle$ can be prepared in time polynomial in $tn \leq m$. Moreover,

$$\langle\phi|\psi\rangle = \frac{1}{\sqrt{t}} \sum_{i=1}^{t} \beta_{z_i}^{*} \alpha_{z_i} \geq \frac{1}{\sqrt{t}} \sum_{i=1}^{t} \frac{|\alpha_{z_i}|}{\sqrt{2}} \geq \sqrt{\frac{k}{2Nn}}.$$

We can therefore set $h := \sqrt{\frac{k}{2Nn}}$, so that $k = 2h^2 Nn$. Hence

$$m \leq (n+2)(k+1) = (n+2)\left(2h^2 Nn + 1\right) = n + 2 + O\left(h^2 2^n n^2\right).$$

■

The following is an immediate consequence of Theorem 4.4.

Corollary 4.5 *Suppose we have an n-qubit quantum oracle U such that either (i) $U = U_\psi$ for some $|\psi\rangle$, or (ii) $U = I$ is the identity. Then given an m-bit classical witness in support of case (i), there exists a quantum algorithm that verifies the witness using $O\left(n\sqrt{2^n/m} + 1\right)$ queries to U, together with $O\left(n^2\sqrt{2^n/m} + \operatorname{poly}(m)\right)$ steps of auxiliary computation.*

It is natural to ask whether we could construct a smaller explicit h-net, and thereby improve the query complexity in Corollary 4.5 from $O\left(n\sqrt{2^n/m} + 1\right)$ to the optimal $O\left(\sqrt{2^n/m} + 1\right)$. We certainly believe that this is possible, but it seems to require more complicated techniques from the theory of sphere coverings.

5 Group Non-Membership

The Group Non-Membership (GNM) problem is defined as follows. We are given a finite group G, a subgroup $H \leq G$, and an element $x \in G$. The problem is to decide whether $x \notin H$.

But how are G, H, and x specified? To abstract away the details of this question, we will use Babai and Szemerédi's model of *black-box groups* [8]. In this model, we know generators for H, and we know how to multiply and invert the elements of G, but we "do not know anything else." More formally, we are given access to a group oracle $\mathcal{O}$, which represents each element $x \in G$ by a randomly-chosen label $\ell(x) \in \{0,1\}^n$ for some $n \gg \log_2 |G|$. We are also given the labels of generators $\langle h_1, \ldots, h_l\rangle$ for H. We are promised that every element has a unique label.

Suppose that our quantum computer's state has the form

$$|\Phi\rangle = \sum_{x,y \in G,\ z} \alpha_{x,y,z} |\ell(x), \ell(y)\rangle |z\rangle,$$

where $\ell(x)$ and $\ell(y)$ are labels of group elements and $|z\rangle$ is a workspace register. Then the oracle $\mathcal{O}$ maps this state to

$$\mathcal{O}|\Phi\rangle = \sum_{x,y \in G,\ z} \alpha_{x,y,z} \left|\ell(x), \ell\left(xy^{-1}\right)\right\rangle |z\rangle.$$

Note that if the first register does not contain valid labels of group elements, then $\mathcal{O}$ can behave arbitrarily. Thus, from now on we will ignore labels, and talk directly about the group elements they represent. Using $\mathcal{O}$, it is easy to see that we can perform group inversion (by putting the identity element e in the x register) and multiplication (by first inverting y, then putting y^{-1} in the y register), as well as any combination of these operations.

We will show that GNM has polynomially-bounded QCMA query complexity. In other words, if $x \notin H$, then Merlin can provide Arthur with a $\operatorname{poly}(n)$-bit classical witness of that fact, which enables Arthur to verify it with high probability using $\operatorname{poly}(n)$ quantum queries to the group oracle $\mathcal{O}$.

To prove this result, we first need to collect various facts from finite group theory. Call $g_1, \ldots, g_k$ an *efficient generating set* for a finite group G if (i) $k = O(\log|G|)$, and (ii) every $x \in G$ is expressible as $g_1^{e_1} \cdots g_k^{e_k}$ where $e_1, \ldots, e_k \in \{0,1\}$. The following lemma was shown by Babai and Erdős [7].

Lemma 5.1 ([7]) *Every finite group G has an efficient generating set.*

Given finite groups Γ and G, we say that functions $f, g : \Gamma \to G$ are *ε-close* if

$$\Pr_{x \in \Gamma} [f(x) \neq g(x)] \leq \varepsilon.$$

Also, recall that $f : \Gamma \to G$ is a homomorphism if $f(xy) = f(x) f(y)$ for all $x, y \in \Gamma$. The following two propositions relate ε-closeness to homomorphisms.

