The complexity of simulating local measurements on quantum systems

Sevag Gharibian[1,3] and Justin Yirka[2,3]

[1]Department of Computer Science, Paderborn University, Germany
[2]Department of Computer Science, The University of Texas at Austin, USA
[3]Department of Computer Science, Virginia Commonwealth University, USA

An important task in quantum physics is the estimation of local quantities for ground states of local Hamiltonians. Recently, [Ambainis, CCC 2014] defined the complexity class $P^{QMA[\log]}$, and motivated its study by showing that the physical task of estimating the expectation value of a local observable against the ground state of a local Hamiltonian is $P^{QMA[\log]}$-complete. In this paper, we continue the study of $P^{QMA[\log]}$, obtaining the following lower and upper bounds.

Lower bounds (hardness results):

- The $P^{QMA[\log]}$-completeness result of [Ambainis, CCC 2014] requires $O(\log n)$-local observables and Hamiltonians. We show that simulating even a *single qubit* measurement on ground states of 5-local Hamiltonians is $P^{QMA[\log]}$-complete, resolving an open question of Ambainis.

- We formalize the complexity theoretic study of estimating two-point correlation functions against ground states, and show that this task is similarly $P^{QMA[\log]}$-complete.

- We identify a flaw in [Ambainis, CCC 2014] regarding a $P^{UQMA[\log]}$-hardness proof for estimating spectral gaps of local Hamiltonians. By introducing a "query validation" technique, we build on [Ambainis, CCC 2014] to obtain $P^{UQMA[\log]}$-hardness for estimating spectral gaps under polynomial-time Turing reductions.

Upper bounds (containment in complexity classes):

- $P^{QMA[\log]}$ is thought of as "slightly harder" than QMA. We justify this formally by exploiting the hierarchical voting technique of [Beigel, Hemachandra, Wechsung, SCT 1989] to show $P^{QMA[\log]} \subseteq PP$. This improves the containment QMA $\subseteq$ PP [Kitaev, Watrous, STOC 2000].

This work contributes a rigorous treatment of the subtlety involved in studying oracle classes in which the oracle solves a *promise* problem. This is particularly relevant for quantum complexity theory, where most natural classes such as BQP and QMA are defined as promise classes.

Sevag Gharibian: sgharibian@upb.de, http://groups.uni-paderborn.de/fg-qi/people.html

Justin Yirka: yirka@utexas.edu, https://www.justinyirka.com/

Contents

1 Introduction

1.1 Background on Quantum Hamiltonian Complexity

The use of computational complexity theory to study the inherent difficulty of computational problems has proven remarkably fruitful over the last decades. For example, the theory of NP-completeness [21, 42, 46] has helped classify the worst-case complexity of hundreds of computational problems which elude efficient classical algorithms. In the quantum setting, the study of a quantum analogue of NP, known as Quantum Merlin Arthur[1] (QMA), was started in 1999 by the seminal "quantum Cook-Levin theorem" of Kitaev [45], which showed that estimating the ground state energy of a given k-local Hamiltonian is QMA-complete for $k \geq 5$. Here, a k-local Hamiltonian H can be thought of as a quantum constraint satisfaction system in which each quantum clause acts non-trivially on k qubits. More formally, $H \in \mathbb{C}^{2^n \times 2^n}$ is an exponentially large Hermitian matrix acting on n qubits, but with a succinct description[2] $H = \sum_i H_i$, where each local clause $H_i \in \mathbb{C}^{2^k \times 2^k}$ acts non-trivially on k qubits. The "largest total weight of satisfiable clauses" is given by the *ground state energy* of H, i.e. the smallest eigenvalue of H. Physically, the ground state energy and its corresponding eigenvector, the *ground state*, are motivated in that they represent the energy level and state of a given quantum system at low temperature, respectively. For this reason, since Kitaev's work [45], a number of physically motivated problems have been shown complete for QMA (this has given rise to the field of *Quantum Hamiltonian Complexity*, see, e.g., [52], [10]

---

[1]More precisely, QMA is Merlin-Arthur (MA) with a quantum proof and quantum verifier.

[2]Implicitly, if H_i acts on a subset $S_i \subseteq [n]$ of qubits non-trivially, then more accurately one writes $H_i \otimes I_{[n] \setminus S_i}$. We write $H = \sum_i H_i$ for simplicity.

and [31] for surveys[3]). Many of these QMA-complete problems focus on estimating ground state energies of local Hamiltonians.

Beyond ground state energies. In recent years, however, new directions in quantum complexity theory involving other physical properties of local Hamiltonians have appeared. We attempt to survey a number of such results here.

Brown, Flammia and Schuch [15] (also Shi and Zhang [59]) introduced a quantum analogue of #P, denoted #BQP, and showed that computing the ground state degeneracy or density of states of local Hamiltonians is #BQP-complete. Gharibian and Sikora [30] showed that determining whether the ground space of a local Hamiltonian has an "energy barrier" is QCMA-complete, where QCMA [2] is Merlin-Arthur (MA) with a classical proof and quantum prover. This was strengthened by Gosset, Mehta, and Vidick [38], who showed that QCMA-completeness holds even for *commuting* local Hamiltonians. Bravyi and Gosset [12] studied the complexity of quantum impurity problems, which involve a bath of free fermions coupled to an interacting subsystem dubbed an "impurity". Cubitt, Montanaro, and Piddock [24] have shown that certain simple spin-lattice models are "universal", in the sense that they can replicate the entire physics of any other quantum many-body system.

From a hardness of approximation perspective, Gharibian and Kempe [29] introduced cq-Σ_2, a quantum generalization of Σ_2^p, and showed that determining the smallest subset of interaction terms of a given local Hamiltonian which yields a frustrated ground space is cq-Σ_2-complete (and additionally, cq-Σ_2-hard to approximate). Aharonov and Zhou [3] studied the task of "Hamiltonian sparsification" or "degree-reduction", in which one attempts to simulate an input local Hamiltonian with a new local Hamiltonian with an interaction graph of bounded degree while preserving only the ground space and spectral gap (in general, [3] show this is impossible). This was in pursuit of answering whether classical proof techniques for the classical PCP theorem carry over to the quantum setting.

Finally, various authors have studied spectral gaps of local Hamiltonians from a *computability theory* perspective. (In computability theory, one asks whether a decision or promise problem can be decided by a Turing machine running in a finite number of steps. In the current paper, our focus is instead on complexity theory, in which problems are typically known to be computable; the question is rather to obtain an estimate of the resources the Turing machine requires to solve the problem.) Gosset and Mozgunov [37] and Bravyi and Gosset [11] have studied spectral gaps for frustration-free 1D translation-invariant systems (the latter, in particular, shows that distinguishing between gapped and gapless phases is decidable in the spin-1/2 chains studied). Cubitt, Perez-Garcia and Wolf [23] have shown undecidability of estimating spectral gaps in the thermodynamic limit for translation-invariant, nearest-neighbor Hamiltonians on a 2D square lattice. This has very recently been improved to undecidability for 1D translation invariant systems by Bausch, Cubitt, Lucia, and

---

[3]Since these surveys were published, the study of ground spaces through the field of Quantum Hamiltonian Complexity has continued to grow. For example, recent variants on circuit-to-Hamiltonian mappings such as the space-time construction of Breuckmann and Terhal [14], of Bausch and Crosson [6] for considering complex weights and branching transitions, of Bausch, Cubitt and Ozols [7] for embedding QMA_{exp} computations into 1D translation invariant systems of local dimension 42, or of Caha, Landau, and Nagaj [17] for achieving an arbitrarily high success probability of extracting a computation result from a history state while only needing to increase the clock size logarithmically, have been given. Bravyi and Hastings [13] have shown that the local Hamiltonian problem for the quantum Ising model is StoqMA-complete, completing the complexity classification scheme of Cubitt and Montanaro [22]. Gilyén and Sattath [35] have given a constructive Quantum Lovasz Local Lemma which efficiently prepares a frustration-free Hamiltonian's ground state under the assumption that the system is "uniformly" gapped. (This list of results is a small sample of recent work.)

Perez-Garcia [8]. Note that both the current paper and [5] also study spectral gaps, but from a complexity theory perspective; in particular, in contrast to the undecidability studies listed above, which consider the thermodynamic limit (i.e. the number of qubits n goes to infinity), in our setting the number of qubits n (expressed in unary) is part of the input to the problem, as is standard in complexity theory.

1.2 Simulating local measurements on low-temperature quantum systems

In Section 1.1, we listed a number of results studying properties of local Hamiltonians beyond ground state energies. We intentionally omitted one particular problem, however, which is extremely well motivated physically and which is the starting point of this work. Suppose one cools a many-body quantum system to low temperature in the lab and *a priori* does not know the system's properties (such as its ground state energy, ground space, etc); this is a natural assumption, as many of these properties are QMA-complete to estimate to begin with. The most "basic" action an experimenter can now take is to perform a local measurement (typically on a constant number of qubits) in an attempt to extract information about the system. The question is: How difficult is it to computationally simulate this "basic" action?

This computational task was formalized by Ambainis [5] (formal definitions in Section 2) and designated *Approximate Simulation (APX-SIM): Given a k-local Hamiltonian H and an l-local observable A, estimate the expectation value of the measurement A against the ground state of H, i.e. estimate $\langle A \rangle := \langle \psi | A | \psi \rangle$ for $| \psi \rangle$ a ground state of H.* It turned out that not only is this problem "hard", but that it is in fact harder then even QMA.

To formalize this, Ambainis introduced [5] the complexity class $P^{QMA[\log]}$, which intuitively is "slightly harder" than QMA, and showed that APX-SIM is $P^{QMA[\log]}$-complete when the Hamiltonian H and observable A are both $O(\log n)$-local. To help set context before stating our results, we now discuss $P^{QMA[\log]}$ and its classical analogue $P^{NP[\log]}$ in further depth.

1.3 Oracle complexity classes, $P^{QMA[\log]}$, and $P^{NP[\log]}$

Formally, $P^{QMA[\log]}$ is the class of decision problems which can be decided by a polynomial-time deterministic Turing machine making $O(\log n)$ queries to an oracle for QMA. Thus, it is an example of an *oracle complexity class*.

Let us make three remarks. First, by "oracle for QMA", one typically means an oracle for a QMA-complete problem Π (since any other problem in QMA may be reduced in polynomial-time to Π); in this paper, we shall set Π as the QMA-complete 2-local Hamiltonian problem (2-LH) [43], an instance (H, a, b) of which asks: *Is the ground state energy of H at most a (YES case), or at least b (NO case), for $b - a \geq 1/\text{poly}(n)$?* Second, although $P^{QMA[\log]}$ uses a QMA oracle, it in fact also contains the complementary class co-QMA. This is because to solve any co-QMA problem, a $P^{QMA[\log]}$ machine can plug the co-QMA problem instance into the QMA oracle, and subsequently flip the oracle's answer in polynomial-time to solve the co-QMA problem. Thus, $P^{QMA[\log]} \neq QMA$ unless co-QMA $\subseteq$ QMA (which is unlikely), and so $P^{QMA[\log]}$ is likely strictly harder than QMA. The third remark is that $P^{QMA[\log]}$ uses an oracle for a class of *promise* problems; this is a subtle but crucial point, which we discuss shortly; first, let us set the stage by reviewing the analogous classical class $P^{NP[\log]}$.

The class $P^{NP[\log]}$. $P^{NP[\log]}$ is defined analogously to $P^{QMA[\log]}$ except it utilizes an NP oracle. Analogous to $P^{QMA[\log]}$, $P^{NP[\log]}$ contains both NP and co-NP and thus is likely strictly harder than

NP. As an upper bound, $P^{NP[\log]} \subseteq NP^{NP} = \Sigma_2^p$, where an NP^{NP} machine is an NP machine which nondeterministically makes up to a polynomial number of calls to an NP oracle, and Σ_2^p is the second level of the Polynomial-Time Hierarchy (PH). In contrast, it is unlikely for $P^{QMA[\log]}$ to be in PH, as even $BQP \subseteq QMA \subseteq P^{QMA[\log]}$ is generally not believed to be in PH [1, 20, 26, 55, 56]. A natural question is whether $P^{QMA[\log]}$ might instead be contained in an appropriate quantum analogue of PH. The answer is not clear, since unlike the fact that $\Sigma_2^p = NP^{NP}$, it is not clear that "quantum Σ_2^p" should equal (say) QMA^{QMA}; see [32] for a discussion and treatment of quantum analogues of PH. Additional references on "quantum PH" are [29, 48, 63].

As far as we are aware, whether $P^{NP[k]}$ (i.e. $k \in O(1)$ NP queries), $P^{NP[\log^k n]}$ (i.e. $O(\log^k n)$ NP queries for $k \geq 1$), and P^{NP} (i.e. polynomially many NP queries) coincide remains open. Notably, if $P^{NP[1]} = P^{NP[2]}$, then $P^{NP[1]} = P^{NP[\log]}$ and PH collapses [39]. But, it *is* known is that if the queries to the NP oracle are *non-adaptive*, meaning they are all made in parallel, then the resulting class $P^{||NP}$ equals $P^{NP[\log]}$ [16, 40]. (The analogous statement $P^{||QMA} = P^{QMA[\log]}$ has recently been shown [33].) Similarly, a P machine making $O(\log^k n)$ adaptive NP queries is equivalent in power to a P machine making $O(\log^{k+1} n)$ non-adaptive queries for all $k \geq 1$ [18]. In terms of complete problems, determining the election winner in Lewis Carroll's 1876 voting system is $P^{NP[\log]}$-complete [41], and model checking for certain branching-time temporal logics is $P^{NP[\log^2 n]}$-complete [57].

Finally, it is important to note that NP is a class of *decision* problems. Formally, this means any NP problem is specified by a language $L \subseteq \{0,1\}^*$ with the corresponding decision problem: Given input $x \in \{0,1\}^*$, is $x \in L$? In the context of $P^{NP[\log]}$, this means *any* string $x \in \{0,1\}^*$ is a valid problem instance or *query* to an NP oracle for language L, since any such x is either in the language or not. Unfortunately, an analogous statement cannot be made about $P^{QMA[\log]}$, complicating its study; this brings us to our third, crucial remark about $P^{QMA[\log]}$ from earlier.

Oracles for promise classes and the issue of invalid queries. In contrast to NP, QMA is a class of *promise* problems. Formally, this means the space of all inputs $\{0,1\}^*$ is partitioned into *three* sets, A, B, C, where A and C are YES and NO instances, respectively, and B is the set of "invalid" instances. (Classes of decision problems are a special case of this in which $A = L$, $B = \emptyset$, and $C = \{0,1\}^* \setminus L$.) Why might this pose a problem for studying $P^{QMA[\log]}$?

