A Full Characterization of Quantum Advice

Scott Aaronson [*]
MIT

Andrew Drucker [†]
MIT

ABSTRACT

We prove the following surprising result: given any quantum state ρ on n qubits, there exists a local Hamiltonian H on $\operatorname{poly}(n)$ qubits (e.g., a sum of two-qubit interactions), such that any ground state of H can be used to simulate ρ on all quantum circuits of fixed polynomial size. In terms of complexity classes, this implies that $\mathsf{BQP/qpoly} \subseteq \mathsf{QMA/poly}$, which supersedes the previous result of Aaronson that $\mathsf{BQP/qpoly} \subseteq \mathsf{PP/poly}$. Indeed, we can exactly characterize quantum advice, as equivalent in power to *untrusted* quantum advice combined with trusted *classical* advice.

Proving our main result requires combining a large number of previous tools—including a result of Alon et al. on learning of real-valued concept classes, a result of Aaronson on the learnability of quantum states, and a result of Aharonov and Regev on 'QMA_+ super-verifiers'—and also creating some new ones. The main new tool is a so-called *majority-certificates lemma*, which is related to boosting in machine learning, and which seems likely to find independent applications. In its simplest version, this lemma says the following. Given any set S of Boolean functions on n variables, any function $f \in S$ can be expressed as the pointwise majority of $m = O(n)$ functions $f_1, \ldots, f_m \in S$, such that each f_i is the unique function in S compatible with $O(\log|S|)$ input/output constraints.

Categories and Subject Descriptors

F.1.3 [**Theory of Computation**]: Computation by Abstract Devices—*Complexity Measures and Classes*

General Terms

Theory

Keywords

quantum computation, learning, compression, advice, local Hamiltonians, nonuniform computation, boosting, Karp-Lipton Theorem

1. INTRODUCTION

How much classical information is needed to specify a quantum state on n qubits?

This question has inspired a rich and varied set of responses, in part because it can be interpreted in many ways. If we want to specify a quantum state ρ *exactly*, then of course the answer is 'an infinite amount,' since amplitudes in quantum mechanics are continuous. A natural compromise is to try to specify ρ *approximately*, i.e., to give a description which yields a state $\widetilde{\rho}$ whose statistical behavior is close to that of ρ under every measurement. (This statement is captured by the requirement that ρ and $\widetilde{\rho}$ are close under the so-called *trace distance* metric.) But it is not hard to see that even for this task, we still need to use an exponential (in n) number of classical bits.

This fact can be viewed as a disappointment, but also as an opportunity, since it raises the prospect that we might be able to encode massive amounts of information in physically compact quantum states: for example, we might hope to store 2^n classical bits in n qubits. But an obvious practical requirement is that we be able to retrieve the information reliably, and this rules out the hope of significant 'quantum compression' of classical strings, as shown by a landmark result of Holevo [21] from 1973. Consider a sender Alice and a recipient Bob, with a one-way quantum channel between them. Then Holevo's Theorem says that, if Alice wants to encode an n-bit classical string x into an m-qubit quantum state ρ_x, in such a way that Bob can retrieve x (with probability $2/3$, say) by measuring ρ_x, then Alice must take $m \geq n - O(1)$ (or $m \geq n/2 - O(1)$, if Alice and Bob share entanglement). In other words, for this communication task, quantum states offer essentially no advantage over classical strings. In 1999, Ambainis et al. [13] generalized Holevo's result as follows: even if Bob wants to learn only a *single bit* x_i of $x = x_1 \ldots x_n$ (for some $i \in [n]$ unknown to Alice), and is willing to destroy the state ρ_x in the process of learning that bit, Alice still needs to send $m = \Omega(n)$ qubits for Bob to succeed with high probability.

[*] Email: aaronson@csail.mit.edu. MIT EECS, Cambridge, MA, USA. Based upon work supported by the National Science Foundation under Grant No. 0844626. Also supported by a DARPA YFA grant, a Sloan Fellowship, and the Keck Foundation.

[†] Email: adrucker@mit.edu. MIT EECS, Cambridge, MA, USA. Supported by a DARPA YFA grant. The author was supported during part of this work by an Akamai Presidential Graduate Fellowship.

STOC'10, June 5–8, 2010, Cambridge, Massachusetts, USA
Copyright 2010 ACM 978-1-4503-0050-6/10/06 ...$10.00.

These results say that the exponential descriptive complexity of quantum states cannot be effectively harnessed for classical data storage, but they do not bound the number of practically meaningful 'degrees of freedom' in a quantum state used for purposes other than storing data. For example, a quantum state could be useful for computation, or it could be a physical system worthy of study in its own right. The question then becomes, what useful information *can* we give about an n-qubit state using a 'reasonable' number (say, $\text{poly}(n)$) of classical bits?

One approach to this question is to identify special subclasses of quantum states for which a faithful approximation can be specified using only $\text{poly}(n)$ bits. This has been done, for example, with matrix product states [30] and 'tree states' [1]. A second approach is to try to describe an *arbitrary* n-qubit state ρ concisely, in such a way that the state $\widetilde{\rho}$ recovered from the description is close to ρ with respect to some natural subclass of *measurements*. This has been done for specific classes like the 'pretty good measurements' of Hausladen and Wootters [20]. A more ambitious goal in this vein, explored by Aaronson in two previous works [2, 6] and continued in the present paper, is to give a description of an n-qubit state ρ which yields a state $\widetilde{\rho}$ that behaves approximately like ρ with respect to all (binary) measurements performable by quantum circuits of 'reasonable' size—say, of size at most n^c, for some fixed $c > 0$. Then if c is taken large enough, $\widetilde{\rho}$ is arguably 'just as good as' ρ for practical purposes.

Certainly we can achieve this goal using $2^{n^{c+O(1)}}$ bits: simply give approximations to the measurement statistics for every size-n^c circuit. However, the results of Holevo [21] and Ambainis et al. [13] suggest that a much more succinct description might be possible. This hope was realized by Aaronson [2], who gave a description scheme in which an n-qubit state can be specified using $\text{poly}(n)$ classical bits. There is a significant catch in Aaronson's result, though: the encoder Alice and decoder Bob both need to invest exponential amounts of computation.

In a subsequent paper [6], Aaronson gave a closely-related result which significantly reduces the computational requirements: now Alice can generate her message in polynomial time (for fixed c). Also, while Bob cannot necessarily construct the state $\widetilde{\rho}$ efficiently on his own, if he is presented with such a state (by an untrusted prover, say), Bob can *verify* the state in polynomial time. The catch in this result is a weakened approximation guarantee: Bob cannot use $\widetilde{\rho}$ to predict the outcomes of *all* the measurements defined by size-n^c circuits, but only *most* of them (with respect to a samplable distribution used by Alice in the encoding process). Aaronson [2, 6] conjectured that the tradeoff between this result and the previous one revealed an inherent limit to quantum compression.

1.1 Our Quantum Information Result

The main result of this paper is that Aaronson's conjecture was false: one really can get the best of both worlds, and simulate an arbitrary quantum state ρ on all small circuits, using a different state $\widetilde{\rho}$ that is easy to recognize. Indeed, we can even take $\widetilde{\rho}$ to be the *ground state of a local Hamiltonian*: that is, the unique pure state $\widetilde{\rho} = |\psi\rangle\langle\psi|$ on $\text{poly}(n)$ qubits that is compatible with $\text{poly}(n)$ local constraints, each involving a constant number of qubits. In a sense, then, this paper completes a 'trilogy' of which [2, 6] were the first two installments.

Here is a formal statement of our result.

THEOREM 1. *Let $c, \varepsilon > 0$, and let ρ be any n-qubit quantum state. Then there exists a 2-local Hamiltonian H on* $\text{poly}\left(n, \frac{1}{\varepsilon}\right)$ *qubits with unique ground state $|\psi\rangle\langle\psi|$, and a transformation $C \longrightarrow C'$ of quantum circuits, computable in time* $\text{poly}(n, 1/\varepsilon)$ *given H, such that the following holds: $|C'(|\psi\rangle\langle\psi|) - C(\rho)| \leq \varepsilon$ for any measurement C definable by a quantum circuit of size n^c. (Here $C(\rho)$ is the probability that C accepts ρ.)*

In other words, the ground states of local Hamiltonians are 'universal quantum states' in a very non-obvious sense. For example, suppose you own a quantum software store, which sells quantum states ρ that can be fed as input to quantum computers. Then our result says that *ground states of local Hamiltonians are the only kind of state you ever need to stock*. What makes this surprising is that being a good piece of quantum software might entail satisfying an exponential number of constraints: for example, if ρ is supposed to help a customer's quantum computer Q evaluate some Boolean function $f : \{0,1\}^n \rightarrow \{0,1\}$, then $Q(\rho, x)$ should output $f(x)$ for *every* input $x \in \{0,1\}^n$. By contrast, any k-local Hamiltonian H can be described as a set of at most $\binom{n}{k} = O(n^k)$ constraints.