Proposition 5.2 *If two homomorphisms $f, g : \Gamma \to G$ are $(1/2 - \varepsilon)$-close for any $\varepsilon > 0$, then $f = g$.*

Proof. Fix $x \in \Gamma$; then for all $y \in \Gamma$, we have $f(x) = f(y) f\left(y^{-1}x\right)$ and $g(x) = g(y) g\left(y^{-1}x\right)$. By the union bound,

$$\Pr_{y \in \Gamma}\left[f(y) = g(y) \wedge f\left(y^{-1}x\right) = g\left(y^{-1}x\right)\right] \geq 1 - \Pr_{y \in \Gamma}[f(y) \neq g(y)] - \Pr_{y \in \Gamma}\left[f\left(y^{-1}x\right) \neq g\left(y^{-1}x\right)\right] > 0.$$

Hence there exists a y such that $f(y) = g(y)$ and $f\left(y^{-1}x\right) = g\left(y^{-1}x\right)$. But this implies that $f(x) = g(x)$. ■

In particular, Proposition 5.2 implies that if a function f is $1/5$-close to a homomorphism, then it is $1/5$-close to a *unique* homomorphism ($1/5$ being an arbitrary constant less than $1/4$).

Proposition 5.3 (Ben-Or et al. [9]) *Given finite groups Γ and G, a function $f : \Gamma \to G$, and a real number $\varepsilon > 0$, if*

$$\Pr_{x, y \in \Gamma} [f(xy) \neq f(x) f(y)] \leq \varepsilon$$

then f is ε-close to a homomorphism.

Together, Propositions 5.2 and 5.3 have the following easy corollary.

Corollary 5.4 *There is a randomized algorithm which, given finite groups Γ and G and a function $f : \Gamma \to G$ as input, makes $O(1)$ oracle queries to f, accepts with probability 1 if f is a homomorphism, and rejects with probability at least $2/3$ if f is not $1/5$-close to a homomorphism. Also, if f is $1/5$-close to some homomorphism $\widetilde{f}$, then there exists a randomized algorithm that, given an input $x \in \Gamma$, makes $O(r)$ oracle queries to f, and outputs $\widetilde{f}(x)$ with probability at least $1 - 1/2^r$.*

In the present context, our algorithms are not limited in space or time, and we can say for simplicity that Γ is represented by its entire multiplication table. It is then easy, as the proof will require, to pick elements of Γ uniformly at random. By contrast, G is represented by oracle access, but there will be no need to choose its elements at random. But note that even if access to Γ were restricted by an oracle model, it is still possible to pick elements at random using an algorithm of Babai [5]; this is relevant to the discussion in Section 5.1.

Proof. The first algorithm simply chooses $O(1)$ pairs $x, y \in \Gamma$ uniformly at random, accepts if $f(xy) = f(x) f(y)$ for all of them, and rejects otherwise. Let $k = O(r)$. Then the second algorithm chooses $z_1, \ldots, z_k \in \Gamma$ uniformly at random, and outputs the plurality answer among $f(z_1) f\left(z_1^{-1}x\right), \ldots, f(z_k) f\left(z_k^{-1}x\right)$ (breaking ties arbitrarily). ■

It follows from the Classification of Finite Simple Groups that there are at most two finite simple groups of any particular order (see [13] for example). The following well-known result is a combination of that fact and of a theorem due to Neumann [24].

Theorem 5.5 *There are $N^{O\left((\log_2 N)^2\right)}$ groups of order N up to isomorphism.*[8]

[8] The most accurate asymptotic result on the number of groups of order N, in terms of the prime factorization of N, appears in a paper by Pyber [25].

Finally, recall that the Hidden Subgroup Problem (HSP) is defined as follows. We are given a finite group G, and oracle access to a function $f : G \to \mathbb{Z}$. We are promised that there exists a "hidden subgroup" $H \leq G$ such that $f(x) = f(y)$ if and only if x and y belong to the same left coset of H. The problem is then to output a set of generators for H. Whether HSP can be solved in quantum polynomial time, for various non-abelian groups G, is one of the most actively studied questions in quantum computing. However, if we only care about query complexity, then Ettinger, Høyer, and Knill [14] proved the following useful result.

Theorem 5.6 ([14]) *There is a quantum algorithm such that, given any finite group G as oracular input, solves* HSP *using only* $\operatorname{polylog}(|G|)$ *quantum queries to f (together with a possibly exponential amount of postprocessing).*[9]

We can now prove Theorem 1.2: that GNM has polynomially-bounded QCMA query complexity.

Proof of Theorem 1.2. Let G be a group of order at most 2^n, and let $\mathcal{O}$ be a group oracle that maps each element of G to an n-bit label. Also, given (the labels of) group elements $x, h_1, \ldots, h_m \in G$, let H be the subgroup of G generated by $\langle h_1, \ldots, h_m \rangle$. Then the problem is to decide if $x \notin H$.

In our QCMA protocol for this problem, Merlin's witness will consist of the following:

- An explicit "model group" Γ, of order at most 2^n.
- A list of elements $\gamma_1, \ldots, \gamma_k \in \Gamma$, where $k = O(\log |\Gamma|)$.
- A corresponding list $g_1, \ldots, g_k \in G$.
- Another list $z, \lambda_1, \ldots, \lambda_m \in \Gamma$.