Recall that we assume all calls by the $P^{QMA[\log]}$ machine to the QMA oracle Q are for instances (H, a, b) of 2-LH: *Is the ground state energy of H at most a (YES case), or at least b (NO case), for $b - a \geq 1/\text{poly}(n)$?* Unfortunately, a P machine cannot in general tell[4] whether the instance (H, a, b) it feeds to Q satisfies the promise conditions of LH; in particular, the ground state energy may lie in the interval (a, b). We call any such instance or query (H, a, b) violating the 2-LH promise "invalid" (these instances belong in the set B of instances above). For any invalid query, the oracle Q is allowed to accept or reject arbitrarily. This raises a potential issue: if the oracle is allowed to respond arbitrarily to invalid queries, how does one ensure a YES instance (or NO instance) of a $P^{QMA[\log]}$ problem is well-defined (normally, the P machine conditions its future actions on

---

[4]By definition, the P machine can only execute polynomial-time computations. Thus, even if we assume without loss of generality that the P machine uses specific circuit-to-Hamiltonian constructions (for preparing queries to the QMA oracle) whose promise gaps are precisely known (e.g. [6, 17]), it is not clear how the machine should know whether the quantum circuit it feeds into said construction satisfies a promise gap to begin with — where such circuit-to-Hamiltonian constructions require the input circuit to satisfy an inverse polynomial promise gap, which is unlikely to be relaxed to more easily verified, say, inverse exponentially small promise gaps, since it would imply $PSPACE \subseteq QMA$, which follows since QMA with exponentially small gap equals PSPACE [25].

the response the oracle returns for each query)? To do so, we stipulate (see, e.g., Definition 3 of Goldreich [36]) that the P machine must output the *same* answer regardless of how any invalid queries are answered by the oracle. We view the issue of formally handling invalid queries as one of the central themes and contributions of this work.

1.4 Our results

Our results fall into two categories: *Lower bounds* (hardness results) and *upper bounds* (containment in complexity classes).

Lower bounds (hardness results). We begin by showing three hardness results; two focus on computational problems introduced in [5] (APX-SIM and SPECTRAL-GAP) and one introduces a new problem (APX-2-CORR).

1. $P^{\text{QMA[log]}}$-completeness of APX-SIM *for $O(1)$-local Hamiltonians and single-qubit observables.* Recall that in [5], Ambainis introduced $P^{\text{QMA[log]}}$ and showed that APX-SIM (i.e. given k-local Hamiltonian H and l-local observable A, simulate measurement A on the ground state of H) is $P^{\text{QMA[log]}}$-complete. This proof required both the Hamiltonian H and observable A to be $O(\log n)$-local. From a physical standpoint, however, it is typically desirable to have $O(1)$-local Hamiltonians and observables — whether $P^{\text{QMA[log]}}$-hardness holds in this regime was left as an open question in [5]. We thus first ask: *Is* APX-SIM *still hard for $O(1)$-local Hamiltonians and 1-local observables?*

Let us develop some intuition before answering this question. Typically, computational problems (such as estimating ground state energies) of 1-local *Hamiltonians* are easy, since the qubits do not interact. This intuition does not, however, carry over to the setting of simulating 1-local *measurements*. For example, by embedding a 3-SAT instance ϕ into a 3-local Hamiltonian and then using the ability to repeatedly measure observable Z against single qubits of the ground state, we can extract a solution to ϕ! Thus, the 3-local Hamiltonian and 1-local observable case is at least NP-hard. Indeed, here we show it is much harder, resolving Ambainis's open question.

Theorem 1.1. *Given a 5-local Hamiltonian H on n qubits and a 1-local observable A, estimating $\langle A \rangle$ (i.e.* APX-SIM*) is $P^{\text{QMA[log]}}$-complete.*

Thus, measuring just a *single* qubit of the ground state of a local Hamiltonian H with a single-qubit observable A (in fact, A is fixed, independent of H in our construction) is harder than QMA (assuming $\text{QMA} \neq P^{\text{QMA[log]}}$, which is likely as otherwise co-QMA $\subseteq$ QMA).

2. $P^{\text{QMA[log]}}$-completeness of APX-2-CORR *for $O(1)$-local Hamiltonians and single-qubit observables.* We next introduce a second natural problem related to APX-SIM, denoted APX-2-CORR. APX-2-CORR is defined similarly to APX-SIM except one is given Hamiltonian H and observables A and B and asked to estimate the *two-point correlation function* $\langle A \otimes B \rangle - \langle A \rangle \langle B \rangle$ (recall $\langle A \rangle := \langle \psi | A | \psi \rangle$ for $|\psi\rangle$ a ground state of H). By modifying the construction behind the proof of Theorem 1.1, we also show APX-2-CORR is $P^{\text{QMA[log]}}$-complete.

Theorem 1.2. *Given a 5-local Hamiltonian H on n qubits and a pair of 1-local observables A and B, estimating $\langle A \otimes B \rangle - \langle A \rangle \langle B \rangle$ (i.e.* APX-2-CORR*) is $P^{\text{QMA[log]}}$-complete.*

3. $P^{\text{UQMA[log]}}$-hardness of estimating spectral gaps. A third well-motivated problem involving Hamiltonians is SPECTRAL-GAP [5]: *Given a k-local Hamiltonian H, estimate $\lambda_1(H) - \lambda_2(H)$, where*

$\lambda_i(H)$ *denotes the i-th smallest eigenvalue of H.* (For clarity, if the ground space of H is degenerate, we define its spectral gap as 0.) In [5], it is shown that SPECTRAL-GAP $\in \mathrm{P}^{\mathrm{QMA[log]}}$, and a claimed proof is given that SPECTRAL-GAP for $O(\log n)$-local Hamiltonians H is $\mathrm{P}^{\mathrm{UQMA[log]}}$-hard. Here, $\mathrm{P}^{\mathrm{UQMA[log]}}$ is $\mathrm{P}^{\mathrm{QMA[log]}}$ except with a Unique QMA oracle, and Unique QMA is roughly QMA with a unique accepting quantum witness in the YES case (see Section 4.3 for formal definitions).

Recall now that, as discussed in Section 1.3, a central theme of this work is the subtlety involved in the study of oracle classes in which the oracle solves a *promise* problem (such as $\mathrm{P}^{\mathrm{QMA[log]}}$), as opposed to a decision problem (such as $\mathrm{P}^{\mathrm{NP[log]}}$). In particular, for 2-LH queries to the QMA oracle which violate the promise gap for 2-LH, the oracle is allowed to give an arbitrary answer. We observe that this point appears to have been missed in [5]'s claimed proofs of $\mathrm{P}^{\mathrm{QMA[log]}}$-hardness of APX-SIM and of $\mathrm{P}^{\mathrm{UQMA[log]}}$-hardness of SPECTRAL-GAP, rendering the proofs incorrect. Though, as shown by our first result, we are able to correct and improve on the proof that APX-SIM is $\mathrm{P}^{\mathrm{QMA[log]}}$-hard.

In our last result, we also overcome this difficulty to recover $\mathrm{P}^{\mathrm{UQMA[log]}}$-hardness of SPECTRAL-GAP, with the tradeoff that we obtain a "slightly weaker" hardness claim, in the sense that it follows from a Turing reduction as opposed to a Karp or mapping reduction[5] ([5] claimed hardness under the latter). In the process, we also improve the locality of H to $O(1)$.

Theorem 1.3. *Given a 4-local Hamiltonian H, estimating its spectral gap (i.e. SPECTRAL-GAP) is $\mathrm{P}^{\mathrm{UQMA[log]}}$-hard under polynomial-time Turing reductions.*

Thus, the ability to solve $\mathrm{P}^{\mathrm{UQMA[log]}}$ problems in polynomial time would imply a polynomial-time algorithm for SPECTRAL-GAP. Note that whether UQMA = QMA remains open. For the cases of NP [60] or MA and QCMA [4], it is known that Unique NP, Unique MA, and Unique QCMA reduce to NP, MA, and QCMA, respectively, under randomized reductions.

Upper bounds (containment in complexity classes). As mentioned earlier, since both QMA and co-QMA are contained in $\mathrm{P}^{\mathrm{QMA[log]}}$, it is likely that $\mathrm{P}^{\mathrm{QMA[log]}}$ is strictly more powerful than both QMA and co-QMA. This raises the question: *What is an upper bound on the power of $\mathrm{P}^{\mathrm{QMA[log]}}$?* Two points of reference are worth mentioning here. First, the classical analogue of $\mathrm{P}^{\mathrm{QMA[log]}}$, $\mathrm{P}^{\mathrm{NP[log]}}$, is known to be upper bounded by PP [9]. Here, PP is the set of promise problems solvable in probabilistic polynomial time with *unbounded* error; in other words, a PP machine accepts YES instances with probability strictly larger than $1/2$ and accepts NO instances with probability less than or equal to $1/2$. The second point of reference is that QMA is bounded by PP [44, 49] ([61] actually shows the slightly stronger containment QMA $\subseteq \mathrm{A_0PP}$). We thus ask whether, like its classical analogue, $\mathrm{P}^{\mathrm{QMA[log]}}$ can bounded by PP, and so is "slightly harder" than QMA. Indeed, we show this is the case.

Theorem 1.4. $\mathrm{P}^{\mathrm{QMA[log]}} \subseteq \mathrm{PP}$.

---

[5] A Karp, many-one, or mapping reduction from a decision problem A to a decision problem B maps any input Π_A for A into an input Π_B for B such that Π_A is a YES (NO) instance of A if and only if Π_B is a YES (NO) instance for B. A Turing reduction generalizes the notion of a Karp reduction; here, one is given access to an oracle solving B, and the goal is to solve Π_A using potentially multiple calls to the oracle. In this paper, both notions of reduction are assumed to be deterministic polynomial-time.

Thus, $\text{QMA} \subseteq \text{P}^{\text{QMA}[\log]} \subseteq \text{PP}$, which rigorously justifies the intuition that $\text{P}^{\text{QMA}[\log]}$ should be thought of as "slightly harder" than QMA.

1.5 Proof techniques

We now outline our proof techniques for both our upper bound and lower bound results.

Lower bounds. We begin with proof techniques techniques for our lower bounds involving APX-SIM (Theorem 1.1), APX-2-CORR (Theorem 1.2), and SPECTRAL-GAP (Theorem 1.3).

1. $\text{P}^{\text{QMA}[\log]}$-hardness of APX-SIM. To show Theorem 1.1, our intuition is simple: To design our local Hamiltonian H so that its ground state encodes a so-called history state[6] $|\psi\rangle$ for a given $\text{P}^{\text{QMA}[\log]}$ instance such that measuring observable Z on the designated "output qubit" of $|\psi\rangle$ reveals the answer of the computation. At a high level, this is achieved by combining a variant of Kitaev's circuit-to-Hamiltonian construction [45] (which forces the ground state to follow the P circuit) with Ambainis's "query Hamiltonian" [5] (which forces the ground state to encode correctly answered queries to the QMA oracle). Making this rigorous requires a careful analysis of the ground space of Ambainis's query Hamiltonian when queries violating the promise gap of the oracle are allowed (Lemma 3.1), showing a simple but useful extension of Kempe, Kitaev, and Regev's Projection Lemma [43] (Lemma 4.1) that any low energy state of H must be close to a valid history state, which thus allows us to use just a single-qubit observable, and applying Kitaev's unary encoding trick [45] to bring the locality of the Hamiltonian H down to $O(1)$ (Lemma 3.3).

2. Containment of APX-2-CORR in $\text{P}^{\text{QMA}[\log]}$. The hardness proof for APX-2-CORR is similar to that of APX-SIM, so we focus instead on the containment proof of APX-2-CORR $\in \text{P}^{\text{QMA}[\log]}$, for which a trick is required. For containment, the naive approach would be to run Ambainis's $\text{P}^{\text{QMA}[\log]}$ protocol [5] for APX-SIM independently for each term $\langle A \otimes B \rangle$, $\langle A \rangle$, and $\langle B \rangle$. However, if a cheating prover does not send the *same* ground state $|\psi\rangle$ for each of these measurements, soundness of the protocol can be violated.

To address this, we exploit a trick of Chailloux and Sattath [19] from the setting of QMA(2): we observe that the correlation function requires only knowledge of the two-body reduced density matrices $\{\rho_{ij}\}$ of $|\psi\rangle$. A prover can send classical descriptions of the $\{\rho_{ij}\}$ (which is possible since each ρ_{ij} is of constant size), along with a "consistency proof" for the QMA-complete Consistency problem [47] to ensure the $\{\rho_{ij}\}$ indeed correspond to some state. The verifier can then freely copy each ρ_{ij} and thus can calculate each term.

3. $\text{P}^{\text{UQMA}[\log]}$-hardness of estimating spectral gaps. As mentioned in Sections 1.3 and 1.4, a central theme of this work is the fact that a $\text{P}^{\text{QMA}[\log]}$ machine can make invalid queries to the QMA oracle, and this point appears to have been missed in [5], where all queries were assumed to satisfy the LH promise. This results in the proofs of two key claims of [5] being incorrect. The first claim was used in the proof of $\text{P}^{\text{QMA}[\log]}$-completeness for APX-SIM (Claim 1 in [5]); we provide a corrected

---

[6]A *history state* can be seen as a quantum analogue of the "tableau" which appears in the proof of the Cook-Levin theorem, i.e. a history state encodes the history of a quantum computation. In contrast to tableaus, however, the history encodes information in quantum superposition.

statement and proof in Lemma 3.1 (which suffices for the $P^{QMA[\log]}$-hardness results in [5] regarding APX-SIM to hold).

The error in the second claim (Claim 2 of [5]), wherein $P^{UQMA[\log]}$-hardness of determining the spectral gap of a local Hamiltonian is shown, appears arguably more serious. The construction of [5] requires a certain "query Hamiltonian" to have a spectral gap, which indeed holds if the $P^{UQMA[\log]}$ machine makes no invalid queries. However, if the machine makes invalid queries, this gap can close, and it is not clear how one can recover $P^{UQMA[\log]}$-hardness under mapping reductions.

To overcome this, we introduce a technique of "query validation": given a query to the UQMA oracle, we would like to determine if the query is valid or "far" from valid. While it is not clear how a P machine alone can perform such "query validation", we show how to use a SPECTRAL-GAP oracle to do so, allowing us to eliminate "sufficiently invalid" queries. Combining this idea with Ambainis's original construction [5], we show Theorem 1.3, i.e. $P^{UQMA[\log]}$-hardness for SPECTRAL-GAP, with the additional improvement that the input Hamiltonians are now $O(1)$-local, versus $O(\log n)$-local as in [5], via the same unary encoding trick from our previous results. Since our "query validation" requires a polynomial number of calls to the SPECTRAL-GAP oracle, this result requires a polynomial-time *Turing* reduction. Whether this can be improved to a mapping reduction is left as an open question.

Upper bounds. We now move to our upper bound result, which is the most technically involved. To show $P^{QMA[\log]} \subseteq PP$ (Theorem 1.4), we exploit the technique of *hierarchical voting*, used by Beigel, Hemachandra, and Wechsung [9] to show $P^{NP[\log]} \subseteq PP$, in conjunction with the QMA strong amplification results of Marriott and Watrous [49].