One can also interpret Theorem 1 as a statement about communication over quantum channels. Suppose Alice (who is computationally unbounded) has a classical description of an n-qubit state ρ. She would like to describe ρ to Bob (who is computationally bounded), at least well enough for Bob to be able to *simulate* ρ on all quantum circuits of some fixed polynomial size. However, Alice cannot just send ρ to Bob, since her quantum communication channel is noisy and there is a chance that ρ might get corrupted along the way. Nor can she send a faithful classical description of ρ, since that would require an exponential number of bits. Our result provides an alternative: Alice can send a different quantum state σ, of $\text{poly}(n)$ qubits, together with a $\text{poly}(n)$-bit classical string x. Then, Bob can use x to *verify* that σ can be used to accurately simulate ρ on all small measurements.

We believe Theorem 1 makes a significant contribution to the study of the effective information content of quantum states. It does, however, leave open whether a quantum state of n qubits can be efficiently encoded *and* decoded in polynomial time, in a way that is 'good enough' to preserve the measurement statistics of measurements defined by circuits of fixed polynomial size. This remains an important problem for future work.

1.2 Impact on Quantum Complexity Theory

The questions addressed in this paper, and our results, are naturally phrased and proved in terms of complexity classes. In recent years, researchers have defined quantum complexity classes as a way to study the 'useful information' embodied in quantum states. One approach is to study the power of nonuniform *quantum advice*. The class $\mathsf{BQP/qpoly}$, defined by Nishimura and Yamakami [27], consists of all languages decidable in polynomial time by a quantum computer, with the help of a $\text{poly}(n)$-qubit advice state that depends only on the input length n. This class is analogous to the classical class $\mathsf{P/poly}$. To understand the role of quantum information in determining the power of $\mathsf{BQP/qpoly}$, a useful benchmark of comparison is the class $\mathsf{BQP/poly}$ of decision problems efficiently solvable by a quantum computer with $\text{poly}(n)$ bits of *classical* advice. It is open whether $\mathsf{BQP/qpoly} = \mathsf{BQP/poly}$.

A second approach studies the power of quantum *proof systems*, by analogy with the classical class NP. Kitaev (unpublished, 1999) defined the complexity class now called QMA, for 'Quantum Merlin-Arthur'. This is the class of decision problems for which a 'yes' answer can be proved by exhibiting a *quantum witness state* (or *quantum proof*) $|\psi\rangle$, on $\text{poly}(n)$ qubits, which is then checked by a skeptical polynomial-time quantum verifier. A natural benchmark class is QCMA (for 'Quantum Classical Merlin-Arthur'), defined by Aharonov and Naveh [9]. This is the class of decision problems for which a 'yes' answer can be checked by a *quantum* verifier who receives a *classical* witness. Here the natural open question is whether $\mathsf{QMA} = \mathsf{QCMA}$.

In this paper we prove a new upper bound on $\mathsf{BQP/qpoly}$:

THEOREM 2. $\mathsf{BQP/qpoly} \subseteq \mathsf{QMA/poly}$.

Previously Aaronson showed in [2] that $\mathsf{BQP/qpoly} \subseteq \mathsf{PP/poly}$, and showed in [6] that $\mathsf{BQP/qpoly}$ is contained in the 'heuristic' class $\mathsf{HeurQMA/poly}$; Theorem 2 supersedes both of these earlier results.

Theorem 2 says that one can always replace polynomial-size quantum advice by polynomial-size *classical* advice, together with a polynomial-size quantum *witness* (or equivalently, *untrusted* quantum advice). Indeed, we can *characterize* the class $\mathsf{BQP/qpoly}$, as equal to the subclass of $\mathsf{QMA/poly}$ in which the quantum witness state $|\psi_n\rangle$ can only depend on the input length n.[1]

Using Theorem 2, we also obtain several other results for quantum complexity theory:

(1) Without loss of generality, every quantum advice state can be taken to be the ground state of some local Hamiltonian H. (This essentially follows by combining our $\mathsf{BQP/qpoly} \subseteq \mathsf{QMA/poly}$ result with the result of Kitaev that LOCAL HAMILTONIANS is QMA-complete.)

(2) It is open whether for every local Hamiltonian H on n qubits, there exists a quantum circuit of size $\text{poly}(n)$ that prepares a ground state of H. It is easy to show that an affirmative answer would imply $\mathsf{QMA} = \mathsf{QCMA}$. As a consequence of Theorem 2, we can show that an affirmative answer would also imply $\mathsf{BQP/qpoly} = \mathsf{BQP/poly}$—thereby establishing a previously-unknown connection between quantum proofs and quantum advice.

(3) In the full version of this paper, we generalize Theorem 2 to show that $\mathsf{QCMA/qpoly} \subseteq \mathsf{QMA/poly}$.

(4) In the full version, we also use our new characterization of $\mathsf{BQP/qpoly}$ to prove a quantum analogue of the Karp-Lipton Theorem [24]. Recall that the Karp-Lipton Theorem says that if $\mathsf{NP} \subset \mathsf{P/poly}$, then the polynomial hierarchy collapses to the second level. Our 'Quantum Karp-Lipton Theorem' says that if $\mathsf{NP} \subset \mathsf{BQP/qpoly}$ (that is, NP-complete problems are efficiently solvable with the help of quantum advice), then $\Pi_2^{\mathsf{P}} \subseteq \mathsf{QMA}^{\mathsf{PromiseQMA}}$. As far as we know, this is the first nontrivial result to derive unlikely consequences from a hypothesis about quantum machines being able to solve NP-complete problems in polynomial time.

1.3 Proof Overview

We now give an overview of the proof Theorem 2, that $\mathsf{BQP/qpoly} \subseteq \mathsf{QMA/poly}$. As we will explain, our proof rests on a new idea we call the 'majority-certificates' technique, which is not specifically quantum and which seems likely to find other applications.

We begin with a language $L \in \mathsf{BQP/qpoly}$ and, for $n > 0$, a $\text{poly}(n)$-size quantum circuit $Q(x, \xi)$ that computes $L(x)$ with high probability when given the 'correct' advice state $\xi = \rho_n$ on $\text{poly}(n)$ qubits. The challenge, then, is to force Merlin to supply a witness state ρ' that behaves like ρ_n on every input $x \in \{0,1\}^n$.

Every potential advice state ξ defines a function $f_\xi : \{0,1\}^n \rightarrow [0,1]$, by $f_\xi(x) := \Pr[Q(x,\xi) = 1]$. For each such ξ, let $\widehat{f_\xi}(x) := [f_\xi(x) \geq 1/2]$ be the Boolean function obtained by rounding f_ξ. As a simplification, suppose that Merlin is restricted to sending an advice state ξ for which $f_\xi(x) \notin (1/3, 2/3)$: that is, an advice state which renders a 'clear opinion' about every input x. (This simplification helps to explain the main ideas, but does not follow the actual proof.) Let S be the set of all Boolean functions $f : \{0,1\}^n \rightarrow \{0,1\}$ that are expressible as $\widehat{f_\xi}$ for some such advice state ξ. Then S includes the 'target function' $f^* := L_n$ (the restriction of L to inputs of length n), as well as a potentially-large number of other functions. However, we claim S is not *too* large: $|S| \leq 2^{\text{poly}(n)}$. This bound on the 'effective information content' of quantum states was derived previously by Aaronson [2, 6], building on the work of Ambainis et al. [13].

One might initially hope that, just by virtue of the size bound on S, we could find some set of $\text{poly}(n)$ values

$$(x_1, f^*(x_1)), \ldots, (x_k, f^*(x_t))$$

which *isolate* f^* in S—that is, which differentiate f^* from all other members of S. In that case, the trusted classical advice could simply specify those values, as 'tests' for Arthur to perform on the quantum state sent by Merlin. Alas, this hope is unfounded in general. For consider the case where f^* is the identically-zero function, and S consists of f^* along with the 'point function' f_y (which equals 1 on y and 0 elsewhere), for all $y \in \{0,1\}^n$. Then f^* can only be isolated in S by specifying its value at *every* point!

Luckily, this counterexample leads us to a key observation. Although f is not isolatable in S by a small number of values, each point function f_y *can* be isolated (by its value at y), and moreover, f_y is quite 'close' to f. In fact, if we choose any three distinct strings x, y, z, then $f^* = \text{MAJ}(f_x, f_y, f_z)$. (Of course if f^* were the identically-zero function, it could be easily specified with classical advice! But f^* could have been any function in this example.)

This suggests a new, more indirect approach to our general problem: we try to express f as the pointwise majority vote

$$f^*(x) \equiv \text{MAJ}(f_1(x), \ldots, f_m(x)),$$

of a small number ($m = O(n)$, say) of *other* functions $f_1, \ldots, f_m$ in S, where each f_i is isolatable in S by specifying at most $k = O(\log|S|)$ of its values. Indeed, we will

[1] We call this restricted class $\mathsf{YQP/poly}$; in another notation it would be $\mathsf{OQMA/poly} \cap \mathsf{coOQMA/poly}$ (where the O stands for 'oblivious').