We should be more explicit about the notion of an "explicit" group Γ, and about the syntax of this witness. By Theorem 5.5, there are at most $2^{\operatorname{poly}(n)}$ groups of order $|\Gamma| \leq 2^n$ up to isomorphism. Since Arthur is allowed unlimited computatoin and is only restricted in queries, he can construct a full multiplication table for Γ using only the name of its isomorphism type. The multiplication table is not unique, because the elements of Γ can be permuted; but for instance Arthur could construct the lexicographically first such table. Since Merlin can anticipate Arthur's construction of Γ, he can then refer to elements of Γ using the same construction. He can also refer to elements of G since he understands the oracle. In conclusion, Merlin can specify the witness using only $\operatorname{poly}(n)$ bits.

If Merlin is honest, then the witness will have the following three properties:

(1) $\gamma_1, \ldots, \gamma_k$ is an efficient generating set for Γ.

(2) $z \notin \Lambda$, where Λ is the subgroup of Γ generated by $\langle \lambda_1, \ldots, \lambda_m \rangle$.

(3) There exists an embedding $\widetilde{f} : \Gamma \to G$, such that (i) $\widetilde{f}(\gamma_i) = g_i$ for all $i \in \{1, \ldots, k\}$, (ii) $\widetilde{f}(\lambda_j) = h_j$ for all $j \in \{1, \ldots, m\}$, and (iii) $\widetilde{f}(z) = x$.

Suppose for the moment that (1)-(3) all hold. Then there exists an embedding $\widetilde{f} : \Gamma \to G$, which maps the set $\langle \gamma_1, \ldots, \gamma_k \rangle$ in Γ to the set $\langle g_1, \ldots, g_k \rangle$ in G. Furthermore, this embedding satisfies $\widetilde{f}(\Lambda) = H$ and $\widetilde{f}(z) = x$. Since $z \notin \Lambda$ by (2), it follows that $x \notin H$ as well, which is what Arthur wanted to check.

So it suffices to verify (1)-(3). In the remainder of the proof, we will explain how to do this using a possibly exponential amount of computation, but only $\operatorname{poly}(n)$ quantum queries to the group oracle $\mathcal{O}$.

First, since properties (1) and (2) only involve the explicit group Γ, not the black-box group G, Arthur can verify these properties "free of cost." In other words, regardless of how much computation he needs, he never has to query the group oracle.

The nontrivial part is to verify (3). It will be convenient to split (3) into the following sub-claims:

(3a) There exists a homomorphism $\widetilde{f} : \Gamma \to G$ such that $\widetilde{f}(\gamma_i) = g_i$ for all $i \in \{1, \ldots, k\}$.

[9] Indeed, for *Normal* HSP (which is the special case we care about), Hallgren, Russell, and Ta-Shma [16] improved this result, showing how to find a hidden subgroup using only $O(\log |G|)$ queries to f (again, with exponential postprocessing).

(3b) $\widetilde{f}$ satisfies $\widetilde{f}(z) = x$ and $\widetilde{f}(\lambda_j) = h_j$ for all $j \in \{1, \ldots, m\}$.

(3c) $\widetilde{f}$ is injective (i.e. is an embedding into G).

To verify (3a), first Arthur fixes a "canonical representation" of each element $\gamma \in \Gamma$. This representation has the form

$$\gamma = \gamma_1^{e_1} \cdots \gamma_k^{e_k},$$

where $\langle \gamma_1, \ldots, \gamma_k \rangle$ is the efficient generating set for Γ, and $e_1, \ldots, e_k \in \{0, 1\}$ are bits depending on γ. Next he defines a function $f : \Gamma \to G$ by

$$f(\gamma) := g_1^{e_1} \cdots g_k^{e_k}$$

for all $\gamma \in \Gamma$. By using the canonical representation of γ, Arthur can evaluate $f(\gamma)$ using at most $k - 1$ queries to the group oracle $\mathcal{O}$. Finally Arthur appeals to Corollary 5.4. If f is not $1/5$-close to a homomorphism, then by using $O(1)$ queries to f, with high probability Arthur can detect that f is not a homomorphism. In that case Merlin has been caught cheating, so Arthur rejects. On the other hand, if f is $1/5$-close to some homomorphism $\widetilde{f}$, then by using $O(\log|\Gamma|)$ queries to f, with high probability Arthur can "correct" f to $\widetilde{f}$. In that case it remains only to check that $\widetilde{f}(\gamma_i) = g_i$ for all $i \in \{1, \ldots, k\}$.

Once Arthur has an efficient procedure for computing $\widetilde{f}$—that is, a procedure that involves only $\operatorname{poly}(n)$ queries to $\mathcal{O}$—he can then verify property (3b) directly.