To give some intuition as to how the technique works, we sketch its application in the classical case of $P^{NP[\log]}$ [9]. There, we have a P machine making queries to an NP oracle for SAT (i.e. the i-th query feeds the oracle some SAT formulae ϕ_i), and the goal is to simulate this by a PP machine. Since the PP machine does not have access to an NP oracle, the best it can do is *guess* the answer to each query; so, it begins by guessing the answers to each NP query by picking a random assignment x_i for each formula ϕ_i in the hope that $\phi_i(x_i) = 1$. Of course, such a guess x_i can be satisfying only if ϕ_i is satisfiable to begin with; with a bit of thought, this implies that the "correct" query string y^* (i.e. the i-th bit of y^* equals 1 if and only if ϕ_i is satisfiable) is the lexicographically *largest* string y^* attainable by this guessing process.

Unfortunately, the PP machine might have a very small probability of guessing y^*; thus, it next "boosts" this probability via several rounds of "hierarchical voting". Conceptually, this technique does not actually *increase* the probability of outputting y^* but rather decreases the probability of outputting other, lexicographically smaller query strings $y' \neq y^*$. This works because the machine is a PP machine (i.e. it can have unbounded error), and so it suffices to reduce the probability of obtaining any such y' being output to be strictly smaller than the probability of y^* being output, in which case the PP machine is more likely to output the correct answer in its simulation of the $P^{NP[\log]}$ computation than not, as needed.

We develop a quantum variant of this scheme which is quite natural. However, its analysis is markedly more involved than the classical setting due to both the probabilistic nature of QMA and the possibility of "invalid queries" violating the QMA promise gap. To give a sense of the problems which arise, for $P^{QMA[\log]}$ it is no longer necessarily true that the lexicographically largest obtainable y^* is a correct query string.

9

1.6 Discussion and open questions

The problems studied here explore the line of research recently initiated by Ambainis [5] on $\mathrm{P}^{\mathrm{QMA[log]}}$ and focus on central problems for local Hamiltonian systems. The complexity theoretic study of such problems is appealing in that it addresses the original motivation of physicist Richard Feynman in proposing quantum computers [27], who was interested in avenues for simulating quantum systems. Indeed, hardness results, such as Kitaev's quantum Cook-Levin theorem, rigorously justify Feynman's intuition that such simulation problems are hard. Our work (e.g. Theorem 1.1) strongly supports this view by demonstrating that even some of the simplest and most natural simulation tasks, such as measuring a *single qubit (!)* of a ground state, can be harder than QMA. Our study of spectral gaps (Theorem 1.3) further highlights another theme: the subtleties which must be carefully treated when studying oracle classes for promise problems (such as $\mathrm{P}^{\mathrm{QMA[log]}}$). As quantum complexity theory commonly focuses on such promise problems, we believe this theme could be of interest to a broader computer science audience.

Moving to open questions, although we resolve one of the open questions from [5], there are others we leave open, along with some new ones. Do our results for APX-SIM and APX-2-CORR hold for more restricted classes of Hamiltonians, such as 2-local Hamiltonians, local Hamiltonians on a 2D lattice, or specific Hamiltonian models of interest (see e.g. [22, 53] for QMA-completeness results for estimating ground state energies of the spin-1/2 Heisenberg anti-ferromagnet)? Is SPECTRAL-GAP $\mathrm{P}^{\mathrm{UQMA[log]}}$-complete or $\mathrm{P}^{\mathrm{QMA[log]}}$-complete (recall SPECTRAL-GAP $\in \mathrm{P}^{\mathrm{QMA[log]}}$ and that [5] and our work together show $\mathrm{P}^{\mathrm{UQMA[log]}}$-hardness)? What is the relationship between $\mathrm{P}^{\mathrm{QMA[log]}}$ and $\mathrm{P}^{\mathrm{UQMA[log]}}$? Finally, what is the complexity of other physical tasks "beyond" estimating ground state energies?

Remark added later: Since the original preprint of this paper appeared, the present authors, together with Stephen Piddock, have partially answered the first open question above by showing [33] that APX-SIM remains $\mathrm{P}^{\mathrm{QMA[log]}}$-complete for any family of local Hamiltonians which can simulate "spatially sparse" [51] Hamiltonians. This, in turn, implies (e.g. using [53], [58], [54]) that APX-SIM remains $\mathrm{P}^{\mathrm{QMA[log]}}$-complete even on physically motivated models, such as the Heisenberg interaction on a 2D square lattice.

Organization. This paper is organized as follows: In Section 2, we give notation and formal definitions. In Section 3, we show various lemmas regarding Ambainis's query Hamiltonian. In Section 4.1, Section 4.2, and Section 4.3 we show Theorem 1.1, Theorem 1.2, and Theorem 1.3, respectively. Section 5 proves Theorem 1.4. A result used in Section 3 is proved in Appendix A.

2 Preliminaries

Notation. For $x \in \{0,1\}^n$, $|x\rangle \in (\mathbb{C}^2)^{\otimes n}$ denotes the computational basis state labeled by x. Let $\mathcal{X}$ be a complex Euclidean space. Then, $\mathrm{L}(\mathcal{X})$ and $\mathrm{D}(\mathcal{X})$ denote the sets of linear and density operators acting on $\mathcal{X}$, respectively. For subspace $\mathcal{S} \subseteq \mathcal{X}$, $\mathcal{S}^\perp$ denotes the orthogonal complement of $\mathcal{S}$. For Hermitian operator H, $\lambda(H)$ and $\lambda(H|_{\mathcal{S}})$ denote the smallest eigenvalue of H and the smallest eigenvalue of H restricted to space $\mathcal{S}$, respectively. The spectral and trace norms are defined $\|A\|_\infty := \max\{\|A|v\rangle\|_2 : \||v\rangle\|_2 = 1\}$ and $\|A\|_{\mathrm{tr}} := \mathrm{Tr}\sqrt{A^\dagger A}$, respectively, where $:=$ denotes a definition. We set $[m] := \{1, \ldots, m\}$.

Definitions and lemmas.

PP *and* PQP. The class PP [34] is the set of decision problems for which there exists a polynomial-time probabilistic Turing machine which accepts any YES instance with probability $> 1/2$ and accepts any NO instance with probability $\leq 1/2$. The quantum analogue of PP, denoted PQP, is defined analogously to BQP except in the YES case the verifier accepts with probability $> 1/2$ and in the NO case the verifier accepts with probability $\leq 1/2$.

$P^{QMA[\log]}$. Defined by Ambainis [5], $P^{QMA[\log]}$ is the set of decision problems decidable by a polynomial-time deterministic Turing machine with the ability to query an oracle for a QMA-complete problem $O(\log n)$ times, where n is the size of the input. For clarity, without loss of generality we assume that for any QMA problem instance Π the P machine wishes to solve, it applies a polynomial-time Karp/many-one reduction to map Π to an instance of the QMA-complete problem 2-LH [43] which it then feeds to the QMA oracle. 2-LH is defined as: Given a 2-local Hamiltonian H and inverse polynomially separated thresholds $a, b \in \mathbb{R}$, decide whether $\lambda(H) \leq a$ (YES-instance) or $\lambda(H) \geq b$ (NO-instance). Note that the P machine is allowed to make queries which violate the promise gap of 2-LH, i.e. with $\lambda(H) \in (a, b)$; in this case, the oracle can output YES or NO arbitrarily. The P machine is nevertheless required to output the same final answer (i.e. accept or reject) regardless of how such "invalid" queries are answered [36].

For any P machine M making m queries to a QMA oracle, we use the following terminology throughout this article. A *valid* (*invalid*) query satisfies (violates) the promise gap of the QMA oracle. A *correct* query string $y \in \{0, 1\}^m$ encodes a sequence of correct answers to all of the m queries. Note that for any invalid query by M, any answer is considered correct, yielding the possible existence of multiple correct query strings. An *incorrect* query string is one which contains at least one incorrect query answer.

The problem APX-SIM.

Definition 2.1 (APX-SIM(H, A, k, l, a, b, δ) (Ambainis [5]))**.** *Given a k-local Hamiltonian H, an l-local observable A, and real numbers a, b, and δ such that $b - a \geq n^{-c}$ and $\delta \geq n^{-c'}$, for n the number of qubits H acts on and $c, c' > 0$ some constants, decide:*

- *If H has a ground state $|\psi\rangle$ satisfying $\langle\psi| A |\psi\rangle \leq a$, output YES.*

- *If for all $|\psi\rangle$ satisfying $\langle\psi| H |\psi\rangle \leq \lambda(H) + \delta$, it holds that $\langle\psi| A |\psi\rangle \geq b$, output NO.*

Kitaev's circuit-to-Hamiltonian construction. Next, we briefly review Kitaev's circuit-to-Hamiltonian construction from the "quantum Cook-Levin theorem" [45]. Given a quantum circuit $U = U_L \cdots U_1$ consisting of 1- and 2-qubit gates U_i and acting on registers Q (proof register) and W (workspace register), this construction maps U to a 5-local Hamiltonian $H = H_{in} + H_{out} + H_{prop} + H_{stab}$ acting on registers Q (proof register), W (workspace register), and C (clock register). Each of H's terms

are defined below:

$$H_{\text{in}} \quad := \quad \left(\sum_{i=1}^{p} |1\rangle\langle 1|_{W_i} \right) \otimes |0\rangle\langle 0|_C$$

$$H_{\text{out}} \quad := \quad |0\rangle\langle 0|_{W_1} \otimes |L\rangle\langle L|_C$$

$$H_{\text{prop}} \quad := \quad \sum_{j=1}^{L} H_j, \text{ where } H_j \text{ is defined as}$$

$$-\frac{1}{2} U_j \otimes |j\rangle\langle j-1|_C - \frac{1}{2} U_j^\dagger \otimes |j-1\rangle\langle j|_C + \frac{1}{2} I \otimes (|j\rangle\langle j| + |j-1\rangle\langle j-1|)_C$$

$$H_{\text{stab}} \quad := \quad \sum_{i=1}^{L-1} |01\rangle\langle 01|_{C_i,C_{i+1}}.$$

Above, the notation R_i refers to the i-th qubit of any given register R. For any candidate proof $|\psi\rangle$ to be evaluated in expression $\text{Tr}(H |\psi\rangle\langle\psi|)$, each penalty term of H forces a structure on any minimizing $|\psi\rangle$: at time zero, H_{in} ensures the workspace register W is set to zero; H_{out} checks that at the end of the computation, i.e. at time step L, the output qubit is close to $|1\rangle$; the propagation Hamiltonian H_{prop} ensures all steps of U appear in superposition in $|\psi\rangle$ with equal weights. Finally, although we have labeled (for ease of exposition) the clock register in H_{out} and H_{prop} in binary, note that in Kitaev's construction it is actually encoded in unary. In other words, time t in clock register C is encoded as $|1^t 0^{L-t}\rangle$ (note for H_{stab} above, register C is already written in unary); H_{stab}'s role is thus to prevent invalid encodings of time steps. (For the interested reader, we remark that instead of working explicitly with Kitaev's construction, one can also consider the more general Quantum Thue System framework [7].)

In this paper, we shall use two key properties of $H_{\text{in}} + H_{\text{prop}} + H_{\text{stab}}$ (note H_{out} is omitted). First, the null space of $H_{\text{in}} + H_{\text{prop}} + H_{\text{stab}}$ is spanned by *history states*, which for any proof $|\psi\rangle$ have form

$$|\psi_{\text{hist}}\rangle = \frac{1}{\sqrt{L+1}} \sum_{t=0}^{L} U_t \cdots U_1 |\psi\rangle_Q |0\cdots 0\rangle_W |t\rangle_C, \tag{1}$$

where C is a clock register keeping track of time [45], and each U_i acts on at most two qubits from registers Q and W. Second, we use the following lower bound[7] on the smallest non-zero eigenvalue of $H_{\text{in}} + H_{\text{prop}} + H_{\text{stab}}$:

Lemma 2.2 (Lemma 3 (Gharibian, Kempe [29])). *The smallest non-zero eigenvalue of $\Delta(H_{\text{in}} + H_{\text{prop}} + H_{\text{stab}})$ is at least $\pi^2 \Delta/(64L^3) \in \Omega(\Delta/L^3)$, for $\Delta \in \mathbb{R}^+$ and $L \geq 1$.*

Miscellaneous useful facts. Let V denote a QMA verification circuit acting on M proof qubits with completeness c and soundness s. If one runs V on "proof" $\rho = I/2^M$, then for a YES instance, V accepts with probability $\geq c/2^M$ (since $I/2^M$ can be viewed as "guessing" a correct proof with probability $\geq 1/2^M$), and in a NO instance, V accepts with probability $\leq s$ (see, e.g., [49, 62]).

A useful fact for complex unit vectors $|v\rangle$ and $|w\rangle$ is (see, e.g., Equation 1.33 of [28])

$$\| |v\rangle\langle v| - |w\rangle\langle w| \|_{\text{tr}} = 2\sqrt{1 - |\langle v|w\rangle|^2} \leq 2 \| |v\rangle - |w\rangle \|_2. \tag{2}$$

---

[7]This bound is stated as $\Omega(\Delta/L^3)$ in [29]; the constant $\pi^2/64$ can be derived from the analysis therein .