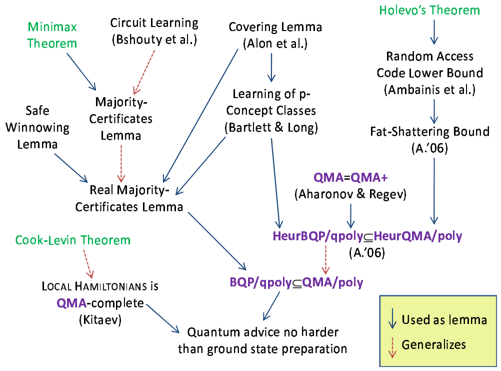
Minimax Theorem
Circuit Learning (Bshouty et al.)
Covering Lemma (Alon et al.)
Holevo's Theorem
Random Access Code Lower Bound (Ambainis et al.)
Safe Winnowing Lemma
Majority-Certificates Lemma
Learning of p-Concept Classes (Bartlett & Long)
Fat-Shattering Bound (A.'06)
QMA=QMA+ (Aharonov & Regev)
Real Majority-Certificates Lemma
HeurBQP/qpoly⊆HeurQMA/poly (A.'06)
Cook-Levin Theorem
BQP/qpoly⊆QMA/poly
LOCAL HAMILTONIANS is QMA-complete (Kitaev)
Quantum advice no harder than ground state preparation
Used as lemma
Generalizes

Figure 1: Dependency structure of our proof that quantum advice states can be expressed as ground states of local Hamiltonians.

show this can *always* be done. We call this key result the *majority-certificates lemma*; we will say more about its proof and its relation to earlier work in Section 1.4.

With this lemma in hand, we can solve our (artificially simplified) problem: in the $\mathsf{QMA/poly}$ protocol for L, we use certificates which isolate $f_1, \ldots, f_m \in S$ as above as the classical advice for Arthur. Arthur requests from Merlin each of the m states $\xi_1, \ldots, \xi_m$ such that $f_i = f_{\xi_i}$, and verifies that he receives appropriate states by checking them against the certificates. This involves multiple measurements of each ξ_i—and an immediate difficulty is that, since measurements are irreversible in quantum mechanics, the process of verifying the witness state might also destroy it. However, we get around this difficulty by appealing to a result of Aharonov and Regev [11]. This result essentially says that a QMA protocol in which Arthur is granted the (physically unrealistic) ability to perform 'non-destructive measurements' on his witness state, can be efficiently simulated by an ordinary QMA protocol.

To build intuition, we will begin (in Section 2) by proving the majority-certificates lemma for Boolean functions, as described above. However, to remove the artificial simplification we made and prove Theorem 2, we will need to generalize the lemma substantially, to a statement about possibly-infinite sets of real-valued functions $f : \{0,1\}^n \to [0,1]$. In the general version, the hypothesis that S is finite and not too large gets replaced by a more subtle assumption: namely, an upper bound on the so-called *fat-shattering dimension* of S. To prove our generalization, we use powerful results of Alon et al. [12] and Bartlett and Long [15] on the learnability of real-valued functions. We then use a bound on the fat-shattering dimension of real-valued functions defined by quantum states (from Aaronson [6], building on Ambainis et al. [13]). Figure 1 shows the overall dependency structure of the proof.

1.4 Majority-Certificates Lemma in Context

The majority-certificates lemma is closely related to the seminal notion of *boosting* [29] from computational learning theory. Boosting is a broad topic with a vast literature, but a common 'generic' form of the boosting problem is as follows: we want to learn some target function f^*, given sample data of the form $(x, f^*(x))$. We assume we have a *weak learning algorithm* $A^{f^*,\mathcal{D}}$, with the property that, for any probability distribution $\mathcal{D}$ over inputs x, with high probability A finds a hypothesis $f \in \mathcal{F}$ which predicts $f^*(x)$ 'reasonably well' when $x \sim \mathcal{D}$. The task is to 'boost' this weak learner into a *strong* learner B^{f^*}. The strong learner should output a collection of functions $f_1, \ldots, f_m \in \mathcal{F}$, such that a (possibly-weighted) majority vote over $f_1(x), \ldots, f_m(x)$ predicts $f^*(x)$ 'extremely well.' It turns out [29, 19] that this goal can be achieved in a very general setting.

Our majority-certificates lemma has strengths and weaknesses compared to boosting. Our assumptions are much milder than those of boosting: rather than needing a weak learner, we assume only that the hypothesis class S is 'not too large.' Also, we represent our target function f^* *exactly* by $\mathrm{MAJ}(f_1, \ldots, f_m)$, not just approximately. On the other hand, we do not give an efficient algorithm to *find* our majority-representation. Also, the f_i's are not 'explicitly given': we only give a way to *recognize* each f_i, under the assumption that the function purporting to be f_i is in fact drawn from the original hypothesis class.

The proof of our lemma also has similarities to boosting. As an analogue of a 'weak learner', we show that for every distribution $\mathcal{D}$, there exists a function $f \in S$ which agrees with the target function f^* on most $x \sim \mathcal{D}$, *and* which is isolatable in S by specifying $O(\log|S|)$ queries. Using the Minimax Theorem, we then nonconstructively 'boost' this fact into the desired majority-representation of f^*. We note that Nisan used the Minimax Theorem for boosting in a similar way, in his alternative proof of Impagliazzo's 'hard-core set theorem' (see [22]).

The majority-certificates lemma is also reminiscent of Bshouty et al.'s algorithm [16], for learning small circuits in the complexity class $\mathsf{ZPP}^{\mathsf{NP}}$. Our lemma lacks the algorithmic component of this earlier work, but unlike Bshouty et al., we do not require the functions being learned to come with any succinct labels (such as circuit descriptions).

1.5 Organization of the Paper

In Section 2, we prove the Boolean majority-certificates-lemma. In Section 3, we give our real-valued generalization of this lemma, and in Section 4 we use it to prove Theorem 2, and state some consequences for quantum complexity theory. Finally, Theorem 1 is proved in Section 4.3.

2. MAJORITY-CERTIFICATES LEMMA

A *Boolean concept class* is a family of sets $\{S_n\}_{n\geq 1}$, where each S_n consists of Boolean functions $f : \{0,1\}^n \to \{0,1\}$ on n variables. Abusing notation, we will often use S to refer directly to a set of Boolean functions on n variables, with the quantification over n being understood.

By a *certificate*, we mean a partial Boolean function $C : \{0,1\}^n \to \{0,1,*\}$. The *size* of C, denoted $|C|$, is the number of inputs x such that $C(x) \in \{0,1\}$. A Boolean function $f : \{0,1\}^n \to \{0,1\}$ is *consistent* with C if $f(x) = C(x)$ whenever $C(x) \in \{0,1\}$. Given a set S of Boolean functions and a certificate C, let $S[C]$ be the set of all functions $f \in S$ that are consistent with C. Say that a function $f \in S$ is *isolated in* S by the certificate C if $S[C] = \{f\}$.

We now prove a lemma that represents one of the main tools of this paper.

LEMMA 3 (MAJORITY-CERTIFICATES LEMMA). *Let S be a set of Boolean functions $f : \{0,1\}^n \to \{0,1\}$, and let $f^* \in S$. Then there exist $m = O(n)$ certificates $C_1, \ldots, C_m$,* *each of size $k = O(\log|S|)$, and functions $f_1, \ldots, f_m \in S$, such that*

(i) $S[C_i] = \{f_i\}$ *all* $i \in [m]$;

(ii) $\mathrm{MAJ}(f_1(x), \ldots, f_m(x)) = f^*(x)$ *for all* $x \in \{0,1\}^n$.

Proof. Our proof of Lemma 3 relies on the following claim.

Claim 4. *Let $\mathcal{D}$ be any distribution over inputs $x \in \{0,1\}^n$. Then there exists a function $f \in S$ such that*

(i) f is isolatable in S by a certificate C of size $k = O(\log|S|)$;

(ii) $\Pr_{x\sim\mathcal{D}}[f(x) \neq f^*(x)] \leq \frac{1}{10}$.

Lemma 3 follows from Claim 4 by a boosting-type argument, as follows. Consider a two-player game where:

- Alice chooses a certificate C of size k that isolates some $f \in S$, and
- Bob simultaneously chooses an input $x \in \{0,1\}^n$.