To verify (3c), Arthur runs the algorithm of Ettinger, Høyer, and Knill [14] for the Hidden Subgroup Problem. Notice that, since $\widetilde{f} : \Gamma \to G$ is a homomorphism, there must be a "hidden subgroup" $K \leq \Gamma$—namely the kernel of $\widetilde{f}$—such that $\widetilde{f}$ is constant on cosets of K and distinct on distinct cosets. Furthermore, $\widetilde{f}$ is injective if and only if K is trivial. But deciding whether K is trivial is just an instance of HSP, and can therefore be solved using $\operatorname{poly}(n)$ quantum queries by Theorem 5.6. ■

5.1 Computational Complexity

Theorem 1.2 showed that one can always verify group non-membership using a polynomial-size classical witness, together with polynomially many quantum queries to the group oracle $\mathcal{O}$. Unfortunately, while the *query* complexity is polynomial, the *computational* complexity might be exponential. However, as mentioned in Section 1.1, we conjecture that this shortcoming of Theorem 1.2 can be removed, and that GNM is in QCMA for any group oracle $\mathcal{O}$.

In our QCMA protocol, the main computational problem that needs to be solved is not the general HSP, but rather the Normal Hidden Subgroup Problem (NHSP)—that is, HSP where the hidden subgroup is normal. This is because the kernel of a homomorphism is always a normal subgroup. Hallgren, Russell, and Ta-Shma [16] showed that NHSP is in BQP for an explicit group Γ, provided that the quantum Fourier transform over Γ can be implemented efficiently (and its output can be interpreted). Furthermore, Moore, Rockmore, and Russell [22] showed that many classes of finite groups G have an explicit model $\Gamma \cong G$ for which this assumption holds.

Even if NHSP is in BQP, there are two remaining obstacles to showing that GNM is in QCMA. First, we need to be able to verify group non-membership in the explicit model group Γ, possibly with the help of additional classical information from Merlin. Second, we need an efficient algorithm to compute the function $\widetilde{f} : \Gamma \to G$ for every $\gamma \in \Gamma$, even though $\widetilde{f}$ is explicitly defined only on the generators $\gamma_1, \ldots, \gamma_k$.

Actually, Laszlo Babai pointed out to us (private communication) that there does not exist a notion of an "explicit" model of a finite group which is suitable for all finite groups and all algorithms. What we really mean by "explicit" is a definition such that all steps of our sketched algorithm are efficient. More precisely, we need that for every finite group G, there should exist an isomorphic group $\Gamma \cong G$ with the following properties:

(i) Γ has a fixed list of generators $\gamma_1, \ldots, \gamma_k \in \Gamma$ with $k = O(\operatorname{polylog}|G|)$.

(ii) There is an algorithm to generate an approximately random element $\gamma \in \Gamma$, expressed as a (polylogarithmically) short product of $\gamma_1, \ldots, \gamma_k$. (This may be essentially established; see Babai [5].)

(iii) NHSP over Γ is in BQP. If the Hallgren-Russell-Ta-Shma algorithm is used, there should be an a string encoding of elements of Γ. There must be an efficient way to prepare the constant pure state $|\Gamma\rangle$, which for instance could be based on an efficient membership test. Then there should be an algorithm for the quantum Fourier transform over Γ. In this case the QFT produces randomly-chosen characters of the quotient group Γ/Λ for some normal subgroup Λ (the hidden subgroup). The characters must be encoded so that it is easy to recognize Λ, or at least confirm that it is trivial.

(iv) GNM over Γ is in QCMA. The class QCMA may even be overkill at this step, since in settled examples GNM is in NP or even in P.

If Γ is the symmetric group S_n, or an abelian group expressed as a product of cyclic groups, or if it is a matrix group $\mathrm{GL}(n, \mathbb{Z}/p)$, then the obvious notation for group elements is amenable to the above requirements for explicitness. Moreover, in each of these examples there is an easy generating set that satisfies (i) and (ii) (exercise for the reader). However, for general finite groups even this much is not known. It is known for finite *simple* groups except for Chevalley groups of type G_2, where it is strongly suspected. We may speculate that there is a good explicit model for any non-simple group using its Jordan-Hölder composition series. However, there no known polylogarithmic description of the "glue" in such a composition series (*i.e.*, the data that determines the group extensions); the current best result for solvable groups is quasipolylogarithmic [17].

Nonetheless, all steps have already been completed for several classes of groups. For example, if $\Gamma = \mathbb{Z}/r_1 \times \cdots \times \mathbb{Z}/r_k$ is abelian, then NHSP is in BQP by the work of Shor [28] and Kitaev [19]; GNM is in P by linear algebra. Another example is $\Gamma = S_n$. In this case NHSP is trivial (since the only normal subgroup is A_n) and GNM is in P by the work of Sims [29]. We optimistically conjecture that all steps can be completed for arbitrary finite groups.

6 Mimicking Random Quantum Oracles

We have seen, on the one hand, that there exists a quantum oracle separating QMA from QCMA; and on the other hand, that separating these classes by a *classical* oracle seems much more difficult. Together, these results raise a general question: how much "stronger" are quantum oracles than classical ones? In particular, are there complexity classes $\mathcal{C}$ and $\mathcal{D}$ that can be separated by quantum oracles, but such that separating them by classical oracles is almost as hard as separating them in the unrelativized world? Whatever the answer, we conjecture that QMA and QCMA are *not* examples of such classes. The reason is that it seems possible, using only classical oracles, to approximate quantum oracles similar to ones that would separate QMA from QCMA.