3 Ambainis's Query Hamiltonian

We now show various results regarding Ambainis's "query Hamiltonian" [5], which intuitively aims to have its ground space contain correct answers to a sequence of QMA queries. Let U be a $\mathrm{P}^{\mathrm{QMA[log]}}$ computation[8], and let $H^{i,y_1\cdots y_{i-1}}_{\mathcal{Y}_i}$ be the 2-local Hamiltonian[9] corresponding to the i-th query made by U given that the answers to the previous $i-1$ queries are given by $y_1\cdots y_{i-1} \in \{0,1\}^{0,1}$. (Without loss of generality, we may assume $H^{i,y_1\cdots y_{i-1}}_{\mathcal{Y}_i} \succeq 0$ by adding multiples of the identity and rescaling.) The oracle query made at step i corresponds to an input $(H^{i,y_1\cdots y_{i-1}}_{\mathcal{Y}_i}, \epsilon, 3\epsilon)$ to 2-LH, with (as in [5]) completeness/"YES case" and soundness/"NO case" parameters ϵ and 3ϵ, respectively, for $\epsilon > 0$ a fixed inverse polynomial. Then, Ambainis's [5] $O(\log(n))$-local query Hamiltonian H acts on $\mathcal{X} \otimes \mathcal{Y}$, where $\mathcal{X} = (\mathcal{X}_i)^{\otimes m} = (\mathbb{C}^2)^{\otimes m}$ and $\mathcal{Y} = \otimes^m_{i=1}\mathcal{Y}_i$, such that $\mathcal{X}_i$ is intended to encode the answer to query i with $\mathcal{Y}_i$ encoding the ground state of the corresponding query Hamiltonian $H^{i,y_1\cdots y_{i-1}}_{\mathcal{Y}_i}$. Specifically,

$$
\begin{aligned}
H &= \sum_{i=1}^{m} \frac{1}{4^{i-1}} \sum_{y_1,\ldots,y_{i-1}} \bigotimes_{j=1}^{i-1} |y_j\rangle\langle y_j|_{\mathcal{X}_j} \otimes \left(2\epsilon |0\rangle\langle 0|_{\mathcal{X}_i} \otimes I_{\mathcal{Y}_i} + |1\rangle\langle 1|_{\mathcal{X}_i} \otimes H^{i,y_1\cdots y_{i-1}}_{\mathcal{Y}_i} \right) \\
&=: \sum_{i=1}^{m} \frac{1}{4^{i-1}} \sum_{y_1,\ldots,y_{i-1}} M_{y_1\cdots y_{i-1}}.
\end{aligned}
\tag{3}
$$

Recall from Section 2 that a sequence of query answers $y = y_1\cdots y_m \in \{0,1\}^m$ is *correct* if it corresponds to a possible execution of U. Since U can make queries to its QMA oracle which violate the QMA promise gap, the set of correct y is generally not a singleton. However, we henceforth assume without loss of generality that U makes at least one valid query (i.e. which satisfies the QMA promise gap). (Assume to the contrary that U makes only invalid queries. Then, since recall we stipulate [36] that the oracle's answer on any invalid query must not affect the final output of U, it follows that for *all* possible query strings, U outputs the same answer. Moreover, since $m \in O(\log n)$, we can efficiently *verify* this property — simply iterate via brute force through all possible polynomially many query strings, and check that U outputs the same answer for each one. Therefore, we can assume that any reduction includes such a check as an implicit subroutine, prepended to the start of the reduction, which will handle any case in which U makes only invalid queries.) We now prove the following about H:

Lemma 3.1. *Define for any $x \in \{0,1\}^m$ the space $\mathcal{H}_{x_1\cdots x_m} := \bigotimes_{i=1}^m |x_i\rangle\langle x_i| \otimes \mathcal{Y}_i$. Then, there exists a correct query string $x \in \{0,1\}^m$ such that the ground state of H lies in $\mathcal{H}_{x_1\cdots x_m}$. Moreover, suppose this space has minimum eigenvalue λ. Then, for any incorrect query string $y_1\cdots y_m$, any state in $\mathcal{H}_{y_1\cdots y_m}$ has energy at least $\lambda + \frac{\epsilon}{4^m}$.*

As discussed in Section 1, Claim 1 of [5] proved a similar statement under the assumption that the correct query string x is unique. In that setting, [5] showed the ground state of H is in $\mathcal{H}_x$, and that for *all* query strings $y \neq x$, the space $\mathcal{H}_y$ has energy at least $\lambda + \frac{\epsilon}{4^{m-1}}$. However, in general invalid

---

[8]In this work, we let U denote a uniformly generated classical circuit, as opposed to a Turing machine; this is to ease the transition to quantum circuits and corresponding application of Kitaev's circuit-to-Hamiltonian construction later.

[9]While each query is encoded as a 2-local Hamiltonian, our hardness results in Section 4 will be for 5- and 4-local Hamiltonians, since the query Hamiltonians are just components of our larger, final Hamiltonian constructions.

queries must be allowed, and in this setting this claim no longer holds — two distinct correct query strings can have eigenvalues which are arbitrarily close if they contain queries violating the promise gap. The key observation we make here is that even in the setting of non-unique x, a spectral gap between the ground space and all *incorrect* query strings can be shown.

Proof of Lemma 3.1. Observe first that H in Equation (3) is block-diagonal with respect to register $\mathcal{X}$, i.e. to understand the spectrum of H, it suffices to understand the eigenvalues in each of the blocks corresponding to fixing $\mathcal{X}_i$ to some string $y \in \{0,1\}^m$. Thus, we can restrict our attention to spaces $\mathcal{H}_y$ for $y \in \{0,1\}^m$. To begin, let $x \in \{0,1\}^m$ denote a correct query string which has lowest energy among all *correct* query strings against H, i.e. the block corresponding to x has the smallest eigenvalue among such blocks. (Note that x is well-defined, though it may not be unique; in this latter case, any such x will suffice for our proof.) For any $y \in \{0,1\}^m$, define λ_y as the smallest eigenvalue in block $\mathcal{H}_y$. We show that for any *incorrect* query string $y = y_1 \cdots y_m$, $\lambda_y \geq \lambda_x + \epsilon/(4^m)$.

We use proof by contradiction, via an exchange argument[10]. Suppose there exists an incorrect query string $y = y_1 \cdots y_m$ such that $\lambda_y < \lambda_x + \epsilon/(4^m)$. Since y is an incorrect query string, there exists an $i \in [m]$ such that y_i is the wrong answer to a valid query $H_{y_i}^{i,y_1 \cdots y_{i-1}}$. Let i denote the first such position. Now, consider operator $M_{y_1 \cdots y_{i-1}}$, which recall is defined as

$$M_{y_1 \cdots y_{i-1}} = \bigotimes_{j=1}^{i-1} |y_j\rangle\langle y_j|_{\mathcal{X}_j} \otimes \left(2\epsilon |0\rangle\langle 0|_{\mathcal{X}_i} \otimes I_{\mathcal{Y}_i} + |1\rangle\langle 1|_{\mathcal{X}_i} \otimes H_{y_i}^{i,y_1 \cdots y_{i-1}} \right)$$

and let $\lambda_{y_1 \cdots y_{i-1}\overline{y}_i}$ denote the smallest eigenvalue of $M_{y_1 \cdots y_{i-1}}$ restricted to space $\mathcal{H}_{y_1 \cdots y_{i-1}\overline{y}_i}$, where $\overline{y}_i$ denotes the complement of bit y_i; for clarity, here we define

$$\mathcal{H}_{y_1 \cdots y_{i-1}\overline{y}_i} := |y_1\rangle\langle y_1|_{\mathcal{X}_1} \otimes \cdots \otimes |\overline{y}_i\rangle\langle \overline{y}_i|_{\mathcal{X}_i} \otimes \left(\bigotimes_{j=i+1}^{m} \mathcal{X}_j \right) \otimes \mathcal{Y}.$$

Note that string $y_1 \cdots y_{i-1}\overline{y}_i$ corresponds to i correct query bits, with $\overline{y}_i$ the correct answer to query i. Then, we claim that any state $|\phi\rangle \in \mathcal{H}_{y_1 \cdots y_i}$ (i.e. $|\phi\rangle$ is of the same dimension as H, with registers $\mathcal{X}_1$ through $\mathcal{X}_i$ fixed to state $|y_1\rangle \otimes \cdots \otimes |y_i\rangle$) satisfies the bound of the following lemma.

Lemma 3.2. *For $|\phi\rangle$, $M_{y_1 \cdots y_{i-1}}$, and $\lambda_{y_1 \cdots y_{i-1}\overline{y}_i}$ as defined above, we have*

$$\langle \phi | M_{y_1 \cdots y_{i-1}} |\phi\rangle \geq \lambda_{y_1 \cdots y_{i-1}\overline{y}_i} + \epsilon.$$

Here, for clarity, $M_{y_1 \cdots y_{i-1}}$ is of the same dimension as $|\phi\rangle$, since we assume $M_{y_1 \cdots y_{i-1}}$ has an implicit tensor product with term $I_{\mathcal{X}_{i+1} \otimes \mathcal{Y}_{i+1} \cdots \otimes \mathcal{X}_m \otimes \mathcal{Y}_m}$. (This is intuitively because $M_{y_1 \cdots y_{i-1}}$ corresponds to making query i given responses to queries 1 through $i-1$, and queries $i+1$ through m are yet to be determined.)

---

[10]At a high level, in an *exchange argument*, one begins with a solution y to some problem instance, where y is lacking some additional desired property P. One then shows how to "exchange" y for another solution y' such that y' is at least as "good" as y for the problem at hand, and so that y' now has the desired property P. In our application here, we will exchange query string y for a query string y' satisfying $\lambda_{y'} < \lambda_y$. The property P in this case will be whether the query string correct.

Proof. Constrained to space $\mathcal{H}_{y_1 \cdots y_{i-1}}$, $M_{y_1 \cdots y_{i-1}}$ reduces to operator $M' := 2\epsilon |0\rangle\langle 0|_{\mathcal{X}_i} \otimes I_{\mathcal{Y}_i} + |1\rangle\langle 1|_{\mathcal{X}_i} \otimes H_{\mathcal{Y}_i}^{i, y_1 \cdots y_{i-1}}$, which is block-diagonal in the standard basis with respect to $\mathcal{X}_i$. Thus, if query i is a YES instance, the smallest eigenvalue of M' lies in the block corresponding to setting $\mathcal{X}_i$ to (the correct query answer) $|1\rangle$, and is at most ϵ (since we are assuming each query has YES and NO energy thresholds ϵ and 3ϵ). On the other hand, the block with $\mathcal{X}_i$ set to $|0\rangle$ by definition reduces to $2\epsilon I_{\mathcal{Y}_i}$, and hence has all eigenvalues equaling 2ϵ. Thus, setting $\mathcal{X}_i$ to the correct query answer $|1\rangle$ yields an energy penalty which is at least ϵ less than the penalty obtained for setting $\mathcal{X}_i$ to the wrong query answer, $|0\rangle$. Analogously, when query i is a NO instance (i.e. the correct query answer is $|0\rangle$), a similar argument shows the $|0\rangle$-block again has eigenvalues equaling 2ϵ and now the $|1\rangle$-block has eigenvalues at least 3ϵ. We conclude that in both YES and NO cases, setting $\mathcal{X}_i$ to the correct query answer achieves an energy penalty at least ϵ less than setting $\mathcal{X}_i$ to the wrong query answer. As a result,

$$\lambda_{y_1 \cdots y_{i-1} \bar{y}_i} \leq \lambda_{y_1 \cdots y_i} - \epsilon,$$

which implies the claim. $\qquad\square$

With Lemma 3.2 in hand, we return to our exchange argument. Let $\widehat{M}_{y_1 \cdots y_t}$ denote the set of terms from Equation (3) which are consistent with some prefix $y_1 \ldots y_t$ (e.g. $M_{y_1 \cdots y_t}$, $M_{y_1 \cdots y_t 0}$, $M_{y_1 \cdots y_t 1}$, etc). Recall that the bits $y_{i+1} \cdots y_m$ form the tail of the string y following the incorrect query answer y_i. Set $y'_{i+1} \cdots y'_m$ so that $y' := y_1 \cdots \bar{y}_i y'_{i+1} \cdots y'_m$ is a correct query string. Care is now required; the new query bits $y'_{i+1} \cdots y'_m$ may lead to different energy penalties than the previous string $y_{i+1} \cdots y_m$ against the Hamiltonian terms in set $\widehat{M}_{y_1 \cdots \bar{y}_i}$. In other words, we must upper bound any possible energy penalty *increase* when exchanging $y_1 \cdots \bar{y}_i y_{i+1} \cdots y_m$ for y'. To do so, recall that all Hamiltonian terms in Equation (3) are positive semidefinite. Thus, for any state $|\psi\rangle$ in space $\mathcal{H}_{y_1 \cdots \bar{y}_i}$, the energy obtained by $|\psi\rangle$ against terms in $\widehat{M}_{y_1 \cdots \bar{y}_i}$ is at least 0. Conversely, in the worst case, since each term in $\widehat{M}_{y_1 \cdots \bar{y}_i}$ has minimum eigenvalue at most 2ϵ, the eigenvector $|\psi\rangle$ of smallest eigenvalue in block $H_{y'}$ incurs an additional penalty against H for queries $i+1$ through m of at most (taking into account the normalization factor $1/4^{i-1}$ in Equation (3))

$$2\epsilon \sum_{k=i}^{\infty} \frac{1}{4^k} = \frac{2\epsilon}{3 \cdot 4^{i-1}}.$$

We conclude that (again taking into account the normalization factor $1/4^{i-1}$ in Equation (3))

$$\lambda_{y'} \leq \lambda_y - \frac{\epsilon}{4^{i-1}} + \frac{2\epsilon}{3 \cdot 4^{i-1}} < \left(\lambda_x + \frac{\epsilon}{4^m} \right) - \frac{\epsilon}{4^{i-1}} + \frac{2\epsilon}{3 \cdot 4^{i-1}} < \lambda_x$$

where the second inequality follows by the assumption $\lambda_y < \lambda_x + \epsilon/4^m$. This is a contradiction. $\quad\square$

The next lemma converts H from an $O(\log n)$-local Hamiltonian to an $O(1)$-local one.

Lemma 3.3. *For any $x \in \{0,1\}^m$, let $\hat{x}$ denote its unary encoding. Then, for any $\mathrm{P}^{\mathrm{QMA}[\log]}$ circuit U acting on n bits and making $m \geq 1$ queries to a QMA oracle, there exists a mapping to a 4-local Hamiltonian H' acting on space $(\mathbb{C}^2)^{\otimes 2^m - 1} \otimes \mathcal{Y}$ such that there exists a correct query string $x = x_1 \cdots x_m$ satisfying:*

1. *The ground state of H' lies in subspace $|\hat{x}\rangle\langle \hat{x}| \otimes \mathcal{Y}$.*

2. *For any state $|\psi\rangle$ in subspace $|\hat{x}'\rangle\langle\hat{x}'| \otimes \mathcal{Y}$ where either $\hat{x}'$ is not a unary encoding of a binary string x' or x' is an incorrect query string, one has $\langle\psi| H' |\psi\rangle \geq \lambda(H') + \epsilon/4^m$, for inverse polynomial ϵ.*

3. *For all strings $x' \in \{0,1\}^m$, H' acts invariantly on subspace $|\hat{x}'\rangle\langle\hat{x}'| \otimes \mathcal{Y}$.*

4. *The mapping can be computed in time polynomial in n (recall $m \in O(\log n)$).*

Proof. We show how to improve the $O(\log n)$-local construction H of Lemma 3.1 to 4-local H'. Specifically, recall that H from Equation (3) acts on $(\mathcal{X} \otimes \mathcal{Y})$. Using a trick of Kitaev [45], we encode the $\mathcal{X} = \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_m$ register in unary. Specifically, we can write

$$M_{y_1 \cdots y_{i-1}} = \sum_{y_{i+1},\ldots,y_m} 2\epsilon \bigotimes_{j=1}^{i-1} |y_j\rangle\langle y_j|_{\mathcal{X}_j} \otimes |0\rangle\langle 0|_{\mathcal{X}_i} \bigotimes_{k=i+1}^{m} |y_k\rangle\langle y_k|_{\mathcal{X}_k} \otimes I_{\mathcal{Y}} +$$

$$\bigotimes_{j=1}^{i-1} |y_j\rangle\langle y_j|_{\mathcal{X}_j} \otimes |1\rangle\langle 1|_{\mathcal{X}_i} \bigotimes_{k=i+1}^{m} |y_k\rangle\langle y_k|_{\mathcal{X}_k} \otimes H_{\mathcal{Y}_i}^{i,y_1\cdots y_{i-1}}.$$

We now replace register $\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_m$ with register $\mathcal{X}' = (\mathbb{C}^2)^{\otimes 2^m - 1}$ and encode each binary string $x \in \{0,1\}^m$ as the unary string $\hat{x} = |1\rangle^{\otimes |x|} |0\rangle^{\otimes 2^m - |x| - 1}$, where $|x|$ is the non-negative integer corresponding to string x. In other words, for $M_{x_1 \cdots x_{i-1}}$, we replace each string $|x\rangle\langle x|_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_m}$ with $|1\rangle\langle 1|_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_{|x|}} \otimes |0\rangle\langle 0|_{\mathcal{X}_{|x|+1} \otimes \cdots \otimes \mathcal{X}_{2^m - 1}}$. Denote the resulting Hamiltonian as H_1.