Alice wins the game if $f(x) = f^*(x)$. Claim 4 tells us that for every mixed strategy of Bob (i.e., distribution $\mathcal{D}$ over inputs), there exists a pure strategy of Alice that succeeds with probability at least 0.9 against $\mathcal{D}$. Then by the Minimax Theorem, there exists a mixed strategy for Alice—that is, a probability distribution $\mathcal{C}$ over certificates—that allows her to win with probability at least 0.9 against *every* pure strategy of Bob. Now suppose we draw $C_1, \ldots, C_m$ independently from $\mathcal{C}$, isolating functions $f_1, \ldots, f_m$ in S. Fix an input $x \in \{0,1\}^n$; then by the success of Alice's strategy against x, and applying a Chernoff bound,

$$\Pr_{f_1,\ldots,f_m\sim\pi}\left[\mathrm{MAJ}(f_1(x), \ldots, f_m(x)) \neq f^*(x)\right] < \frac{1}{2^n},$$

provided we choose $m = O(n)$ suitably. But by the union bound, this means there must be a *fixed* choice of $C_1, \ldots, C_m$ such that $\mathrm{MAJ}(f_1, \ldots, f_m) \equiv f^*(x)$, where each f_i is isolated in S by C_i. This proves Lemma 3, modulo the Claim. □

Proof of Claim 4. By symmetry, we can assume without loss of generality that f^* is the identically-zero function. Given the mixed strategy $\mathcal{D}$ of Bob, we construct the certificate C as follows. Initially C is empty: that is, $C(x) = *$ for all $x \in \{0,1\}^n$. In the first stage, we draw $t = O(\log|S|)$ inputs $x_1, \ldots, x_t$ independently from $\mathcal{D}$. For any $f : \{0,1\}^n \to \{0,1\}$, let

$$w_f := \Pr_{x\sim\mathcal{D}}[f(x) = 1].$$

Now suppose f is such that $w_f > 0.1$. Then

$$\Pr_{x_1,\ldots,x_t\sim\mathcal{D}}[f(x_1) = 0 \wedge \cdots \wedge f(x_t) = 0] < 0.9^t \leq \frac{1}{|S|},$$

provided $t \geq \log_{10/9}|S|$. So by the union bound, there must be a *fixed* choice of $x_1, \ldots, x_t$ that kills off every $f \in S$ such that $w_f > 0.1$—that is, such that $f(x_1) = \cdots = f(x_t) = 0$ implies $w_f \leq 0.1$. Fix that $x_1, \ldots, x_t$, and set $C(x_i) := 0$ for all $i \in [t]$. In the second stage, our goal is just to isolate some *particular* function $f \in S[C]$. We do this recursively as follows. If $|S[C]| = 1$ then we are done. Otherwise, there exists an input x such that $f(x) = 0$ for some $f \in S[C]$ and $f(x) = 1$ for other $f \in S[C]$. If setting $C(x) := 0$ decreases $|S[C]|$ by at least a factor of 2, then set $C(x) := 0$; otherwise set $C(x) := 1$. Since $S[C]$ can halve in size at most $\log_2|S|$ times, this procedure terminates after at most $\log_2|S|$ steps with $|S[C]| = 1$. The end result is a certificate C of size $O(\log|S|)$, which isolates a function f in S for which $w_f \leq 1/10$. We have therefore found a pure strategy for Alice that fails with probability at most $1/10$ against $\mathcal{D}$, as desired. □

3. EXTENSION TO REAL FUNCTIONS

In this section, we extend the majority-certificates lemma from Boolean functions to real-valued functions $f : \{0,1\}^n \to [0,1]$. We will need this extension for the application to quantum advice in Section 4. In proving our extension we will have to confront several new difficulties. Firstly, the concept classes S that we want to consider can now contain a *continuum* of functions—so Lemma 3, which assumed that S was finite and constructed certificates of size $O(\log|S|)$, is not going to work. In Section 3.1, we review notions from computational learning theory, including fat-shattering dimension and ε-covers, which (combined with results of Alon et al. [12] and Bartlett and Long [15]) can be used to get around this difficulty. Secondly, it is no longer enough to isolate a function $f_i \in S$ that we are interested in; instead we will need to 'safely' isolate f_i, which roughly speaking means that (i) f_i is consistent with some certificate C, and (ii) any $f \in S$ that is even *approximately* consistent with C is close to f_i. In Section 3.2, we prove a 'safe winnowing lemma' that can be used for this purpose. Finally, in Section 3.3, we put the pieces together to prove a real-valued majority-certificates lemma.

3.1 Background from Learning Theory

A *p-concept class* S is a family of functions $f : \{0,1\}^n \to [0,1]$ (as usual, quantification over all n is understood). Given functions $f, g : \{0,1\}^n \to [0,1]$ and a subset of inputs $X \subseteq \{0,1\}^n$, we will be interested in three measures of the distance between f and g restricted to X:

$$\Delta_\infty(f,g)[X] := \max_{x\in X}|f(x) - g(x)|,$$

$$\Delta_1(f,g)[X] := \sum_{x\in X}|f(x) - g(x)|,$$

$$\Delta_2(f,g)[X] := \sqrt{\sum_{x\in X}(f(x) - g(x))^2}.$$

For convenience, we define $\Delta_\infty(f,g) := \Delta_\infty(f,g)[\{0,1\}^n]$, and similarly for $\Delta_1(f,g)$ and $\Delta_2(f,g)$. Also, given a distribution $\mathcal{D}$ over $\{0,1\}^n$, define

$$\Delta_1(f,g)\langle\mathcal{D}\rangle := \mathop{\mathrm{E}}_{x\sim\mathcal{D}}[|f(x) - g(x)|].$$

Finally we will need the notions of coverability and fat-shattering dimension.

Definition 5 (Coverability). *Let S be a p-concept class. The subset $C \subseteq S$ is an ε-cover for S if for all $f \in S$, there exists a $g \in C$ such that $\Delta_\infty(f,g) \leq \varepsilon$.*

Definition 6 (Fat-Shattering Dimension). *Let S be a p-concept class and $\varepsilon > 0$ be a real number. We say the set $A \subseteq \{0,1\}^n$ is ε-shattered by S if there exists a function $r : A \to [0,1]$ such that for all $2^{|A|}$ Boolean functions*

$g : A \to \{0,1\}$, there exists a p-concept $f \in S$ such that for all $x \in A$, we have $f(x) \le r(x) - \varepsilon$ whenever $g(x) = 0$ and $f(x) \ge r(x) + \varepsilon$ whenever $g(x) = 1$. Then the ε-fat-shattering dimension of S, or $\operatorname{fat}_\varepsilon(S)$, is the size of the largest set ε-shattered by S.

The following central result was shown by Alon et al. [12] (see also [23]).

THEOREM 7 ([12]). *Every p-concept class S has an ε-cover of size $\exp\left[O\left((n + \log 1/\varepsilon)\operatorname{fat}_{\varepsilon/4}(S)\right)\right]$.*

Building on the work of Alon et al. [12], Bartlett and Long [15] then proved the following:

THEOREM 8 ([15]). *Let S be a p-concept class and $\mathcal{D}$ be a distribution over $\{0,1\}^n$. Fix an $f : \{0,1\}^n \to [0,1]$ (not necessarily in S) and an error parameter $\alpha > 0$. Suppose we form a set $X \subseteq \{0,1\}^n$ by choosing m inputs independently with replacement from $\mathcal{D}$. Then there exists a positive constant K such that, with probability at least $1-\delta$ over X, any hypothesis $h \in S$ that minimizes $\Delta_1(h,f)[X]$ also satisfies*

$$\Delta_1(h,f)\langle\mathcal{D}\rangle \le \alpha + \inf_{g\in S}\Delta_1(g,f)\langle\mathcal{D}\rangle,$$

provided $m \ge \frac{K}{\alpha^2}\left(\operatorname{fat}_{\alpha/5}(S)\log^2\frac{1}{\alpha} + \log\frac{1}{\delta}\right)$.

Theorem 8 has the following corollary, which is similar to Corollary 2.4 of Aaronson [6], but more directly suited to our purposes here.

COROLLARY 9. *Let S be a p-concept class and $\mathcal{D}$ be a distribution over $\{0,1\}^n$. Fix an $f \in S$ and an error parameter $\varepsilon > 0$. Suppose we form a set $X \subseteq \{0,1\}^n$ by choosing m inputs independently with replacement from $\mathcal{D}$. Then there exists a positive constant K such that, with probability at least $1-\delta$ over X, any hypothesis $h \in S$ that satisfies $\Delta_\infty(h,f)[X] \le \varepsilon$ also satisfies $\Delta_1(h,f)\langle\mathcal{D}\rangle \le 11\varepsilon$, provided $m \ge \frac{K}{\varepsilon^2}\left(\operatorname{fat}_\varepsilon(S)\log^2\frac{1}{\varepsilon} + \log\frac{1}{\delta}\right)$.*

PROOF. In the full version. □

3.2 Safe Winnowing Lemma

To prove the real-valued majority-certificates lemma, the first step is to prove a so-called 'safe winnowing lemma.' This lemma says intuitively that, given any set S of real-valued functions with a small ε-cover (or equivalently, with polynomially-bounded fat-shattering dimension), it is possible to find a set of $k = \operatorname{poly}(n)$ constraints $|f(x_1) - a_1| \le \epsilon$, $\ldots$, $|f(x_k) - a_k| \le \epsilon$ that are *essentially* compatible with one and only one function $f \in S$. Here 'essentially' means that (i) any function that satisfies the constraints is close to f in L_∞-norm, and (ii) f itself not only satisfies the constraints, but does so with a 'margin to spare.'

LEMMA 10 (SAFE WINNOWING LEMMA). *Let S be a set of functions $f : \{0,1\}^n \to [0,1]$. Fix a function $f^* \in S$ and subset $Y \subseteq \{0,1\}^n$. For some parameter $\varepsilon > 0$, let C be a finite ε-cover for S. Then there exists an $f \in S$, as well as a subset $Z \subseteq \{0,1\}^n$ of size at most $k = \log_2|C|$, such that:*

(i) Every $g \in S$ that satisfies $\Delta_\infty(f,g)[Y \cup Z] \le \frac{\varepsilon}{5k}$ also satisfies $\Delta_\infty(f,g) \le 3\varepsilon$.