To illustrate, let σ be the uniform probability measure over $2^n \times 2^n$ unitary diagonal matrices. (In other words, each diagonal entry of $D \in \sigma$ is a random complex number with norm 1.) Also, let $H^{\otimes n}$ be a tensor product of n Hadamard matrices. Then let ς_k be the probability measure over $2^n \times 2^n$ unitary matrices

$$U = D_k H^{\otimes n} D_{k-1} H^{\otimes n} \cdots H^{\otimes n} D_1 H^{\otimes n}$$

induced by drawing each D_i independently from σ. In other words, $U \in \varsigma_k$ is obtained by first applying a Hadamard gate to each qubit, then a random $2^n \times 2^n$ diagonal matrix, then Hadamard gates again, then another random diagonal matrix, and so on k times.

Note that we can efficiently apply such a U—at least to polynomially many bits of precision—if given a classical random oracle A. To do so, we simply implement the random diagonal matrix D_i as

$$\sum_{x \in \{0,1\}^n} \alpha_x |x\rangle \rightarrow \sum_{x \in \{0,1\}^n} \omega^{A(i,x)} \alpha_x |x\rangle ,$$

where $A(i,x)$ is a uniformly random n-bit integer indexed by i and x, and $\omega = e^{2\pi i/2^n}$.

Now let μ be the uniform probability measure over $2^n \times 2^n$ unitary matrices. If $k \ll 2^n$, then ς_k is *not* close to μ in variation distance, since the former has only $\Theta(k2^n)$ degrees of freedom while the latter has $\Theta(k4^n)$.[10] On the

[10]Admittedly, it is still conceivable that the finite-precision version of ς_k is close in variation distance to the finite-precision version of μ. However, a more sophisticated argument that counts distinguishable unitaries rules out that possibility as well.

other hand, we conjecture that a U drawn from ς_k will "look random" to any polynomial-time algorithm, and that this property can be used to prove a classical oracle separation between QMA and QCMA.

Let us explain what we mean in more detail. Suppose we are given access to an n-qubit unitary oracle U, and want to decide whether

(i) U was drawn uniformly at random (that is, from μ), or

(ii) U was drawn uniformly at random conditioned on there existing $n/2$-qubit pure states $|\psi\rangle$ and $|\varphi\rangle$ such that $U\left(|0\rangle^{\otimes n/2}|\psi\rangle\right) \approx |0\rangle^{\otimes n/2}|\varphi\rangle$.

In case (i), the states $|\psi\rangle$ and $|\varphi\rangle$ will exist only with negligible probability.[11] It follows that the above problem is in QMA^U—since if case (ii) holds, then a succinct quantum proof of that fact is just $|\psi\rangle$ itself. We now state three conjectures about this problem, in increasing order of difficulty.

Conjecture 6.1 *The above problem is not in QCMA^U. In other words, if case (ii) holds, there is no succinct classical proof of that fact that can be verified with high probability using $\operatorname{poly}(n)$ quantum queries to U.*

Presumably Conjecture 6.1 can be proved using ideas similar to those in Section 3. If so, then the next step is to replace the uniform measure μ by the "pseudorandom" measure ς_k.

Conjecture 6.2 *Suppose that instead of being drawn from μ, the unitary U is drawn from ς_k for some $k = \Omega(n)$. Then the probability that there exist $n/2$-qubit states $|\psi\rangle$ and $|\varphi\rangle$ such that $U\left(|0\rangle^{\otimes n/2}|\psi\rangle\right) \approx |0\rangle^{\otimes n/2}|\varphi\rangle$ is still negligibly small.*

Now suppose we want to decide whether

(i') U was drawn from ς_k, or

(ii') U was drawn from ς_k conditioned on there existing $n/2$-qubit states $|\psi\rangle$ and $|\varphi\rangle$ such that $U\left(|0\rangle^{\otimes n/2}|\psi\rangle\right) \approx |0\rangle^{\otimes n/2}|\varphi\rangle$.

Also, let A be a classical oracle that encodes the diagonal matrices $D_1, \ldots, D_k$ such that

$$U = D_k H^{\otimes n} D_{k-1} H^{\otimes n} \cdots H^{\otimes n} D_1 H^{\otimes n}.$$

If Conjecture 6.2 is true, then case (ii') can be verified in QMA^A. So to obtain a classical oracle separation between QMA and QCMA, the one remaining step would be to prove the following.

Conjecture 6.3 *Case (ii') cannot be verified in QCMA^A.*

6.1 From Random Oracles to Random Unitaries

The previous discussion immediately suggests even simpler questions about the ability of classical oracles to mimic quantum ones. In particular, could a BQP machine use a classical random oracle to prepare a uniformly random n-qubit pure state? Also, could it use such an oracle to apply a random n-qubit unitary?

In this section we answer the first question in the affirmative, and present partial results about the second question. We first need a notion that we call the "ε-smoothing" of a probability measure.