To ensure states in $\mathcal{X}'$ follow this encoding, add a weighted version of Kitaev's [45] penalty Hamiltonian, introduced in Section 2,

$$H_{\text{stab}} = 3\epsilon \sum_{j=1}^{2^m - 2} |0\rangle\langle 0|_j \otimes |1\rangle\langle 1|_{j+1},$$

i.e., our final Hamiltonian is $H' = H_1 + H_{\text{stab}}$. To show that H' satisfies the same properties as H as stated in the claim, we follow the analysis of Kitaev [45]. Namely, partition the space $\mathcal{X}' \otimes \mathcal{Y}$ into orthogonal spaces $\mathcal{S}$ and $\mathcal{S}^\perp$ corresponding to the space of valid and invalid unary encodings of $\mathcal{X}'$, respectively. Since H_1 and H_{stab} act invariantly on $\mathcal{S}$ and $\mathcal{S}^\perp$, we can consider $\mathcal{S}$ and $\mathcal{S}^\perp$ separately. In $\mathcal{S}$, H' is identical to H, implying the claim. In $\mathcal{S}^\perp$, the smallest non-zero eigenvalue of H_{stab} is at least 3ϵ. Thus, since $H_1 \succeq 0$, if we can show that the smallest eigenvalue of H is at most $3\epsilon - \epsilon/4^m$, we have shown the claim (since, in particular, we will have satisfied statement 2 of our claim). To show this bound on the smallest eigenvalue, suppose x is all zeros, i.e. set register $\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_m$ for H to all zeros. Then, each term $M_{0_1 \cdots 0_{i-1}}$ yields an energy penalty of exactly 2ϵ, yielding an upper bound on the smallest eigenvalue of H of $2\epsilon \sum_{k=0}^{m-1} \frac{1}{4^k} \leq \frac{8}{3}\epsilon = 3\epsilon - \epsilon/3$. □

4 Lower bounds (hardness results)

4.1 $\mathsf{P}^{\mathsf{QMA[log]}}$-completeness of APX-SIM

In this section, we restate and prove Theorem 1.1. For this, we first show an extension of the Projection Lemma of Kempe, Kitaev, Regev [43].

Lemma 4.1 (Extended Projection Lemma (*cf. [43]*))**.** *Let* $H = H_1 + H_2$ *be the sum of two Hamiltonians operating on some Hilbert space* $\mathcal{H} = \mathcal{S} + \mathcal{S}^\perp$. *The Hamiltonian* H_1 *is such that* $\mathcal{S}$ *is a zero eigenspace and the eigenvectors in* $\mathcal{S}^\perp$ *have eigenvalue at least* $J > 2\|H_2\|_\infty$. *Let* $K := \|H_2\|_\infty$. *Then, for any* $\delta \geq 0$ *and* $|\psi\rangle$ *satisfying* $\langle\psi| H |\psi\rangle \leq \lambda(H) + \delta$, *there exists a* $|\psi'\rangle \in \mathcal{S}$ *such that:*

- *(Ground state energy bound)*

$$\lambda(H_2|_{\mathcal{S}}) - \frac{K^2}{J - 2K} \leq \lambda(H) \leq \lambda(H_2|_{\mathcal{S}}),$$

  where $\lambda(H_2|_{\mathcal{S}})$ *denotes the smallest eigenvalue of* H_2 *restricted to space* $\mathcal{S}$.

- *(Ground state deviation bound)*

$$|\langle\psi|\psi'\rangle|^2 \geq 1 - \left(\frac{K + \sqrt{K^2 + \delta(J - 2K)}}{J - 2K}\right)^2.$$

- *(Energy obtained by perturbed state against* H*)*

$$\langle\psi'| H |\psi'\rangle \leq \lambda(H) + \delta + 2K\frac{K + \sqrt{K^2 + \delta(J - 2K)}}{J - 2K}.$$

Proof. The first claim is the original Projection Lemma [43] (which we do not require in this work, but which we also state above for completeness). The second and third claims we prove here. Consider arbitrary $|\psi\rangle$ such that $\langle\psi| H |\psi\rangle \leq \lambda(H) + \delta$. We can write $|\psi\rangle = \alpha_1 |\psi_1\rangle + \alpha_2 |\psi_2\rangle$ for unit vectors $|\psi_1\rangle \in \mathcal{S}$, $|\psi_2\rangle \in \mathcal{S}^\perp$, $\alpha_1, \alpha_2 \in \mathbb{R}$, $\alpha_1, \alpha_2 \geq 0$, and $\alpha_1^2 + \alpha_2^2 = 1$. Following the proof of the Projection Lemma of [43], we have

$$
\begin{aligned}
\langle\psi| H |\psi\rangle &\geq \langle\psi| H_2 |\psi\rangle + J\alpha_2^2 \\
&= (1 - \alpha_2^2) \langle\psi_1| H_2 |\psi_1\rangle + 2\alpha_1\alpha_2 \operatorname{Re} \langle\psi_1| H_2 |\psi_2\rangle + \\
&\quad \alpha_2^2 \langle\psi_2| H_2 |\psi_2\rangle + J\alpha_2^2 \\
&\geq \langle\psi_1| H_2 |\psi_1\rangle - K(\alpha_2^2 + 2\alpha_2 + \alpha_2^2) + J\alpha_2^2 \\
&= \langle\psi_1| H |\psi_1\rangle + (J - 2K)\alpha_2^2 - 2K\alpha_2, \quad\quad (4)
\end{aligned}
$$

where the last equality follows since $|\psi_1\rangle \in \mathcal{S}$. Since by assumption $\langle\psi| H |\psi\rangle \leq \lambda(H) + \delta$, Equation (4) implies $\lambda(H) + \delta \geq \langle\psi_1| H |\psi_1\rangle + (J - 2K)\alpha_2^2 - 2K\alpha_2$. Combined with the fact $\lambda(H) \leq \langle\psi_1| H |\psi_1\rangle$, we have

$$0 \geq \lambda(H) - \langle\psi_1| H |\psi_1\rangle \geq (J - 2K)\alpha_2^2 - 2K\alpha_2 - \delta,$$

which holds only if $|\alpha_2| \leq \frac{K + \sqrt{K^2 + \delta(J - 2K)}}{J - 2K}$. Thus, setting $|\psi'\rangle := |\psi_1\rangle$ yields the first claim. To see that $|\psi'\rangle$ also satisfies the second claim, by Equation (4) we have

$$\lambda(H) + \delta \geq \langle\psi| H |\psi\rangle \geq \langle\psi_1| H |\psi_1\rangle + (J - 2K)\alpha_2^2 - 2K\alpha_2.$$

The assumption $J > 2K$ implies $(J - 2K)\alpha_2^2 \geq 0$, and so

$$\langle\psi_1| H |\psi_1\rangle \leq \lambda(H) + \delta + 2K |\alpha_2|.$$

Plugging in our upper bound on $|\alpha_2|$ now yields the second claim. $\qquad\square$

With Lemma 4.1 in hand, we now prove the main result of this section.

Theorem 1.1. APX-SIM *is* $P^{QMA[\log]}$-*complete for* $k = 5$ *and* $l = 1$, *i.e., for* 5-*local Hamiltonian* H *and* 1-*local observable* A.

Proof. Containment in $P^{QMA[\log]}$ was shown[11] for $k, l \in O(\log n)$ in [5], for n the input size. We show $P^{QMA[\log]}$-hardness of APX-SIM via a polynomial-time mapping/Karp reduction.

Let U' be an arbitrary $P^{QMA[\log]}$ circuit[12] for instance Π, such that U' acts on workspace register W and query result register Q. Suppose U' consists of L' gates and makes $m = c \log(n)$ queries, for $c \in O(1)$. Without loss of generality, U' can be simulated with a similar unitary U which treats Q as a *proof* register which it does not alter at any point: Namely, U does not have access to a QMA oracle, but rather reads bit Q_i whenever it desires the answer to the i-th query. Thus, if a correct query string $y_1 \cdots y_m$ corresponding to an execution of U' on input x is provided in Q as a "proof", then the output statistics of U' and U are identical. We can also assume that Q is encoded not in binary, but in unary. Thus, Q consists of $2^m - 1 \in \text{poly}(n)$ bits. For simplicity, however, in our discussion we will speak of m-bit query strings $y = y_1 \cdots y_m$ in register Q.

The construction. We now construct our instance of APX-SIM. First, we map U to a 5-local Hamiltonian H_1 via a modification of the circuit-to-Hamiltonian construction of Kitaev [45], such that H_1 acts on registers W (workspace register), Q (proof register), and C (clock register). Recall (Section 2) that Kitaev's construction outputs Hamiltonian terms $H_{\text{in}} + H_{\text{prop}} + H_{\text{stab}} + H_{\text{out}}$. Set $H_1 = \Delta(H_{\text{in}} + H_{\text{prop}} + H_{\text{stab}})$ for Δ to be set as needed. In contrast to [45], it is crucial that H_{out} be omitted from H_1, as we require our final Hamiltonian H to enforce a certain structure on the ground space *regardless* of whether the computation should accept or reject. The job of "checking the output" will instead be delegated to the observable A. Formally, following the discussion in Section 2 regarding Equation (1), H_1 has a non-trivial null space, which is its ground space, consisting of history states $|\psi_{\text{hist}}\rangle$ (see Equation (1)) which simulate U on registers W and Q. These history states correctly simulate U' *assuming that* Q is initialized to a correct proof.

To thus enforce that Q is initialized to a correct proof, let H_2 be our variant H' of Ambainis's query Hamiltonian from Lemma 3.3, such that H_2 acts on registers Q and Q', where for clarity $Q = (\mathbb{C}^2)^{\otimes 2^m - 1}$ for $m \in O(\log n)$; moreover, setting $Q' = \mathcal{Y}$ from Lemma 3.3 enforces the ground space of H_2 to contain unary encodings of valid strings of query answers $y_1 \cdots y_m$ for input Π in register Q, as desired. Our final Hamiltonian is $H = H_1 + H_2$, which is 5-local since H_1 is 5-local and H_2 is 4-local.

Suppose without loss of generality that U's output qubit is W_1, which is set to $|0\rangle$ until the final time step, in which the correct output is copied to it. Then, set observable $A = (I + Z)/2$ such that A acts on qubit W_1. Set $a = 1 - 1/(L+1)$, and $b = 1 - 1/2L$ for L the number of gates in U. Fix $\eta \geq \max(\|H_2\|_\infty, 1)$ (such an η can be efficiently computed by applying the triangle

---

[11] Intuitively, the idea is as follows. First, the P machine uses $O(\log n)$ queries to perform a version of binary search tailored to promise problems to estimate the ground state energy of the given Hamiltonian H to within additive inverse polynomial error; call this estimate g. Then, the P machine makes one final QMA query: Does H have a ground state $|\psi\rangle$ with ground state energy approximately g and with small expectation against the given observable A?

[12] As noted in Section 3, the base class P is normally defined using a polynomial-time deterministic Turing machine. In this paper, we assume the action of the P machine is, without loss of generality, instead expressed via a polynomial-time uniform family of quantum circuits.

inequality and summing the spectral norms of each term of H_2 individually). Set $\Delta = 128L^3\eta\gamma$ for γ a monotonically increasing polynomial function of L to be set as needed. Finally, set $\delta = 1/\Delta$. This completes the construction.

Correctness. We now prove correctness, meaning that if Π is a YES (NO) instance of a $\mathrm{P}^{\mathrm{QMA[log]}}$ problem, then APX-SIM$(H, A, 5, 1, a, b, \delta)$ is a YES (NO) instance of APX-SIM. The following pair of lemmas together show this. They both assume the terminology set up in this proof thus far.

Lemma 4.2. *Suppose Π is a YES instance of a $\mathrm{P}^{\mathrm{QMA[log]}}$ problem. Then, $(H, A, k, l, a, b, \delta)$ is a YES instance of* APX-SIM.

Proof. By Lemma 3.3, the ground space of H_2 is spanned by states of the form $|\hat{x}\rangle_Q \otimes |\phi\rangle_{Q'}$ where $\hat{x}$ is a correct query string encoded in unary. Fix an arbitrary such ground state $|\hat{x}\rangle_Q \otimes |\phi\rangle_{Q'}$. Since Π is a YES instance, setting Q to $\hat{x}$ in this manner causes $U = U_L \cdots U_1$ to accept with certainty. Consider a version of the history state $|\psi_{\mathrm{hist}}\rangle$ from Equation (1) on registers W (workspace), C (clock), Q, and Q' (Q and Q' together are the "proof register", where recall U reads query answers from Q but does not access Q' at any point),

$$|\psi_{\mathrm{hist}}\rangle = \frac{1}{\sqrt{L+1}} \sum_{t=0}^{L} U_t \cdots U_1 |\hat{x}\rangle_Q |\phi\rangle_{Q'} |0\cdots0\rangle_W |\hat{t}\rangle_C,$$

which by construction lies in the ground space of H_1. Since U can read but does not alter the contents of Q, the history state has the tensor product form

$$|\psi_{\mathrm{hist}}\rangle = |\psi'_{\mathrm{hist}}(x)\rangle_{W,C} \otimes |\hat{x}\rangle_Q \otimes |\phi\rangle_{Q'}$$

for appropriately defined $|\psi'_{\mathrm{hist}}(x)\rangle_{W,C}$. Since the state encodes a correct query string in register Q, $|\psi_{\mathrm{hist}}\rangle$ lies in the ground space of H_2. We conclude that $|\psi_{\mathrm{hist}}\rangle$ is in the joint ground space of H_1 and H_2, and hence in the ground space of $H = H_1 + H_2$. Finally, by assumption, the output qubit W_1 of the construction is set to $|0\rangle$ in timesteps 0 to $L - 1$. In timestep L, since U accepts with certainty given the correct query string $\hat{x}$, W_1 is set to $|1\rangle$. It follows that, as desired,

$$\langle\psi_{\mathrm{hist}}| A |\psi_{\mathrm{hist}}\rangle = \frac{1}{2} \langle\psi_{\mathrm{hist}}| (I + Z) |\psi_{\mathrm{hist}}\rangle = 1 - \frac{1}{L+1} = a.$$

$\square$

Lemma 4.3. *Suppose Π is a NO instance of a $\mathrm{P}^{\mathrm{QMA[log]}}$ problem. Then, $(H, A, k, l, a, b, \delta)$ is a NO instance of* APX-SIM.