(ii) $\Delta_\infty(f,f^*)[Y] \le \varepsilon/5$.

PROOF. Let $\delta := \frac{\varepsilon}{5k}$. We construct (f,Z) by an iterative procedure. Initially let $S_0 := S$, let $f_0 := f^*$, and let $Z_0 := Y$. We will form new sets $S_1, S_2, \ldots$ by repeatedly adding constraints of the form $f(x) \le \alpha$ or $f(x) \ge \alpha$ for various x, α, maintaining the invariant that $f_t \in S_t$. At iteration t, suppose there exists a function $g \in S_{t-1}$ such that $\Delta_\infty(f_{t-1},g)[Y \cup Z_{t-1}] \le \delta$, but nevertheless

$$|f_{t-1}(z_t) - g(z_t)| > 3\varepsilon$$

for some input z_t. Then first set $Z_t := Z_{t-1} \cup \{z_t\}$ (i.e., add z_t into our set of inputs, if it is not already there). Let $v := \frac{1}{2}[f_{t-1}(z_t) + g(z_t)]$, let A be the set of all functions $h \in S_{t-1}$ such that $h(z_t) < v$, and let B be the set of all $h \in S_{t-1}$ such that $h(z_t) \ge v$. Also, for any given set M, let $M^\Diamond := M \cap C$. Then clearly $\min\{|A^\Diamond|, |B^\Diamond|\} \le |S_{t-1}^\Diamond|/2$. If $|A^\Diamond| < |B^\Diamond|$, then set $S_t := A$; otherwise set $S_t := B$. Then set $f_t := f_{t-1}$ if $f_{t-1} \in S_t$ and $f_t := g$ otherwise. Since $|S_t^\Diamond|$ can halve at most $k = \log_2|C|$ times, it is clear that after $T \le k$ iterations we have $|S_T^\Diamond| \le 1$. Set $f := f_T$ and $Z := Z_T$. Then by the triangle inequality,

$$\Delta_\infty(f,f^*)[Y] \le T\delta \le \frac{\varepsilon}{5},$$

and also

$$|f(z_t) - f_t(z_t)| \le (T-t)\delta < \frac{\varepsilon}{5}$$

for all $t \in [T]$. So suppose by contradiction that there still exists a function $g \in S_T$ such that $\Delta_\infty(f,g)[Y \cup Z] \le \delta$ but $|f(x) - g(x)| > 3\varepsilon$ for some x, and consider functions $p, q \in C$ in the cover such that $\Delta_\infty(f,p) \le \varepsilon$ and $\Delta_\infty(g,q) \le \varepsilon$. Then $p, q \in S_T^\Diamond$ but $p \ne q$, which contradicts the fact that $|S_T^\Diamond| \le 1$. Also notice that for all $g \in S$, if $\Delta_\infty(f,g)[Y \cup Z] \le \delta$ then $g \in S_T$. Thus $\Delta_\infty(f,g)[Y \cup Z] \le \delta$ implies $\Delta_\infty(f,g) \le 3\varepsilon$ as desired. □

Note that Lemma 10 is still interesting in the special case $Y = \varnothing$, in which case f^* is irrelevant, and the problem reduces to finding a Z such that every $g \in S$ that satisfies $\Delta_\infty(f,g)[Z] \le \frac{\varepsilon}{5k}$ also satisfies $\Delta_\infty(f,g) \le 3\varepsilon$.

3.3 Real-Valued Majority-Certificates Lemma

We are finally ready to generalize Lemma 3 to the case of real-valued functions.

LEMMA 11 (REAL MAJORITY-CERTIFICATES). *Let S be a p-concept class, let $f^* \in S$, and let $\varepsilon > 0$. Then for some $m = O(n/\varepsilon^2)$, there exist functions $f_1, \ldots, f_m \in S$, sets $X_1, \ldots, X_m \subseteq \{0,1\}^n$ each of size*

$$k = O\left(\left(n + \frac{\log^2\frac{1}{\varepsilon}}{\varepsilon^2}\right)\operatorname{fat}_{\frac{\varepsilon}{48}}(S)\right),$$

and an $\alpha = \Omega\left(\frac{\varepsilon}{(n+\log 1/\varepsilon)\operatorname{fat}_{\varepsilon/48}(S)}\right)$ for which the following holds. All $g_1, \ldots, g_m \in S$ that satisfy $\Delta_\infty(f_i,g_i)[X_i] \le \alpha$ for $i \in [m]$ also satisfy $\Delta_\infty(f^*,g) \le \varepsilon$, where $g(x) := \sum_{i=1}^m g_i(x)$.

PROOF. Let $\beta := \varepsilon/48$, let $t := C\left(n + \log\frac{1}{\beta}\right)\operatorname{fat}_\beta(S)$ where C is a suitably large constant, and let $\alpha := 0.4\beta/t$. Also, let S_{fin} be a finite α-cover for S: that is, a finite subset $S_{\text{fin}} \subseteq S$ such that for all $f \in S$, there exists a $g \in S_{\text{fin}}$ such that $\Delta_\infty(f,g) \le \alpha$.[2] Given f and X, let $S[f,X]$

[2] We will need S_{fin} for the technical reason that the basic Minimax Theorem only works with finite strategy spaces.

be the set of all $g \in S$ such that $\Delta_\infty (f,g)[X] \leq \alpha$. Now consider a two-player game where Alice chooses a function $f \in S_{\text{fin}}$ and a set $X \subseteq \{0,1\}^n$ of size k, and Bob simultaneously chooses an input $x \in \{0,1\}^n$. Alice's *penalty* in this game (the number she is trying to minimize) equals $\sup_{g\in S[f,X]} |f^*(x) - g(x)|$. We claim that there exists a mixed strategy for Alice—that is, a probability distribution $\mathcal{P}$ over (f,X) pairs—that gives her an expected penalty of at most $\varepsilon/2$ against every pure strategy of Bob. Let us see why the lemma follows from this claim. Fix an input $x \in \{0,1\}^n$, and suppose Alice draws $(f_1,X_1),\ldots,(f_m,X_m)$ independently from $\mathcal{P}$. Then for all $i \in [m]$,

$$\operatorname*{E}_{(f_i,X_i)\sim\mathcal{P}}\left[\sup_{g\in S[f,X]} |f^*(x) - g(x)|\right] \leq \frac{\varepsilon}{2}.$$

Thus, letting $z_1,\ldots,z_m$ be independent random variables in $[0,1]$, each with expectation at most $\varepsilon/2$, the expression

$$\Pr_{(f_i,X_i)\sim\mathcal{P}\forall i}\left[\begin{array}{c}\exists g_1 \in S[f_1,X_1],\ldots,g_m \in S[f_m,X_m] \quad : \\ \left|f^*(x) - \frac{g_1(x)+\cdots+g_m(x)}{m}\right| > \varepsilon\end{array}\right]$$

is at most $\Pr[z_1 + \cdots z_m > \varepsilon m]$ by the triangle inequality. This, in turn, is less than $2\exp\left(-2(\varepsilon m)^2/m\right) < 2^{-n}$ by Hoeffding's inequality, provided we choose $m = O\left(n/\varepsilon^2\right)$ suitably. By the union bound, this means that there must be a fixed choice of $f_1,\ldots,f_m$ and $X_1,\ldots,X_m$ such that

$$\left|f^*(x) - \frac{g_1(x)+\cdots+g_m(x)}{m}\right| \leq \varepsilon$$

for all $g_1 \in S[f_1,X_1],\ldots,g_m \in S[f_m,X_m]$ and all inputs $x \in \{0,1\}^n$ simultaneously, as desired. We now prove the claim. By the Minimax Theorem, our task is equivalent to the following: given any mixed strategy $\mathcal{D}$ of Bob, find a *pure* strategy of Alice that achieves a penalty of at most $\varepsilon/2$ against $\mathcal{D}$. In other words, given any distribution $\mathcal{D}$ over inputs $x \in \{0,1\}^n$, we want a fixed function $f \in S_{\text{fin}}$, and a set $X \subseteq \{0,1\}^n$ of size k, such that

$$\operatorname*{E}_{x\sim\mathcal{D}}\left[\sup_{g\in S[f,X]} |f^*(x) - g(x)|\right] \leq \frac{\varepsilon}{2}.$$

We construct this (f,X) pair as follows. In the first stage, we let Y be a set, of size at most

$$M := \frac{K}{\beta^2}\left(\operatorname{fat}_\beta(S)\log^2\frac{1}{\beta} + \log\frac{1}{\delta}\right),$$

formed by choosing M inputs independently with replacement from $\mathcal{D}$. Here $\beta = \varepsilon/48$ as defined earlier, $\delta = 1/2$, and K is the constant from Corollary 9. Then by Corollary 9, with probability at least $1-\delta = 1/2$ over the choice of Y, any $g \in S$ that satisfies $\Delta_\infty(f^*,g)[Y] \leq \beta$ also satisfies $\Delta_1(f^*,g)\langle\mathcal{D}\rangle \leq 11\beta$. So there must be a *fixed* choice of Y with that property. Fix that Y, and let S' be the set of all $g \in S$ such that $\Delta_\infty(f^*,g)[Y] \leq \beta$. In the second stage, our goal is just to winnow S' down to a particular function f. More precisely, we want to find an $f \in S' \cap S_{\text{fin}}$, and a set $X \subseteq \{0,1\}^n$ containing Y, such that any $g \in S$ that satisfies $\Delta_\infty(f,g)[X] \leq \alpha$ also satisfies $\Delta_\infty(f,g) \leq 11\beta$. We find this (f,X) pair as follows. By Theorem 7, the class S' has a 4β-cover of size

$$N \leq \exp\left[O\left(\left(n + \log\frac{1}{\beta}\right)\operatorname{fat}_\beta(S)\right)\right].$$

Let $t := \log_2 N$. Then by Lemma 10, there exists a function $u \in S'$, as well as a subset $Z \subseteq \{0,1\}^n$ of size at most t, such that:

(i) $\Delta_\infty(u,f^*)[Y] \leq 0.8\beta$.