[11] Indeed, the reason we did not ask for $(n-1)$-qubit states $|\psi\rangle$ and $|\varphi\rangle$ such that $U(|0\rangle|\psi\rangle) \approx |0\rangle|\varphi\rangle$ is that such states will exist (almost) generically. For the choice of $|\psi\rangle$ gives us $2^{n-1}-1$ independent complex variables, whereas the requirement that $U(|0\rangle|\psi\rangle)$ have the form $|0\rangle|\varphi\rangle$ imposes only 2^{n-1} constraints. Asking for $(n-2)$-qubit states $|\psi\rangle$ and $|\varphi\rangle$ such that $U(|00\rangle|\psi\rangle) \approx |00\rangle|\varphi\rangle$ might suffice (since now we have $2^{n-2}-1$ variables versus $3 \cdot 2^{n-2}$ constraints), but we wish to stay on the safe side.

Definition 6.4 *Let σ be a probability measure over $|\psi\rangle \in \mathbb{CP}^{2^n-1}$. Then the ε-smoothing of σ, or $\mathcal{S}_\varepsilon(\sigma)$, is the probability measure obtained by first drawing a state $|\psi\rangle$ from σ, and then drawing a state $|\varphi\rangle$ uniformly at random subject to $\langle\varphi|\psi\rangle \geq 1-\varepsilon$.*

Let μ be the uniform measure over $\mathbb{CP}^{2^n-1}$. Also, let Q be a quantum algorithm that queries a classical oracle A. Suppose that, given 0^n as input, Q^A outputs the pure state $|\psi_A\rangle \in \mathbb{CP}^{2^n-1}$. Then we say that Q "approximates the uniform measure within ε" if, as we range over uniform random $A \subseteq \{0,1\}^n$, the induced probability measure σ over $|\psi_A\rangle$ satisfies $\|\mathcal{S}_\varepsilon(\sigma) - \mu\| \leq \varepsilon$.

Theorem 6.5 *For all polynomials p, there exists a quantum algorithm Q that runs in polynomial time, and that approximates the uniform measure within $2^{-p(n)}$.*

Proof Sketch. The algorithm Q is as follows: first prepare a uniform superposition over n-bit strings. Then, using the classical random oracle A as a source of random bits, map this state to

$$|\Psi\rangle = \frac{1}{2^{n/2}} \sum_{x\in\{0,1\}^n} |x\rangle \left(\sqrt{1-|\alpha_x|^2}\,|0\rangle + \alpha_x\,|1\rangle\right),$$

where each α_x is essentially a Gaussian random variable. More precisely, let $q(n) = (n+p(n))^2$. Then each α_x is drawn independently from a complex Gaussian distribution with mean 0 and variance $1/q(n)$, with the two technicalities that (1) α_x is rounded to $q(n)$ bits of precision, and (2) the cutoff $|\alpha_x| \leq 1$ is imposed. (By a tail bound, with overwhelming probability we will have $|\alpha_x| \leq 1$ for all x anyway.)

Next measure the second register of $|\Psi\rangle$ in the standard basis. The outcome $|1\rangle$ will be observed with probability $\Omega(1/q(n))$. Furthermore, conditioned on $|1\rangle$ being observed, one can check that the distribution σ over the reduced state of the first register satisfies $\|\mathcal{S}_{2^{-p(n)}}(\sigma) - \mu\| \leq 2^{-p(n)}$. (We omit the calculation.) Hence it suffices to repeat the algorithm $O(q(n))$ times. ■

Theorem 6.5 shows that, by using a classical random oracle A, we can efficiently prepare a uniformly random n-qubit state $|\psi_A\rangle$. But what if we want to use a random oracle to apply a uniformly random n-qubit *unitary* U_A? It is clear that we can do this if we have exponential time: given an oracle A, we simply query an exponentially long prefix A^* of A, and then treat A^* as an explicit description of a quantum circuit for U_A. But what if we can make only polynomially many quantum queries to A? We do not know whether that suffices for applying a random unitary; indeed, we do not even have a conjecture about this.

What we *can* show is that a single quantum query to the classical oracle A does not suffice for applying a random unitary. In particular, suppose every entry of an n-qubit unitary matrix U_A is a degree-1 polynomial in the bits of A (as it must be, if U_A is the result of a single quantum query). Then U_A can assume at most 4^{2^n} distinct values as we range over the possible A's, as opposed to the $\Omega\left(c^{2^{2n}}\right)$ that would be needed to approximate every n-qubit unitary. To prove this statement, we first need a lemma about matrices satisfying a certain algebraic relation.