Proof. Consider any low energy state $|\psi\rangle$ satisfying $\langle\psi| H |\psi\rangle \leq \lambda(H) + \delta$. By Lemma 2.2, the smallest non-zero eigenvalue of H_1 is at least $J = \pi^2\Delta/(64L^3) = \pi^2\eta\gamma/64$. Recalling that $\delta = 1/\Delta$, apply Lemma 4.1 to obtain that there exists a valid history state $|\psi'\rangle$ on W, C, Q, and Q' such that $|\langle\psi|\psi'\rangle|^2 \geq 1 - O(\gamma^{-2})$ (see Lemma A.1 in Appendix A for a proof), which by Equation (2) implies

$$\| |\psi\rangle\langle\psi| - |\psi'\rangle\langle\psi'| \|_{\mathrm{tr}} \leq \frac{c}{\gamma} \tag{5}$$

for some constant $c > 0$. By definition, such a history state $|\psi'\rangle$ simulates U given "quantum proof" $|\phi\rangle_{Q,Q'}$ in registers Q and Q', i.e.

$$|\psi'\rangle = \sum_t U_t \cdots U_1 |\phi\rangle_{Q,Q'} |0 \cdots 0\rangle_W |t\rangle_C.$$

Lemma 4.1 and the proof of Lemma A.1 also give that

$$\langle\psi'| H |\psi'\rangle \leq \lambda(H) + \delta + O\left(\frac{\eta}{\gamma^2}\right) =: \lambda(H) + \delta + \gamma'.$$

Unfortunately, *a priori* the state $|\phi\rangle$ in the proof register (Q, Q') of $|\psi'\rangle$ is arbitrary. Let us now approximate $|\psi'\rangle$ with another history state $|\psi''\rangle$, which contains a string of *correct* query answers in Q. This is accomplished by the following lemma, which assumes the definitions introduced in this proof thus far.

Lemma 4.4. *There exists a history state $|\psi''\rangle$ such that*

$$\||\psi\rangle\langle\psi| - |\psi''\rangle\langle\psi''|\|_{\mathrm{tr}} \leq \frac{c}{\gamma} + 2\sqrt{\frac{4^m(\delta + \gamma')}{\epsilon}} \tag{6}$$

and where register Q of $|\psi''\rangle$ contains only correct query answers when written in the standard basis.

Proof. By Lemma 3.3, the ground space $\mathcal{G}$ of H_2 is contained in the span of states of the form $|\hat{x}\rangle_Q \otimes |\phi'\rangle_{Q'}$ where $\hat{x}$ is a correct query string encoded in unary. Since the ground spaces of H_1 and H_2 have non-empty intersection, i.e. history states acting on "quantum proofs" from $\mathcal{G}$ (which lie in the null space of H_1 and obtain energy $\lambda(H_2)$ against H_2), we know $\lambda(H) = \lambda(H_2)$. Thus, since $H_1 \succeq 0$,

$$\langle\psi'| H_2 |\psi'\rangle \leq \langle\psi'| H |\psi'\rangle \leq \lambda(H_2) + (\delta + \gamma'). \tag{7}$$

Write $|\phi\rangle = \alpha |\phi_1\rangle + \beta |\phi_2\rangle$ for $|\phi_1\rangle \in \mathrm{Span}\{ |\hat{x}\rangle_Q \otimes |\phi'\rangle_{Q'} \mid \text{correct query string } x \}$ and $|\phi_2\rangle \in \mathrm{Span}\{ |\hat{x}\rangle_Q \otimes |\phi'\rangle_{Q'} \mid \text{incorrect query string } x \}$ ($|\phi_1\rangle, |\phi_2\rangle$ normalized), $\alpha, \beta \in \mathbb{C}, |\alpha|^2 + |\beta|^2 = 1$. Since any history state $|\psi'\rangle$, for any amplitudes α_x and unit vectors $|\phi'_x\rangle$, has form

$$\sum_{t,x} \alpha_x U_t \cdots U_1 |0 \cdots 0\rangle_W |t\rangle_C |\hat{x}\rangle_Q |\phi'_x\rangle_{Q'} = \sum_x \alpha_x |\psi'_{\mathrm{hist}}(x)\rangle_{W,C} |\hat{x}\rangle_Q |\phi'_x\rangle_{Q'}$$

(i.e. for any fixed x, $|\hat{x}\rangle_Q$ is not altered), and since H_2 is block-diagonal with respect to the standard basis in Q, by Equation (7) and Lemma 3.3 we have

$$\begin{aligned}
\lambda(H_2) + (\delta + \gamma') &\geq \langle\psi'| H_2 |\psi'\rangle = |\alpha|^2 \langle\phi_1| H_2 |\phi_1\rangle + |\beta|^2 \langle\phi_2| H_2 |\phi_2\rangle \\
&\geq |\alpha|^2 \lambda(H_2) + |\beta|^2 \left(\lambda(H_2) + \frac{\epsilon}{4^m}\right),
\end{aligned}$$

which implies $|\beta|^2 \leq 4^m(\delta + \gamma')/\epsilon$. Thus, defining $|\psi''\rangle$ as the history state for "proof" $|\phi_1\rangle_{Q,Q'}$, we have that $\||\psi\rangle\langle\psi| - |\psi''\rangle\langle\psi''|\|_{\mathrm{tr}}$ is at most

$$\||\psi\rangle\langle\psi| - |\psi'\rangle\langle\psi'|\|_{\mathrm{tr}} + \||\phi\rangle\langle\phi| - |\phi_1\rangle\langle\phi_1|\|_{\mathrm{tr}} \leq \frac{c}{\gamma} + 2\sqrt{\frac{4^m(\delta + \gamma')}{\epsilon}},$$

which follows from the triangle inequality, Equation (5), and the structure of the history state. $\quad\square$

With Lemma 4.4 in hand, we are ready to complete the proof of Lemma 4.3. Observe that increasing γ by a polynomial factor decreases γ' by a polynomial factor,[13] so $\delta + \gamma'$ in Equation (6) also decreases by a polynomial factor. Thus, set γ as a large enough polynomial in L such that

$$\frac{c}{\gamma} + 2\sqrt{\frac{4^m(\delta + \gamma')}{\epsilon}} \leq \frac{1}{2L}. \tag{8}$$

Since U rejects any correct query string (with certainty) in the NO case, and since $|\psi''\rangle$ is a valid history state whose Q register is a superposition over correct query strings (all of which must lead to reject), we conclude that $\langle\psi''|\,A\,|\psi''\rangle = 1$. Moreover, we have that

$$\left|\mathrm{Tr}(A\,|\psi\rangle\langle\psi|) - \mathrm{Tr}(A\,|\psi''\rangle\langle\psi''|)\right| \leq \|A\|_\infty \,\big\||\psi\rangle\langle\psi| - |\psi''\rangle\langle\psi''|\big\|_{\mathrm{tr}} \leq \frac{1}{2L},$$

where the first inequality follows from Hölder's inequality, and the second by Equations (6) and (8). We conclude that $\langle\psi|\,A\,|\psi\rangle \geq 1 - 1/(2L)$, completing the proof. $\qquad\square$

$\square$

4.2 $\mathrm{P}^{\mathrm{QMA[log]}}$-completeness of APX-2-CORR

We now define APX-2-CORR and show that it is $\mathrm{P}^{\mathrm{QMA[log]}}$-complete using similar techniques to Section 4.1. For brevity, define $f(|\psi\rangle, A, B) := \langle\psi|\,A \otimes B\,|\psi\rangle - \langle\psi|\,A\,|\psi\rangle\,\langle\psi|\,B\,|\psi\rangle$.

Definition 4.5 (APX-2-CORR$(H, A, B, k, l, a, b, \delta)$)**.** *Given a k-local Hamiltonian H, l-local observables A and B, and real numbers a, b, and δ such that $a - b \geq n^{-c}$ and $\delta \geq n^{-c'}$, for n the number of qubits H acts on and $c, c' \geq 0$ some constants, decide:*

- *If H has a ground state $|\psi\rangle$ satisfying $f(|\psi\rangle, A, B) \geq a$, output YES.*

- *If for any $|\psi\rangle$ satisfying $\langle\psi|\,H\,|\psi\rangle \leq \lambda(H) + \delta$ it holds that $f(|\psi\rangle, A, B) \leq b$, output NO.*

We now prove Theorem 1.2 by showing $\mathrm{P}^{\mathrm{QMA[log]}}$-hardness in Lemma 4.6 and containment in $\mathrm{P}^{\mathrm{QMA[log]}}$ in Lemma 4.7.

Lemma 4.6. APX-2-CORR *is $\mathrm{P}^{\mathrm{QMA[log]}}$-hard for $k = 5$ and $l = 1$, i.e., for 5-local Hamiltonian H and 1-local observables A and B.*

Proof. For an arbitrary $\mathrm{P}^{\mathrm{QMA[log]}}$ circuit U', define U as in the proof of Theorem 1.1, consisting of L one- and two-qubit gates. We modify U as follows. Let U's output qubit be denoted W_1. We add two ancilla qubits, W_2 and W_3, which are set to $|00\rangle$ throughout U's computation. We then append to U a sequence of six 2-qubit gates which, controlled on W_1, map $|00\rangle$ in $W_2 W_3$ to $|\phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$, e.g. apply a controlled Hadamard gate and the 5-gate Toffoli construction from Figure 4.8 of [50]. Appending six identity gates on W_1, we obtain a circuit $V = V_{L+12} \cdots V_1$ which has $L+12$ gates. Finally, we construct $H = H_1 + H_2$ as in the proof of Theorem 1.1, mapping V to a 5-local Hamiltonian H_1 on registers W, Q, and C, and we set $A = Z_{W_2}$ and $B = Z_{W_3}$ for

---

[13]In a previous version of this article, $\gamma' \in \Theta(\|H\|_\infty /\gamma)$. But this was circularly defined (recall $H = H_1 + H_2$), since $\|H_1\|_\infty$ scales with Δ, which in turn scales with γ. In the present version of the article, $\gamma' \in O\left(\eta/\gamma^2\right)$, where η depends only on $\|H_2\|_\infty$, and is thus independent of γ. Hence, increasing γ now correctly decreases γ'.

Pauli Z. Similar to the proof of Theorem 1.1, set $\Delta = 128L^3\eta\gamma$ and $\delta = 1/\Delta$, for γ large enough so that

$$\frac{c}{\gamma} + 2\sqrt{\frac{4^m(\delta + \gamma')}{\epsilon}} \leq \frac{1}{2(L+13)}, \tag{9}$$

for γ' as defined in the proof of Theorem 1.1. Set $a = 3/(L+13)$ and $b = 1/(L+13)$. This completes the construction.

To set up the correctness proof, consider history state $|\psi_{\text{hist}}\rangle$ for V given quantum proof $|\phi\rangle_{Q,Q'}$, and define for brevity $|\phi_t\rangle := V_t \cdots V_1 |\phi\rangle_{Q,Q'} |0 \cdots 0\rangle_W |00\rangle_{W_2 W_3}$. Then,

$$\langle\psi_{\text{hist}}| Z_{W_2} \otimes Z_{W_3} |\psi_{\text{hist}}\rangle = \frac{1}{L+13} \sum_{t=0}^{L+12} \text{Tr}\left((|\phi_t\rangle\langle\phi_t|_{Q,Q',W} \otimes |t\rangle\langle t|_C) Z_{W_2} \otimes Z_{W_3}\right), \tag{10}$$

since $Z_{W_2} \otimes Z_{W_3}$ acts invariantly on the clock register. Defining $|v\rangle := \sum_{t=L+1}^{L+12} |\phi_t\rangle_{Q,Q',W} |t\rangle_C$, we have that since $W_2 W_3$ is set to $|00\rangle$ for times $0 \leq t \leq L$, Equation (10) simplifies to

$$\frac{1}{L+13}\left(L + 1 + \langle v| Z_{W_2} \otimes Z_{W_3} |v\rangle\right).$$

Thus, via similar reasoning

$$f(|\psi_{\text{hist}}\rangle, Z_{W_2}, Z_{W_3}) = \frac{1}{L+13}\left[(L+1) + \langle v| Z_{W_2} \otimes Z_{W_3} |v\rangle\right] -$$
$$\frac{1}{(L+13)^2}\left[(L+1) + \langle v| Z_{W_2} |v\rangle\right]\left[(L+1) + \langle v| Z_{W_3} |v\rangle\right]. \tag{11}$$

Suppose now that Π is a YES instance. Then there exists a history state $|\psi_{\text{hist}}\rangle$ in the ground space of H (i.e. with quantum proof $|\phi\rangle_{Q,Q'} = |\hat{x}\rangle_Q \otimes |\phi'\rangle_{Q'}$ for a correct query string x) for which $W_2 W_3$ is set to $|\phi^+\rangle$ in the final seven timesteps (since U' is deterministic). Since $\langle\phi^+| Z \otimes Z |\phi^+\rangle = 1$ and $\langle\phi^+| Z \otimes I |\phi^+\rangle = 0$, via Equation (11) we have

$$f(|\psi_{\text{hist}}\rangle, Z_{W_2}, Z_{W_3}) \geq \frac{(L+1) - 5 + 7}{L+13} - \frac{((L+1)+5)^2}{(L+13)^2} = \frac{1}{L+13}\left(4 - \frac{49}{L+13}\right),$$

where the ± 5 terms correspond to timesteps $t = L+1, \ldots, L+5$ and use the fact that $\|Z\|_\infty = 1$.

Conversely, suppose Π is a NO instance, and consider any $|\psi\rangle$ satisfying $\langle\psi| H |\psi\rangle \leq \lambda(H) + \delta$. Then, as argued in the proof of Theorem 1.1, there exists a history state $|\psi''\rangle$ on "proof" $|\phi_1\rangle_{Q,Q'}$ (consisting of a superposition of correct query strings) satisfying

$$\left\| |\psi\rangle\langle\psi| - |\psi''\rangle\langle\psi''| \right\|_{\text{tr}} \leq \frac{1}{2(L+13)}, \tag{12}$$

by Equations (6), (8) and (9). Since the history state $|\psi''\rangle$ has $W_2 W_3$ set to $|00\rangle$ in all time steps, we know because $Z \otimes Z |00\rangle = |00\rangle$ that $f(|\psi''\rangle, Z_{W_2}, Z_{W_3}) = 0$. Thus, using Equation (11) and applying the Hölder inequality to the second term of $f(|\psi\rangle, Z_{W_2}, Z_{W_3})$ yields

$$f(|\psi\rangle, Z_{W_2}, Z_{W_3}) \leq 1 - \left(1 - \frac{1}{2(L+13)}\right)^2 = \frac{1}{L+13}\left(1 - \frac{1}{4(L+13)}\right).$$

$\square$

Lemma 4.7. APX-2-CORR *is in* $P^{QMA[\log]}$.

Proof. The proof combines ideas from Ambainis's original proof of APX-SIM $\in P^{QMA[\log]}$ [5] (see Theorem 6 therein) and a trick of Chailloux and Sattath [19] from the study of QMA(2). We give a proof sketch here. Specifically, let $\Pi = (H, A, B, k, l, a, b, \delta)$ be an instance of APX-2-CORR. Similar to [5], the $P^{QMA[\log]}$ verification procedure proceeds, at a high level, as follows:

1. Use logarithmically many QMA oracle queries to perform a binary search to obtain an estimate $\gamma \in \mathbb{R}$ satisfying $\lambda(H) \in [\gamma, \gamma + \frac{\delta}{2}]$.