(ii) Every $g \in S'$ that satisfies $\Delta_\infty(u,g)[Y\cup Z] \leq \frac{0.8\beta}{t}$ also satisfies $\Delta_\infty(u,g) \leq 12\beta$.

Let $X := Y \cup Z$, and observe that

$$\begin{aligned}|X| &= O\left(\frac{1}{\beta^2}\operatorname{fat}_\beta(S)\log^2\frac{1}{\beta} + \left(n + \log\frac{1}{\beta}\right)\operatorname{fat}_\beta(S)\right)\\ &= O\left(\left(n + \frac{\log^2 1/\varepsilon}{\varepsilon^2}\right)\operatorname{fat}_{\varepsilon/48}(S)\right)\end{aligned}$$

as desired. Now let f be a function in S_{fin} such that $\Delta_\infty(f,u) \leq \alpha$. Let us check that f has the properties we want. First,

$$\begin{aligned}\Delta_\infty(f^*,f)[Y] &\leq \Delta_\infty(f^*,u)[Y] + \Delta_\infty(u,f)[Y]\\ &\leq 0.8\beta + \alpha < 0.9\beta,\end{aligned}$$

hence $f \in S'$ as desired. Next, any $g \in S$ that satisfies $\Delta_\infty(f,g)[X] \leq \alpha$ also satisfies

$$\Delta_\infty(f^*,g)[Y] \leq \Delta_\infty(f^*,f)[Y] + \Delta_\infty(f,g)[Y] \leq 0.9\beta + \alpha < \beta,$$

hence $g \in S'$, hence $\Delta_1(f^*,g)\langle\mathcal{D}\rangle \leq 11\beta$. So any $g \in S$ that satisfies $\Delta_\infty(f,g)[X] \leq \alpha$ satisfies

$$\Delta_\infty(u,g)[Z] \leq \Delta_\infty(u,f)[Z] + \Delta_\infty(f,g)[Z] \leq 2\alpha = \frac{0.8\beta}{t},$$

hence $\Delta_\infty(u,g) \leq 12\beta$ (since such a g must belong to S'), hence

$$\Delta_\infty(f,g) \leq \Delta_\infty(f,u) + \Delta_\infty(u,g) \leq \alpha + 12\beta \leq 13\beta.$$

To conclude,

$$\begin{aligned}&\operatorname*{E}_{x\sim\mathcal{D}}\left[\sup_{g\in S[f,X]} |f^*(x) - g(x)|\right]\\ &\leq \Delta_1(f^*,f)\langle\mathcal{D}\rangle + \sup_{g\in S[f,X]}\Delta_\infty(f,g)\\ &\leq 11\beta + 13\beta = \frac{\varepsilon}{2}\end{aligned}$$

as desired. This proves the claim and hence the lemma. □

4. APPLICATION TO QUANTUM ADVICE

In this section, we use the real-valued majority-certificates lemma to prove Theorem 2.

4.1 Bestiary of Quantum Complexity Classes

Given a language $L \subseteq \{0,1\}^*$, let $L : \{0,1\}^* \to \{0,1\}$ be the characteristic function of L. We now give a formal definition of the class $\mathsf{BQP/qpoly}$.

Definition 12. *A language L is in $\mathsf{BQP/qpoly}$ if there exists a poly-time quantum algorithm A and polynomial p such that for all n, there exists an advice state ρ_n on $p(n)$ qubits such that $A(x,\rho_n)$ outputs $L(x)$ with probability $\geq 2/3$ for all $x \in \{0,1\}^n$.*

Closely related to quantum advice are *quantum proofs*. We now recall the definition of QMA (Quantum Merlin-Arthur), a quantum version of NP.

Definition 13. *A language L is in* QMA *if there exists a polynomial-time quantum algorithm A and polynomial p such that for all $x \in \{0,1\}^n$:*

(i) If $x \in L$ then there exists a witness ρ_x on $p(n)$ qubits such that $A(x, \rho_x)$ accepts with probability $\geq 2/3$.

(ii) If $x \notin L$ then $A(x, \rho)$ accepts with probability $\leq 1/3$ for all ρ.

We will actually need a generalization of QMA, which was called QMA_+ by Aharonov and Regev [10].[3]

Definition 14. *A language L is in* QMA_+ *if there exists a polynomial-time algorithm A, which takes $x \in \{0,1\}^n$ as input and produces quantum circuits $C_{x,1}, \ldots, C_{x,m}$ and rational numbers $r_{x,1}, \ldots, r_{x,m}$ as output, as well as polynomials p, q such that for all $x \in \{0,1\}^n$:*

(i) If $x \in L$ then there exists a witness ρ_x on $p(n)$ qubits such that $|\Pr[C_{x,i}(\rho_x) \text{ accepts}] - r_{x,i}| \leq 1/q(n)$ for all $i \in [m]$.

(ii) If $x \notin L$ then for all ρ, there exists an $i \in [m]$ such that $|\Pr[C_{x,i}(\rho) \text{ accepts}] - r_{x,i}| > 5/q(n)$.

Aharonov and Regev [10] made the following extremely useful observation, which we prove in the full version for completeness.

Theorem 15 ([11]). $\mathsf{QMA}_+ = \mathsf{QMA}$.

To state our results, it will be helpful to have the further notion of *untrusted advice*, which is like advice in that it depends only on the input length n, but like a witness in that it cannot be trusted. This notion has been studied before: Chakaravarthy and Roy [17] and Fortnow and Santhanam [18] defined the complexity class ONP ('Oblivious NP'), which is like NP except that the witness can depend only on the input length. Independently, Aaronson [6] defined the complexity class YP,[4] which is easily seen to equal $\mathsf{ONP} \cap \mathsf{coONP}$. We will adopt the 'Y' notation in this paper, because it is much easier to write $\mathsf{YQP/poly}$ (for example) than $\mathsf{OQMA/poly} \cap \mathsf{coOQMA/poly}$.

We now give a formal definition of YP, as well as a slight variant called YP^*.

Definition 16. *A language L is in* YP *if there exist polytime algorithms A, B and a polynomial p such that:*

(i) For all n, there exists an advice string $y_n \in \{0,1\}^{p(n)}$ such that $A(x, y_n) = 1$ for all $x \in \{0,1\}^n$.

(ii) If $A(x, y) = 1$, then $B(x, y) = L(x)$.

L is in YP^* *if A ignores x, depending only on y.*

Clearly $\mathsf{P} \subseteq \mathsf{YP}^* \subseteq \mathsf{YP} \subseteq \mathsf{P/poly} \cap \mathsf{NP} \cap \mathsf{coNP}$. Also, Aaronson [6] showed that $\mathsf{ZPP} \subseteq \mathsf{YP}$. We will be interested in the natural quantum analogues of YP and YP^*:

Definition 17. *A language L is in* YQP *if there exist polynomial-time quantum algorithms A, B and a polynomial p such that:*

(i) For all n, there exists an advice state ρ_n on $p(n)$ qubits such that $A(x, \rho_n)$ accepts with probability $\geq 2/3$ for all $x \in \{0,1\}^n$.

(ii) If $A(x, \rho)$ accepts with probability $\geq 1/3$, then $B(x, \rho)$ outputs $L(x)$ with probability $\geq 2/3$.

L is in YQP^* *if A ignores x, depending only on ρ.*

Clearly $\mathsf{BQP} \subseteq \mathsf{YQP}^* \subseteq \mathsf{YQP} \subseteq \mathsf{BQP/qpoly} \cap \mathsf{QMA} \cap \mathsf{coQMA}$. By direct analogy to QMA_+, we can define the following generalizations of YQP and YQP^*:

Definition 18. *A language L is in* YQP_+ *if there exists a polynomial-time algorithm A, which takes $x \in \{0,1\}^n$ as input and produces quantum circuits $C_{x,1}, \ldots, C_{x,m}$ and rational numbers $r_{x,1}, \ldots, r_{x,m}$ as output; a polynomial-time quantum algorithm B; and polynomials p, q such that:*

(i) For all n, there exists an advice state ρ_n on $p(n)$ qubits such that $|\Pr[C_{x,i}(\rho_n) \text{ accepts}] - r_{x,i}| \leq 1/q(n)$ for all $i \in [m]$ and $x \in \{0,1\}^n$.