Lemma 6.6 *Let $E_1, \ldots, E_M$ be nonzero $N \times N$ matrices over $\mathbb{C}$, and suppose that $E_i E_j^\dagger + E_j E_i^\dagger = 0$ for all $i \neq j$. Then $M \leq 2N$.*

Proof. Suppose by contradiction that $M > 2N$. Let $e_i^{(k)}$ be vector in $\mathbb{C}^N$ corresponding to the k^{th} row of E_i. Then the condition $E_i E_j^\dagger + E_j E_i^\dagger = 0$ implies that

$$e_i^{(k)} \cdot e_j^{(l)} + e_j^{(k)} \cdot e_i^{(l)} = 0$$

for all $i \neq j$ and k, l, where $\cdot$ denotes the complex inner product. Now for all i, let $k(i)$ be the minimum k such that $e_i^{(k)} \neq 0$, and consider the vectors $e_1^{(k(1))}, \ldots, e_M^{(k(M))} \in \mathbb{C}^N$. Certainly these vectors are not all orthogonal—indeed, since $M > 2N$, there must exist $i \neq j$ such that $\mathrm{Re}\left(e_i^{(k(i))} \cdot e_j^{(k(j))}\right) \neq 0$. There are now two cases: if $k(i) = k(j)$, then

$$e_i^{(k(i))} \cdot e_j^{(k(i))} + e_j^{(k(i))} \cdot e_i^{(k(i))} \neq 0$$

and we are done. On the other hand, if $k(i) \neq k(j)$, then

$$e_j^{(k(i))} \cdot e_i^{(k(j))} = -e_i^{(k(i))} \cdot e_j^{(k(j))}$$

is nonzero. Hence $e_j^{(k(i))}$ and $e_i^{(k(j))}$ must themselves be nonzero. But if $k(i) > k(j)$, then this contradicts the minimality of $k(i)$, while if $k(i) < k(j)$ then it contradicts the minimality of $k(j)$. ■

We can now prove the main result.

Theorem 6.7 *Let $U(X)$ be an $N \times N$ matrix, every entry of which is a degree-1 complex polynomial in variables $X = (x_1, \ldots, x_k)$. Suppose $U(X)$ is unitary for all $X \in \{0,1\}^k$. Then $U(X)$ can assume at most 4^N distinct values as we range over $X \in \{0,1\}^k$.*

Proof. By suitable rotation, we can assume without loss of generality that $U(0^k)$ is the $N \times N$ identity I. Let X_i be the k-bit string with a '1' only in the i^{th} position, and let $E_i := U(X_i) - I$. Then for all i,

$$\begin{aligned} E_i E_i^\dagger &= (U(X_i) - I)\left(U(X_i)^\dagger - I^\dagger\right) \\ &= I - U(X_i) - U(X_i)^\dagger + I \\ &= -E_i - E_i^\dagger. \end{aligned}$$

Next, for all $i \neq j$, let X_{ij} be the k-bit string with '1's only in the i^{th} and j^{th} positions. Since $U(X)$ is an affine function of X, we have

$$\begin{aligned} U(X_{ij}) &= U\left(0^k\right) + \left(U(X_i) - U\left(0^k\right)\right) + \left(U(X_j) - U\left(0^k\right)\right) \\ &= I + E_i + E_j. \end{aligned}$$

Therefore

$$\begin{aligned} 0 &= U(X_{ij})\, U(X_{ij})^\dagger - I \\ &= (I + E_i + E_j)\left(I^\dagger + E_i^\dagger + E_j^\dagger\right) - I \\ &= \left(E_i E_i^\dagger + E_j E_j^\dagger\right) + \left(E_i E_j^\dagger + E_j E_i^\dagger\right) + \left(E_i + E_i^\dagger\right) + \left(E_j + E_j^\dagger\right) \\ &= E_i E_j^\dagger + E_j E_i^\dagger. \end{aligned}$$

Here the first line uses unitarity, and the fourth line uses the fact that $E_i + E_i^\dagger = -E_i E_i^\dagger$ and $E_j + E_j^\dagger = -E_j E_j^\dagger$. Lemma 6.6 now implies that there can be at most $2N$ nonzero E_i's. Hence $U(X)$ can depend nontrivially on at most $2N$ bits of X, and can assume at most 2^{2N} values. ■

7 Open Problems

The most obvious problems left open by this paper are, first, to prove a classical oracle separation between QMA and QCMA, and second, to prove that the Group Non-Membership problem is in QCMA. We end by listing four other problems.

(1) The class QMA(2) is defined similarly to QMA, except that now there are two quantum provers who are guaranteed to share no entanglement. Is there a quantum oracle relative to which QMA$(2) \neq$ QMA?

(2) Is there a quantum oracle relative to which BQP/qpoly $\not\subset$ QMA/poly? This would show that the containment BQP/qpoly $\subseteq$ PP/poly proved in [2] is in some sense close to optimal.

(3) Can we use the ideas of Section 6 to give a classical oracle relative to which $\mathsf{BQP} \not\subset \mathsf{PH}$? What about a classical oracle relative to which $\mathsf{NP} \subseteq \mathsf{BQP}$ but $\mathsf{PH} \not\subset \mathsf{BQP}$?[12]

(4) Is there a polynomial-time quantum oracle algorithm Q, such that for every n-qubit unitary transformation U, there exists a classical oracle A such that Q^A approximately implements U? Alternatively, would any such algorithm require more than $\operatorname{poly}(n)$ queries to A?[13]

8 Acknowledgments

We thank the anonymous reviewers for their suggestions, and Dorit Aharonov, Laci Babai, Robert Beals, Robert Guralnick, Bill Kantor, and Cris Moore for helpful correspondence.

References

[1] S. Aaronson. Quantum copy-protection. In preparation.