2. Use a single QMA oracle query to verify the statement: "There exists $|\psi\rangle$ satisfying (1) $\langle\psi| H |\psi\rangle \leq \lambda(H) + \delta$ and (2) $f(|\psi\rangle, A, B) \geq a$".

The first of these steps is performed identically to the proof of APX-SIM $\in P^{QMA[\log]}$ [5]; we do not elaborate further here. The second step, however, differs from [5] for the following reason. Intuitively, [5] designs a QMA protocol which takes in many copies of a proof $|\psi\rangle$, performs phase estimation on each copy, postselects to "snap" each copy of $|\psi\rangle$ into a low-energy state $|\psi_i\rangle$ of H, and subsequently uses states $\{ \, |\psi_i\rangle \, \}$ to estimate the expectation against an observable A. If the ground space of H is degenerate, the states $\{ \, |\psi_i\rangle \, \}$ may not all be identical. This does not pose a problem in [5], as there soundness of the protocol is guaranteed since all low energy states have high expectation against A. In our setting, however, if we use this protocol to individually estimate each of the terms $\langle\psi| A \otimes B |\psi\rangle$, $\langle\psi| A |\psi\rangle$, and $\langle\psi| B |\psi\rangle$, soundness *can* be violated if each of these three terms are not estimated using the same state $|\psi_i\rangle$, since the promise gap of the input does not necessarily say anything about the values of each of these three terms individually.

To circumvent this, we observe that to evaluate $f(|\psi\rangle, A, B)$, we do not need the ground state $|\psi\rangle$ itself, but only a classical description of its local reduced density matrices (a similar idea was used in [19] to verify the energy of a claimed product state proof against a local Hamiltonian in the setting of QMA(2)). Specifically, suppose Π consists of a k-local Hamiltonian H acting on n qubits. Then, the prover sends classical descriptions of k-qubit density matrices $\{ \rho_S \}$ for each subset $S \subseteq [n]$ of size $|S| = k$, along with a QMA proof that the states $\{ \rho_S \}$ are consistent with a global n-qubit pure state $|\psi\rangle$ (recall the problem of verifying consistency is QMA-complete [47]). The verifier runs the QMA circuit for consistency, and assuming this check passes it uses the classical $\{ \rho_S \}$ to classically verify that (1) $\langle\psi| H |\psi\rangle \leq \lambda(H) + \delta$ and (2) $f(|\psi\rangle, A, B) \geq a$ (since both of these depend only on the local states $\{ \rho_S \}$). □

4.3 $P^{UQMA[\log]}$-hardness of SPECTRAL-GAP

We now restate and prove Theorem 1.3. We begin by defining SPECTRAL-GAP and UQMA.

Definition 4.8 (SPECTRAL-GAP(H, α) (Ambainis [5]))**.** *Given a Hamiltonian H and a real number $\alpha \geq n^{-c}$ for n the number of qubits H acts on and $c > 0$ some constant, decide:*

- *If $\lambda_2 - \lambda_1 \leq \alpha$, output YES.*

- *If $\lambda_2 - \lambda_1 \geq 2\alpha$, output NO.*

where λ_2 and λ_1 denote the second and first smallest eigenvalues of H, respectively.

For clarity, if the ground space of H is degenerate, then we define its spectral gap as 0.

Definition 4.9 (Unique QMA (UQMA) (Aharonov *et al.* [4])). *We say a promise problem $A = (A_{\text{yes}}, A_{\text{no}})$ is in Unique QMA if and only if there exist polynomials p, q and a polynomial-time uniform family of quantum circuits $\{Q_n\}$, where Q_n takes as input a string $x \in \Sigma^n$, a quantum proof $|y\rangle \in (\mathbb{C}^2)^{\otimes p(n)}$, and $q(n)$ ancilla qubits in state $|0\rangle^{\otimes q(n)}$, such that:*

- *(Completeness) If $x \in A_{\text{yes}}$, then there exists a proof $|y\rangle \in (\mathbb{C}^2)^{\otimes p(n)}$ such that Q_n accepts $(x, |y\rangle)$ with probability at least $2/3$, and for all $|\hat{y}\rangle \in (\mathbb{C}^2)^{\otimes p(n)}$ orthogonal to $|y\rangle$, Q_n accepts $(x, |\hat{y}\rangle)$ with probability at most $1/3$.*

- *(Soundness) If $x \in A_{\text{no}}$, then for all proofs $|y\rangle \in (\mathbb{C}^2)^{\otimes p(n)}$, Q_n accepts $(x, |y\rangle)$ with probability at most $1/3$.*

The main theorem of this section is the following.

Theorem 1.3. SPECTRAL-GAP *is* $\mathrm{P}^{\mathrm{UQMA[log]}}$*-hard for 4-local Hamiltonians H under polynomial-time Turing reductions (i.e. Cook reductions).*

We remark that Ambainis [5] showed that SPECTRAL-GAP $\in \mathrm{P}^{\mathrm{QMA[log]}}$, and gave a claimed proof that SPECTRAL-GAP is $\mathrm{P}^{\mathrm{UQMA[log]}}$-hard for $O(\log)$-local Hamiltonians under mapping reductions. ($\mathrm{P}^{\mathrm{UQMA[log]}}$ is defined as $\mathrm{P}^{\mathrm{QMA[log]}}$, except with a UQMA oracle in place of a QMA oracle.) As discussed in Section 1, however, Ambainis' proof of the latter result does not hold if the $\mathrm{P}^{\mathrm{UQMA[log]}}$ machine makes invalid queries (which in general is the case). Here, we build on Ambainis' approach [5] to show $\mathrm{P}^{\mathrm{UQMA[log]}}$-hardness of SPECTRAL-GAP under Turing reductions even when invalid queries are allowed, and we also improve the hardness to apply to $O(1)$-local Hamiltonians.

We begin by showing the following modified version of Lemma 3.3 tailored to UQMA (instead of QMA). It is crucial to observe that the mapping of Lemma 3.3 which produces Hamiltonian H is *efficient*, i.e. H can be computed in polynomial time. However, in Lemma 4.10 below, our mapping to a Hamiltonian H is *not* clearly efficient, meaning the lemma only shows the *existence* of H. Roughly, this is because the proof of Lemma 4.10 proceeds by replacing invalid queries with "dummy" NO queries to obtain the desired spectral gap. But a polynomial-time machine alone cannot identify such invalid queries, and hence cannot simulate this mapping.

Henceforth, we assume that all calls to the UQMA oracle Q are for an instance (H, a, b) of the Unique-Local Hamiltonian Problem (U-LH) (which is complete for UQMA [5]): Is the ground state energy of H at most ϵ with all other eigenvalues at least 3ϵ (YES case), or is the ground state energy at least 3ϵ (NO case), for $\epsilon \geq 1/\mathrm{poly}(n)$?

Lemma 4.10. *For any $x \in \{0, 1\}^m$, let $\hat{x}$ denote its unary encoding. Then, for any $\mathrm{P}^{\mathrm{UQMA[log]}}$ circuit U acting on n bits and making m queries to a UQMA oracle, there exists a 4-local Hamiltonian H acting on space $(\mathbb{C}^2)^{\otimes 2^m - 1} \otimes \mathcal{Y}$ such that there exists a correct query string $x = x_1 \cdots x_m$ such that:*

1. *The unique ground state of H lies in subspace $|\hat{x}\rangle\langle\hat{x}| \otimes \mathcal{Y}$.*

2. *The spectral gap of H is at least $(\epsilon - \delta)/4^m$ for inverse polynomial ϵ, δ with $\epsilon - \delta \geq 1/\mathrm{poly}(n)$.*

3. *For all strings $x' \in \{0, 1\}^m$, H acts invariantly on subspace $|\hat{x}'\rangle\langle\hat{x}'| \otimes \mathcal{Y}$.*

Proof. We begin by giving the construction of H.

Construction. As done in [5], we begin with $O(\log n)$-local Hamiltonian

$$H' = \sum_{i=1}^{m} \frac{1}{4^{i-1}} \sum_{y_1,\ldots,y_{i-1}} \bigotimes_{j=1}^{i-1} |y_j\rangle\langle y_j|_{\mathcal{X}_j} \otimes G'_{y_1\cdots y_{i-1}}, \tag{13}$$

where we define

$$G'_{y_1\cdots y_{i-1}} := |0\rangle\langle 0|_{\mathcal{X}_i} \otimes A_{\mathcal{Y}_i} + |1\rangle\langle 1|_{\mathcal{X}_i} \otimes H^{i,y_1\cdots y_{i-1}}_{\mathcal{Y}_i} \tag{14}$$

with A any fixed 2-local Hermitian operator with unique ground state of eigenvalue 2ϵ and spectral gap ϵ. Our approach is intuitively now as follows. We first run a *query validation* phase, in which we modify H' to obtain a new Hamiltonian H'' by replacing "sufficiently invalid" queries $H^{i,y_1\cdots y_{i-1}}_{\mathcal{Y}_i}$ with high-energy dummy queries. This creates the desired spectral gap. We then apply the technique of Lemma 3.3 to reduce the locality of H'', obtaining a 4-local Hamiltonian H, as desired. Note that our proof shows *existence* of H; unlike Lemma 3.3, however, it is not clear how to construct H in polynomial-time given H', as detecting invalid UQMA queries with a P machine seems difficult.

The query validation runs as follows. Fix any $1/\text{poly}(n) < \delta < \epsilon$ such that $\epsilon - \delta \geq 1/\text{poly}(n)$; for example, set $\delta = \epsilon/2$. For any $G'_{y_1\cdots y_{i-1}}$ whose spectral gap is at most $\epsilon - \delta$, replace it with

$$G_{y_1\cdots y_{i-1}} := |0\rangle\langle 0|_{\mathcal{X}_i} \otimes A_{\mathcal{Y}_i} + |1\rangle\langle 1|_{\mathcal{X}_i} \otimes 3\epsilon I \tag{15}$$

in H', denoting the new Hamiltonian as H''. Two remarks: By Case 2 of Lemma 4.11 below, the validation phase does not catch *all* invalid queries. Second, setting $1/\text{poly}(n) < \delta$ and $\epsilon - \delta \geq 1/\text{poly}(n)$ (as opposed to, say, $1/\exp(n)$) is required for our proof of Theorem 1.3 later.

Correctness. Intuitively, the query validation phase roughly replaces each "sufficiently invalid" query i with a valid NO query. This is formalized by Lemma 4.11, shown subsequently, which should be roughly interpreted as saying a "sufficiently invalid" query i is one for which the spectral gap of $G'_{y_1\cdots y_{i-1}}$ is at most $\epsilon - \delta$. For any such replaced query i, henceforth in this proof, a query string which answers YES (i.e. $|1\rangle$) to query i is considered incorrect with respect to H'' (since i is now a valid NO query in H''). Crucially, if a string x is a correct query string for H'', then it is also a correct query string for H'. The converse is false; nevertheless, H'' has at least one correct query string (since any invalid query would have allowed both $|0\rangle$ and $|1\rangle$ as answers), which suffices for our purposes.

As in the proof of Lemma 3.1, observe that H'' is block-diagonal with respect to register $\bigotimes_{i=1}^{m} \mathcal{X}_i$. Let $x \in \{0,1\}^m$ denote a correct query string which has minimal energy among all *correct* query strings against H'', and for any $y \in \{0,1\}^m$, define λ_y as the smallest eigenvalue in block $\mathcal{H}_y$. A similar analysis to that of Lemma 3.1 shows that for any incorrect query string y, $\lambda_y \geq \lambda_x + \epsilon/4^m$. This is because replacing the term $2\epsilon I$ in $M_{y_1\cdots y_{i-1}}$ from Lemma 3.1 with A in $G_{y_1\cdots y_{i-1}}$ here preserves the property that answering NO on query i yields minimum energy 2ϵ.

We now argue that x is in fact *unique*, and all other eigenvalues of H'' corresponding to correct query strings have energy at least $\lambda_x + (\epsilon - \delta)/4^m$. There are two cases to consider: eigenvalues arising from different query strings and eigenvalues arising from the same query string.

Case 1: Eigenvalues from different query strings. Let $y = y_1 \cdots y_m \neq x$ be a correct query string for H''. Since both x and y are correct strings, there must exist an invalid query i where $x_i \neq y_i$.

First consider the case where $G'_{y_1 \cdots y_{i-1}}$ has spectral gap at most $\epsilon - \delta$. Then, after the validation phase, query i is replaced with a valid NO query $G_{y_1 \cdots y_{i-1}}$. Thus, whichever of x or y has a 1 as bit i is an incorrect string for H'', and from our previous analysis has energy at least $\lambda_x + \epsilon/4^m$ against H''. (This, in particular, implies $x_i = 0$ and $y_i = 1$, by the minimality of x.)

Alternatively, suppose $G'_{y_1 \cdots y_{i-1}} = G_{y_1 \cdots y_{i-1}}$ has spectral gap strictly larger than $\epsilon - \delta$. By Lemma 4.11, the invalid query i must be a "YES instance violating the uniqueness promise" with corresponding query Hamiltonian $H_{\mathcal{Y}_i}^{i,y_1 \cdots y_{i-1}}$ satisfying $\lambda(H_{\mathcal{Y}_i}^{i,y_1 \cdots y_{i-1}}) \leq \epsilon$. Thus, the 0-block of Equation (14) has minimum eigenvalue 2ϵ by construction, whereas the 1-block of Equation (14) has minimum eigenvalue at most ϵ. It follows from our previous analysis that whichever of x or y has a 0 as bit i has energy at least $\lambda_x + \epsilon/4^m$ against H''. (This, in particular, implies $x_i = 1$ and $y_i = 0$, by the minimality of x.)

Case 2: Eigenvalues from the same query string. Let us next consider eigenvalues in the minimal block $\mathcal{H}_x$ with correct query string x. In this block, H'' is equivalent to operator

$$M := \sum_{i=1}^{m} \frac{1}{4^{i-1}} B_{x_1 \cdots x_{i-1}},$$

where $B_{x_1 \cdots x_{i-1}} = A$ if $x_i = 0$ and $B_{x_1 \cdots x_{i-1}}$ can equal either $H_{\mathcal{Y}_i}^{i,y_1 \cdots y_{i-1}}$ or $3\epsilon I$ (depending on how the validation phase proceeded) if $x_i = 1$. To show our claim, it suffices to show that M has spectral gap $(\epsilon - \delta)/4^m$.