(ii) If $|\Pr[C_{x,i}(\rho) \text{ accepts}] - r_{x,i}| \leq 5/q(n)$ for all $i \in [m]$, then $B(x, \rho)$ outputs $L(x)$ with probability $\geq 2/3$.

L is in YQP_+^* *if moreover A ignores x.*

Then we have the following counterpart to Theorem 15:

Theorem 19. $\mathsf{YQP}_+ = \mathsf{YQP}$ *and* $\mathsf{YQP}_+^* = \mathsf{YQP}^*$.

Proof. In the full version. □

4.2 Characterizing Quantum Advice

Fix a polynomial-size quantum circuit Q. For a given advice state ρ, let $f_\rho(x) := \Pr[Q \text{ accepts } x, \rho]$. Let S be the p-concept class consisting of f_ρ for all $p(n)$-qubit mixed states ρ. Then Aaronson [6] proved the following.

Theorem 20 ([6]). $\operatorname{fat}_\gamma(S) = O\left(p(n)/\gamma^2\right)$.

We now prove the following characterization of $\mathsf{BQP/qpoly}$, which immediately implies (and strengthens) Theorem 2:

Theorem 21. $\mathsf{BQP/qpoly} = \mathsf{YQP/poly}$.

Proof. One direction ($\mathsf{YQP/poly} \subseteq \mathsf{BQP/qpoly}$) is obvious, since untrusted quantum advice and trusted classical advice can both be simulated by trusted quantum advice. We prove that $\mathsf{BQP/qpoly} \subseteq \mathsf{YQP/poly}$. It suffices to show that $\mathsf{BQP/qpoly} \subseteq \mathsf{YQP}_+\mathsf{/poly}$, since $\mathsf{YQP} = \mathsf{YQP}_+$ by Theorem 19. Let $L \in \mathsf{BQP/qpoly}$, let Q be a quantum algorithm that decides L with completeness and soundness errors $1/5$, and let $x \in \{0,1\}^n$ be the input. Also, let $f_\xi(z) := \Pr[Q(z, \xi) \text{ accepts}]$, where ξ is a $p(n)$-qubit quantum advice state for Q. Then by definition, there exists a 'true' advice state ρ_n such that

$$|f_{\rho_n}(z) - L(z)| \leq 0.2$$

for all $z \in \{0,1\}^n$. Let S be the p-concept class consisting of f_ξ for all $p(n)$-qubit mixed states ξ. Then Theorem 20 implies that $\operatorname{fat}_\gamma(S) = O\left(p(n)/\gamma^2\right)$ for all $\gamma > 0$. Set $\gamma := 1/480$. Then by Lemma 11, for some $m = O(n)$, there exist $p(n)$-qubit mixed states $\rho[1], \ldots, \rho[m]$, sets $X_1, \ldots, X_m \subseteq \{0,1\}^n$ each of size $k = O(n \cdot p(n))$, and an $\alpha = \Omega\left(\frac{1}{n \cdot p(n)}\right)$ for which the following holds:

[3] Aharonov and Regev actually defined QMA_+ in a slightly more general way. However, the definition below is all we need; note that all these classes turn out to equal QMA anyway.

[4] YP stands for 'Yoda Polynomial-Time,' a nomenclature that seems to make neither more nor less sense than 'Arthur-Merlin.'

(*) *All $p(n)$-qubit states $\sigma[1], \ldots, \sigma[m]$ that for $i \in [m]$ satisfy $\Delta_\infty\left(f_{\rho[i]}, f_{\sigma[i]}\right)[X_i] \leq 5\alpha$, satisfy $\Delta_\infty\left(f_{\rho_n}, f_\sigma\right) \leq 0.1$ as well, where $\sigma := \frac{1}{m}\left(\sigma[1] + \cdots + \sigma[m]\right)$.*

Our $\mathsf{YQP}_+/\mathsf{poly}$ simulation is now the following. The classical /poly advice encodes the sets $X_1, \ldots, X_m$, as well as a rational approximation $r_{i,z}$ to $f_{\rho[i]}(z)$ for each $i \in [m]$ and $z \in X_i$. The untrusted quantum advice ρ_n' consists of m registers of $p(n)$ qubits each; in the honest case, ρ_n' is simply $\rho[1] \otimes \cdots \otimes \rho[m]$. Let $\sigma[i]$ be the i^{th} register of ρ_n'. Then given the advice, the YQP_+ machine A outputs a circuit $C_{i,z}$ that runs $Q(z, \sigma[i])$ and outputs the result, for each $i \in [m]$ and $z \in X_i$. The machine B chooses $i \in [m]$ uniformly at random, then runs $Q(x, \sigma[i])$ and outputs the result. We are interested in the difference between $\Pr\left[C_{i,z}(\rho_n') \text{ accepts}\right]$ and $r_{i,z}$. In the honest case,

$$\Pr\left[C_{i,z}\left(\rho_n'\right) \text{ accepts}\right] = \Pr\left[Q\left(z, \rho[i]\right) \text{ accepts}\right] = f_{\rho[i]}(z)$$

for all i, z. Moreover, we can easily arrange each $r_{i,z}$ to be within α of $f_{\rho[i]}(z)$, by using $O(\log n)$ bits to specify each $r_{i,z}$. For the soundness case, suppose

$$\left|\Pr\left[C_{i,z}\left(\rho_n'\right) \text{ accepts}\right] - r_{i,z}\right| \leq 5\alpha$$

for all $i \in [m]$ and $z \in X_i$. Then by (*), we have $\Delta_\infty\left(f_{\rho_n}, f_\sigma\right) \leq 0.1$. Notice that by linearity of expectation,

$$\Pr[B \text{ accepts}] = \operatorname*{E}_{i \in [m]}\left[\Pr\left[Q\left(x, \sigma[i]\right) \text{ accepts}\right]\right] = f_\sigma(x),$$

and that this holds regardless of what entanglement might be present among the m registers $\sigma[1], \ldots, \sigma[m]$. Hence

$$\begin{aligned} &\left|\Pr[B \text{ accepts}] - L(x)\right| \\ &\leq \left|\Pr[B \text{ accepts}] - f_{\rho_n}(x)\right| + \left|f_{\rho_n}(x) - L(x)\right| \\ &\leq 0.1 + 0.2 \end{aligned}$$

which is less than 1/3, so $L \in \mathsf{YQP}_+/\mathsf{poly} = \mathsf{YQP}/\mathsf{poly}$. □

Theorem 21 actually yields the stronger result $\mathsf{BQP}/\mathsf{qpoly} \subseteq \mathsf{YQP}^*/\mathsf{poly}$, since the machine A had no dependence on the input x. We therefore have $\mathsf{BQP}/\mathsf{qpoly} = \mathsf{YQP}^*/\mathsf{poly} = \mathsf{YQP}/\mathsf{poly}$: the two definitions of YQP collapse in the presence of polynomial-size classical advice. Since we never needed the assumption that the $\mathsf{BQP}/\mathsf{qpoly}$ machine computes a *language* (i.e., a total Boolean function), another strengthening we can easily observe is $\mathsf{PromiseBQP}/\mathsf{qpoly} = \mathsf{PromiseYQP}/\mathsf{poly}$.

4.3 The Complexity of Preparing Quantum Advice States

If we combine Theorem 21 with known QMA-completeness results, we can obtain a striking consequence for quantum complexity theory. Namely, *the preparation of quantum advice states can always be reduced to the preparation of ground states of local Hamiltonians*—despite the fact that quantum advice states involve an exponential number of constraints, while ground states of local Hamiltonians involve only a polynomial number. In particular, if ground states of local Hamiltonians can be prepared by polynomial-size circuits, then we have not only $\mathsf{QMA} = \mathsf{QCMA}$, but also $\mathsf{BQP}/\mathsf{qpoly} = \mathsf{BQP}/\mathsf{poly}$. The following theorem makes this connection precise.

THEOREM 22. *Let Q be a polynomial-size quantum circuit that takes an advice state ρ_n. Then there exists another polynomial-size quantum circuit Q' with the following property. For all n and $\varepsilon > 0$, there exists a 2-local Hamiltonian H on $\operatorname{poly}(n, 1/\varepsilon)$ qubits, such that for all ground states $|\phi\rangle$ of H and inputs $x \in \{0,1\}^n$,*

$$\left|\Pr\left[Q' \text{ accepts } x, |\phi\rangle\right] - \Pr\left[Q \text{ accepts } x, \rho_n\right]\right| \leq \varepsilon.$$

Furthermore, Q' can be efficiently generated given Q together with a description of H.