[2] S. Aaronson. Limitations of quantum advice and one-way communication. *Theory of Computing*, 1:1–28, 2005. quant-ph/0402095. Conference version in Proceedings of CCC'2004.

[3] D. Aharonov and T. Naveh. Quantum NP - a survey. quant-ph/0210077, 2002.

[4] L. Babai. Trading group theory for randomness. In *Proc. ACM STOC*, pages 421–429, 1985.

[5] L. Babai. Local expansion of vertex-transitive graphs and random generation in finite groups. In *Proc. ACM STOC*, pages 164–174, 1991.

[6] L. Babai. Bounded round interactive proofs in finite groups. *SIAM J. Discrete Math*, 5(1):88–111, 1992.

[7] L. Babai and P. Erdős. Representation of group elements as short products. *Annals of Discrete Math.*, 12:27–30, 1982.

[8] L. Babai and E. Szemerédi. On the complexity of matrix group problems I. In *Proc. IEEE FOCS*, pages 229–240, 1984.

[9] M. Ben-Or, D. Coppersmith, M. Luby, and R. Rubinfeld. Non-abelian homomorphism testing, and distributions close to their self-convolutions. In *Proceedings of RANDOM*, pages 273–285. Springer-Verlag, 2004. ECCC TR04-052.

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

[11] K. Böröczky Jr. and G. Wintsche. Covering the sphere by equal spherical balls. In *Discrete and Computational Geometry: The Goodman-Pollack Festschrift*, pages 237–253. 2003.

[12] G. Brassard, P. Høyer, M. Mosca, and A. Tapp. Quantum amplitude amplification and estimation. In S. J. Lomonaco and H. E. Brandt, editors, *Quantum Computation and Information*, Contemporary Mathematics Series. AMS, 2002. quant-ph/0005055.

[13] J. H. Conway, R. T. Curtis, S. P. Norton, R. A. Parker, and R. A. Wilson. *Atlas of Finite Groups*. Clarendon Press, Oxford, 1985.

[12] Note that a simple relativizing argument shows that if $\mathsf{NP} \subseteq \mathsf{BPP}$ then $\mathsf{PH} \subseteq \mathsf{BPP}$.

[13] We do not even know whether a *single* query suffices. Note that Theorem 6.7 does not apply here, since we have dropped the requirement that Q^A must implement *some* n-qubit unitary (as opposed to a more general superoperator) for every oracle A.

[14] M. Ettinger, P. Høyer, and E. Knill. The quantum query complexity of the hidden subgroup problem is polynomial. *Inform. Proc. Lett.*, 91(1):43–48, 2004. quant-ph/0401083.

[15] L. K. Grover. A framework for fast quantum mechanical algorithms. In *Proc. ACM STOC*, pages 53–62, 1998. quant-ph/9711043.

[16] S. Hallgren, A. Russell, and A. Ta-Shma. The hidden subgroup problem and quantum computation using group representations. *SIAM J. Comput.*, 32(4):916–934, 2003. Conference version in STOC'2000, p. 627-635.

[17] B. Höfling. Efficient multiplication algorithms for finite polycyclic groups, 2007. Submitted. www-public.tu-bs.de/ bhoeflin/preprints/collect.pdf.

[18] J. Kempe, A. Kitaev, and O. Regev. The complexity of the Local Hamiltonian problem. *SIAM J. Comput.*, 35(5):1070–1097, 2006. quant-ph/0406180.

[19] A. Kitaev. Quantum measurements and the abelian stabilizer problem. ECCC TR96-003, quant-ph/9511026, 1996.

[20] A. Kitaev. Quantum computation: algorithms and error correction. *Russian Math. Surveys*, 52(6):1191–1249, 1997.

[21] C. Marriott and J. Watrous. Quantum Arthur-Merlin games. *Computational Complexity*, 14(2):122–152, 2005.

[22] C. Moore, D. N. Rockmore, and A. Russell. Generic quantum Fourier transforms. In *Proc. ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 778–787, 2004. quant-ph/0304064.

[23] M. Mosca and D. Stebila. Unforgeable quantum money. In preparation, 2006.

[24] P. M. Neumann. An enumeration theorem for finite groups. *Quart. J. Math. Ser.*, 2(20):395–401, 1969.

[25] L. Pyber. Enumerating finite groups of given order. *Annals of Mathematics*, 137:203–220, 1993.

[26] R. Raz and A. Shpilka. On the power of quantum proofs. In *Proc. IEEE Conference on Computational Complexity*, pages 260–274, 2004.

[27] A. Shamir. IP=PSPACE. *J. ACM*, 39(4):869–877, 1992.

[28] P. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.*, 26(5):1484–1509, 1997. Earlier version in IEEE FOCS 1994. quant-ph/9508027.

[29] C. Sims. Computational methods in the study of permutation groups. In *Computational Problems in Abstract Algebra*, pages 169–183. Pergamon Press, 1970.

[30] J. Watrous. Succinct quantum proofs for properties of finite groups. In *Proc. IEEE FOCS*, pages 537–546, 2000. cs.CC/0009002.