PROOF. Kempe, Kitaev, and Regev [25] proved that the 2-LOCAL HAMILTONIANS problem is QMA-complete. Furthermore, examining their proof, we find that it yields the following stronger result. Let V be a QMA verification procedure with completeness and soundness errors δ. Then there exists a 2-local Hamiltonian H, as well as a polynomial-time 'recovery procedure' R, such that if $|\phi\rangle$ is any ground state of H, then with $\Omega(1/\operatorname{poly}(n))$ probability, $R(|\phi\rangle)$ outputs a state $|\varphi\rangle$ such that $\Pr[V \text{ accepts } |\varphi\rangle] \geq 1 - \delta$. To prove the stronger result: consider a ground state of H, which Kempe et al. show to be a *history state* of the form

$$|\phi\rangle = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} |t\rangle |\phi_t\rangle.$$

Then R can simply measure the clock register $|t\rangle$, postselect on obtaining the outcome $t = 1$, and then retrieve $|\varphi\rangle$ from the computation register $|\phi_1\rangle$. Now let Q be a polynomial-size quantum circuit that takes advice state ρ_n, and let (A, B) be the $\mathsf{YQP}/\mathsf{poly}$ checking algorithm (with error parameter δ) from Theorem 21. Then by the above, there exists a 2-local Hamiltonian H on $\operatorname{poly}(n, 1/\delta)$ qubits, as well as a polynomial-time algorithm R, such that

(i) If $|\phi\rangle$ is any ground state of H, then with at least $1 - \delta$ probability, $R(|\phi\rangle)$ outputs a state $|\varphi\rangle$ such that $\Pr[A \text{ accepts } |\varphi\rangle] \geq 1 - \delta$.

(ii) This $|\varphi\rangle$ satisfies $\left|B_\varphi(x) - Q_{\rho_n}(x)\right| \leq \delta$ for all $x \in \{0,1\}^n$, where $B_\varphi(x) := \Pr[B \text{ accepts } x, |\varphi\rangle]$ and

$$Q_{\rho_n}(x) := \Pr\left[Q \text{ accepts } x, \rho_n\right].$$

We can now combine R and B into a single algorithm Q', such that $\left|Q_\phi'(x) - Q_{\rho_n}(x)\right| \leq 2\delta$ for all $x \in \{0,1\}^n$. Setting $\delta := \varepsilon/2$ then yields the corollary. □

Let us make two remarks about Theorem 22. First, as a 'free byproduct,' we get that

$$\left|\Pr\left[Q' \text{ accepts } x, |\phi\rangle\right] - \Pr\left[Q \text{ accepts } x, \rho_n\right]\right| \leq 2\varepsilon$$

for all $|\phi\rangle$ that are *ε-close* in trace distance to a ground state of H. Second, there is nothing special here about 2-LOCAL HAMILTONIANS. So far as we know, *all* existing QMA-completeness reductions have the property we needed for Theorem 22: namely, the property that any ground state of the new instance can be transformed into a QMA witness for the original instance, with $\Omega(1/\operatorname{poly}(n))$ success probability. As one example, Aharonov et al. [8] showed that even finding the ground state energy of a nearest-neighbor Hamiltonian on the line is QMA-complete, provided the line is composed of qudits with $d \geq 12$. We can combine their result with Theorem 21 to show that for all $L \in \mathsf{BQP}/\mathsf{qpoly}$, there exists a nearest-neighbor qudit Hamiltonian H on the line, such that any ground state of H is a valid quantum advice state for L.

Proof of Theorem 1. Fix $c, \varepsilon > 0$, and let ρ be the n-qubit state in Theorem 1. Let $Q(C, \xi)$ be an (efficiently constructible) polynomial-size quantum circuit that takes a description of a quantum measurement circuit C of size n^c, as well as a quantum state ξ of n qubits, and that outputs the measurement result $C(\xi)$.

Fix $\rho_n := \rho$. Let H be the 2-local Hamiltonian given by Theorem 22, with ground state $|\psi\rangle$, and let $Q'(C, \xi)$ be the circuit in that Theorem, efficiently derivable given Q and H. Then, if we define the measurement C' as $C'(\xi) := Q'(C, \xi)$, we have

$$\left|C'\left(|\psi\rangle\langle\psi|\right) - C(\rho)\right| = \left|Q'\left(C, |\psi\rangle\langle\psi|\right) - Q(C, \rho)\right| \leq \varepsilon.$$

□

5. OPEN PROBLEMS

One open problem is simply to find more applications of the majority-certificates lemma, which seems likely to have uses outside of quantum complexity theory. Also, can we improve the parameters of the majority-certificates lemma (the size of the certificates or the number $O(n)$ of certificates), or alternatively, show that the current parameters are essentially optimal? Can we prove the real-valued majority-certificates lemma with an error tolerance α that depends only on the desired accuracy ε of the final approximation, not on n or the fat-shattering dimension of S?

On the quantum complexity side, we mention several questions. First, in Theorem 22, is the polynomial blowup in the number of qubits unavoidable? Second, can we use the ideas in this paper to prove any upper bound on the class $\mathsf{QMA/qpoly}$ better than the $\mathsf{PSPACE/poly}$ upper bound shown by Aaronson [5]? Third, if $\mathsf{NP} \subset \mathsf{BQP/qpoly}$, then does $\mathsf{QMA}^{\mathsf{PromiseQMA}}$ contain not just Π_2^{P} but the entire polynomial hierarchy? Finally, is $\mathsf{BQP/qpoly} = \mathsf{BQP/poly}$?

6. ACKNOWLEDGMENTS

We thank Sanjeev Arora, Kai-Min Chung, Avinatan Hassidim, John Watrous, and Colin Zheng for helpful discussions, and the anonymous reviewers for their comments.

7. REFERENCES

[1] S. Aaronson. Multilinear formulas and skepticism of quantum computing. *STOC*, p. 118–127, 2004.

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

[3] S. Aaronson. Quantum computing, postselection, and probabilistic polynomial-time. *Proc. Roy. Soc. London*, A461(2063):3473–3482, 2005.

[4] S. Aaronson. Oracles are subtle but not malicious. *Complexity (CCC)*, p. 340–354, 2006.

[5] S. Aaronson. QMA/qpoly is contained in PSPACE/poly: de-Merlinizing quantum protocols. *Complexity (CCC)*, p. 261–273, 2006.

[6] S. Aaronson. The learnability of quantum states. *Proc. Roy. Soc. London*, 463(2088):3089–3114, 2007.

[7] S. Aaronson and G. Kuperberg. Quantum versus classical proofs and advice. *Theory of Computing*, 3(7):129–157, 2007.

[8] D. Aharonov, D. Gottesman, S. Irani, and J. Kempe. The power of quantum systems on a line. *Comm. Math. Phys.*, 287(1):41–65, 2009.

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

[10] D. Aharonov and O. Regev. A lattice problem in Quantum NP. *FOCS*, p. 210–219, 2003.

[11] D. Aharonov and O. Regev. Lattice problems in NP intersect coNP. *J. ACM*, 52(5):749–765, 2005.

[12] N. Alon, S. Ben-David, N. Cesa-Bianchi, and D. Haussler. Scale-sensitive dimensions, uniform convergence, and learnability. *J. ACM*, 44(4):615–631, 1997.

[13] A. Ambainis, A. Nayak, A. Ta-Shma, and U. V. Vazirani. Quantum dense coding and quantum finite automata. *J. ACM*, 49:496–511, 2002.

[14] L. Babai. Local expansion of vertex-transitive graphs and random generation in finite groups. *STOC*, p. 164–174, 1991.

[15] P. L. Bartlett and P. M. Long. Prediction, learning, uniform convergence, and scale-sensitive dimensions. *J. Comput. Sys. Sci.*, 56(2):174–190, 1998.

[16] N. H. Bshouty, R. Cleve, R. Gavaldà, S. Kannan, and C. Tamon. Oracles and queries that are sufficient for exact learning. *J. Comput. Sys. Sci.*, 52(3):421–433, 1996.

[17] V. Chakaravarthy and S. Roy. Oblivious symmetric alternation. *STACS*, p. 230–241, 2006.

[18] L. Fortnow and R. Santhanam. Fixed-polynomial size circuit bounds, 2006. ECCC TR06-157.

[19] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Sys. Sci.*, 55(1):119–139, 1997.

[20] P. Hausladen and W. K. Wootters. A 'pretty good' measurement for distinguishing quantum states. *J. Modern Optics*, 41(12):2385–2390, 1994.

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

[22] R. Impagliazzo. Hard-core distributions for somewhat hard problems. *FOCS*, p. 538–545, 1995.

[23] S. Kakade and A. Tewari. Learning theory lecture notes, 2008. ttic.uchicago.edu/˜tewari/LT_SP2008.html.

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

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

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

[27] H. Nishimura and T. Yamakami. Polynomial time quantum computation with advice. *Inform. Proc. Lett.*, 90:195–204, 2003.

[28] R. Raz. Quantum information and the PCP theorem. *FOCS*, p. 459–468, 2005.

[29] R. E. Schapire. The strength of weak learnability. *Machine Learning*, 5(2):197–227, 1990.

[30] G. Vidal. Efficient classical simulation of slightly entangled quantum computations. *Phys. Rev. Lett.*, 91, 2003.

[31] J. Watrous. Succinct quantum proofs for properties of finite groups. *FOCS*, p. 537–546, 2000.