Recall now that each $B_{x_1 \cdots x_{i-1}}$ acts non-trivially only on space $\mathcal{Y}_i$. Therefore, the minimum eigenvalue of M is obtained by setting each space $\mathcal{Y}_i$ to the ground state vector $|\psi_i\rangle$ of $B_{x_1 \cdots x_{i-1}}$. If we can now show that each $B_{x_1 \cdots x_{i-1}}$ has a spectral gap of $\epsilon - \delta$, it will hence follow that swapping $|\psi_i\rangle$ for any excited state $|\psi_i^{\perp}\rangle$ orthogonal to $|\psi_i\rangle$ on $\mathcal{Y}_i$ yields an additive increase in energy of at least $\epsilon - \delta$ against any $B_{x_1 \cdots x_{i-1}}$. It would then follow that the spectral gap of M is at least $(\epsilon - \delta)/4^m$, as desired.

So let us argue that each $B_{x_1 \cdots x_{i-1}}$ has spectral gap at least $\epsilon - \delta$. There are three cases to consider:

- If $B_{x_1 \cdots x_{i-1}} = A$, it has spectral gap ϵ by definition of A.

- If $B_{x_1 \cdots x_{i-1}} = H_{\mathcal{Y}_i}^{i,y_1 \cdots y_{i-1}}$, then consider first the case where query i was valid. Since $B_{x_1 \cdots x_{i-1}} = H_{\mathcal{Y}_i}^{i,y_1 \cdots y_{i-1}}$ only if $x_i = 1$, it follows that the correct answer to query i is 1. Thus, $H_{\mathcal{Y}_i}^{i,y_1 \cdots y_{i-1}}$ is a valid YES instance of U-LH, and so has a spectral gap of at least 2ϵ by definition.

  If instead query i was invalid, then since $B_{x_1 \cdots x_{i-1}} = H_{\mathcal{Y}_i}^{i,y_1 \cdots y_{i-1}}$ and not $B_{x_1 \cdots x_{i-1}} = 3\epsilon I$, it must be that the query validation phase did not catch this invalid query, implying the spectral gap of $G'_{y_1 \cdots y_{i-1}}$ was strictly larger than $\epsilon - \delta$. But then Lemma 4.11 implies that the spectral gap of $H_{\mathcal{Y}_i}^{i,y_1 \cdots y_{i-1}}$ must be at least $\epsilon - \delta$, as desired.

- Suppose $B_{x_1 \cdots x_{i-1}} = 3\epsilon I$; this happens only when the query validation phase replaced an invalid query i with a valid NO query $3\epsilon I$. But since it also holds that $B_{x_1 \cdots x_{i-1}} = 3\epsilon I$ only if $x_i = 1$, this implies that x is not a correct query string with respect to H'' (since it answered 1 on query i instead of 0), which is a contradiction.

This completes the correctness proof.

Getting H'' down to 4-local H. Finally, the approach of Lemma 3.3 allows us to convert $O(\log n)$-local H'' to 4-local H. $\qquad\square$

The following lemma is used in the proof of Lemma 4.10 above, and assumes the notation introduced therein.

Lemma 4.11. *In Equation (14), suppose $H_{\mathcal{Y}_i}^{i,y_1\cdots y_{i-1}}$ is invalid. Precisely one of the following cases holds.*

1. *(Neither YES or NO case) If there exists $0 < \delta \le \epsilon$ such that $\lambda(H_{\mathcal{Y}_i}^{i,y_1\cdots y_{i-1}}) = \epsilon + \delta$ or $\lambda(H_{\mathcal{Y}_i}^{i,y_1\cdots y_{i-1}}) = 3\epsilon - \delta$, then the spectral gap of $G'_{y_1\cdots y_{i-1}}$ is at most $\epsilon - \delta$.*

2. *(YES case violating uniqueness promise) If $\lambda(H_{\mathcal{Y}_i}^{i,y_1\cdots y_{i-1}}) \le \epsilon$ and $\lambda_2(H_{\mathcal{Y}_i}^{i,y_1\cdots y_{i-1}}) < 3\epsilon$, then for any $0 < \delta < \epsilon$, either the spectral gap of $G'_{y_1\cdots y_{i-1}}$ is at most $\epsilon - \delta$, or the spectral gap of $H_{\mathcal{Y}_i}^{i,y_1\cdots y_{i-1}}$ is least $\epsilon - \delta$.*

Proof. Observe that $G'_{y_1\cdots y_{i-1}}$ is block-diagonal with respect to register $\mathcal{X}_i$; we will refer to each of these blocks as the 0- and 1-block, respectively. For clarity, when we refer to $\lambda_2(X)$ for an operator X below, if the ground space of X is degenerate then we define $\lambda_2(X) = \lambda(X)$.

Case 1. Suppose first that $\lambda(H_{\mathcal{Y}_i}^{i,y_1\cdots y_{i-1}}) = \epsilon + \delta$. Then, since the smallest eigenvalue in the 0-block is 2ϵ, we have that $\lambda(G'_{y_1\cdots y_{i-1}}) = \lambda(H_{\mathcal{Y}_i}^{i,y_1\cdots y_{i-1}})$, and so the spectral gap of $G'_{y_1\cdots y_{i-1}}$ equals

$$\min(2\epsilon, \lambda_2(H_{\mathcal{Y}_i}^{i,y_1\cdots y_{i-1}})) - (\epsilon + \delta) \le \epsilon - \delta.$$

Suppose next that $\lambda(H_{\mathcal{Y}_i}^{i,y_1\cdots y_{i-1}}) = 3\epsilon - \delta$. Then, $\lambda(G'_{y_1\cdots y_{i-1}}) = 2\epsilon$ by the 0-block. Since the spectral gap of the 0-block is ϵ by construction, we thus have that $\lambda_2(G'_{y_1\cdots y_{i-1}}) = \lambda(H_{\mathcal{Y}_i}^{i,y_1\cdots y_{i-1}})$, and so the spectral gap of $G'_{y_1\cdots y_{i-1}}$ equals $(3\epsilon - \delta) - 2\epsilon = \epsilon - \delta$.

Case 2. Since the smallest eigenvalue of the 0-block is 2ϵ, we have $\lambda(G'_{y_1\cdots y_{i-1}}) = \lambda(H_{\mathcal{Y}_i}^{i,y_1\cdots y_{i-1}})$. Therefore, the spectral gap of $G'_{y_1\cdots y_{i-1}}$ equals

$$\min(2\epsilon, \lambda_2(H_{\mathcal{Y}_i}^{i,y_1\cdots y_{i-1}})) - \lambda(H_{\mathcal{Y}_i}^{i,y_1\cdots y_{i-1}}). \tag{16}$$

For any δ as defined in the claim, suppose now that the spectral gap of $G'_{y_1\cdots y_{i-1}}$ is at least $\epsilon - \delta$. Then if in Equation (16) $\min(2\epsilon, \lambda_2(H_{\mathcal{Y}_i}^{i,y_1\cdots y_{i-1}})) = 2\epsilon$, since $\lambda(H_{\mathcal{Y}_i}^{i,y_1\cdots y_{i-1}}) \le \epsilon$ by assumption, we conclude the spectral gap of $H_{\mathcal{Y}_i}^{i,y_1\cdots y_{i-1}}$ is at least ϵ. Otherwise, if $\min(2\epsilon, \lambda_2(H_{\mathcal{Y}_i}^{i,y_1\cdots y_{i-1}})) = \lambda_2(H_{\mathcal{Y}_i}^{i,y_1\cdots y_{i-1}})$, then the smallest two eigenvalues of $H_{\mathcal{Y}_i}^{i,y_1\cdots y_{i-1}}$ and $G'_{y_1\cdots y_{i-1}}$ coincide, and so the spectral gap of the former is at least $\epsilon - \delta$. Thus, we conclude that either the spectral gap of $G'_{y_1\cdots y_{i-1}}$ is at most $\epsilon - \delta$, or the spectral gap of $H_{\mathcal{Y}_i}^{i,y_1\cdots y_{i-1}}$ is least $\epsilon - \delta$. $\qquad\square$

We now prove the main theorem of this section.

Proof of Theorem 1.3. As done in [5], we start with the Hamiltonian H' from Equation (13). In [5], it was shown (Section A.3, Claim 2) that if all query Hamiltonians $H_{\mathcal{Y}_i}^{i,y_1\cdots y_{i-1}}$ correspond to valid UQMA queries, H' has a unique ground state and spectral gap at least $\epsilon/4^m$. When invalid queries are allowed, however, the spectral gap of H' can vanish, invalidating the $\text{P}^{\text{UQMA[log]}}$-hardness proof of [5]. Thus, we require a technique for identifying invalid queries and "removing them" from H'. Unfortunately, it is not clear how a P machine alone can achieve such a "property testing" task of checking if a query is sufficiently invalid. However, the key observation is that an oracle Q for SPECTRAL GAP can help.

We proceed as follows. Given an arbitrary $\text{P}^{\text{UQMA[log]}}$ circuit U acting on n bits, construct $O(\log n)$-local H' from Equation (13). For each term $G'_{y_1\cdots y_{i-1}}$ appearing in H', perform binary search using $O(\log n)$ queries to Q to obtain an estimate Δ for the spectral gap of $G'_{y_1\cdots y_{i-1}}$ to within sufficiently small but fixed additive error $\delta \in 1/\text{poly}(n)$. (A similar procedure involving a QMA oracle is used in Ambainis' proof of containment of $\text{APX-SIM} \in \text{P}^{\text{QMA[log]}}$ to estimate the smallest eigenvalue of a local Hamiltonian; we hence omit further details here.) As done in the proof of Lemma 4.10, if $\Delta \leq \epsilon - \delta$, we conclude $H_{\mathcal{Y}_i}^{i,y_1\cdots y_{i-1}}$ is "sufficiently invalid", and replace $G'_{y_1\cdots y_{i-1}}$ with $G_{y_1\cdots y_{i-1}}$ from Equation (15). Following the construction of Lemma 4.10, we hence can map H' to a 4-local Hamiltonian H such that H has a unique ground state and spectral gap $(\epsilon - \delta)/4^m$, and the ground state of H corresponds to a correct query string for H'. Note that implementing the mapping from H' to H requires polynomially many queries to the oracle, hence yielding a polynomial-time *Turing* reduction.

Next, following [5], let $T := \sum_{y_1\ldots y_m} \bigotimes_{i=1}^m |y_i\rangle\langle y_i| \in \text{L}(\mathcal{Y})$, where we sum over all query strings $y_1\ldots y_m$ which cause U to output 0. Unlike [5], as done in Lemma 3.3, we apply Kitaev's unary encoding trick [45] and implicitly encode the query strings in T in unary. (We remark the term H_{stab} contained in H will enforce the correct unary encoding in register $\mathcal{X}$). Finally, introduce a single-qubit register $\mathcal{B}$, and define

$$H_{\text{final}} := I_B \otimes H_{\mathcal{X},\mathcal{Y}} + 4\epsilon |0\rangle\langle 0|_B \otimes T_{\mathcal{X}} \otimes I_{\mathcal{Y}}.$$

The claim now follows via an analysis similar to [5]. Let $|\psi\rangle_{\mathcal{X},\mathcal{Y}}$ denote the unique ground state of H, whose $\mathcal{X}$ register contains the (unary encoding of) a correct query string for U. If U accepts, then $|i\rangle_{\mathcal{B}} \otimes |\psi\rangle_{\mathcal{X},\mathcal{Y}}$ for $i \in \{0,1\}$ are degenerate ground states of H_{final}, implying H_{final} has no spectral gap. Conversely, if U rejects, observe that the smallest eigenvalue of H_{final} lies in the $|1\rangle_{\mathcal{B}}$ block of H_{final}. This is because H_{final} is block-diagonal with respect to register $\mathcal{X}$, and we have from the proof of Lemma 3.3 that $\lambda(H) < 3\epsilon$. Restricted to this $|1\rangle_{\mathcal{B}}$ block, the spectral gap of H_{final} is at least $(\epsilon - \delta)/4^m$ by Lemma 4.10. Alternatively, restricted to the $|0\rangle_{\mathcal{B}}$ block, any correct query string in $\mathcal{X}$ leads to spectral gap at least 4ϵ (by construction of T, since U outputs 0 in this case), and any incorrect query string in $\mathcal{X}$ leads to spectral gap at least $(\epsilon - \delta)/4^m$ by Lemma 4.10. Hence, H_{final} has an inverse polynomial spectral gap, as desired. $\square$

5 Upper bounds (containment in PP)

We now restate and prove Theorem 1.4. Our approach is to use the strong error reduction technique of Marriott and Watrous [49] to develop a variant of the hierarchical voting scheme used in the proof of $\text{P}^{\text{NP[log]}} \subseteq \text{PP}$ [9]. We also require a more involved analysis than present in [9], since QMA is a class of promise problems, not decision problems.

Theorem 1.4. $\text{P}^{\text{QMA}[\log]} \subseteq \text{PP}$.

Proof. Let Π be a P machine which makes $m = c \log n$ queries to an oracle for 2-LH, for $c \in O(1)$ and n the input size. Without loss of generality, we assume all queries involve Hamiltonians on M qubits (M some fixed polynomial in n). Define $q := (M + 2)m$. We give a PQP computation simulating Π (i.e. the PQP computation, which does not have access to a 2-LH oracle, will accept if and only if Π is a YES instance). Since PQP = PP [62], this yields the claim. Let $|y|$ denote the non-negative integer with binary encoding y, and let V denote the verification circuit for 2-LH. The PQP computation is (intuition to follow):

1. For i from 1 to m:

    (a) Prepare $\rho = I/2^M \in \text{D}\left((\mathbb{C}^2)^{\otimes M}\right)$.

    (b) Run V on the i-th query Hamiltonian $H^{i,y_1\cdots y_{i-1}}_{y_i}$ (see Equation (3)) and proof ρ, and measure the output qubit in the standard basis. Set bit y_i to the result.

2. Let $y = y_1 \cdots y_m$ be the concatenation of bits set in Step 1(b).

3. For i from 1 to $n^c - 1$:

    (a) If $|y| < i$, then with probability $1 - 2^{-q}$, set $y = \#$, and with probability 2^{-q}, leave y unchanged.

4. If $y = \#$, output a bit uniformly at random. Else, run Π on query string y and output Π's answer.

Intuition. In Step 1, one tries to determine the correct answer to query i by guessing a satisfying quantum proof for verifier V. Suppose for the moment that V has zero error, i.e. has completeness 1 and soundness 0, and that Π only makes valid queries. Then, if Step 1(b) returns $y_i = 1$, one knows with certainty that the query answer should be 1. And, if the correct answer to query i is 0, then Step 1(b) returns $y_i = 0$ with certainty. Thus, analogous to the classical case of an NP oracle (as done in [9]), it follows that the lexicographically *largest* query string y^* obtainable by this procedure must be the (unique) correct query string (note that $y^* \neq 1^m$ necessarily[14]). Thus, ideally one wishes to obtain y^*, simulate Π on y^*, and output the result. To this end, Step 3 ensures that among all values of $y \neq \#$, y^* is more likely to occur than all other $y \neq y^*$ combined. We now make this intuition rigorous (including in particular the general case where V is not zero-error and Π makes invalid queries).

Correctness. To analyze correctness of our PQP computation, it will be helpful to refine our partition of the set of query strings $\{0, 1\}^m$ into three sets:

- **(Correct query strings)** Let $A \subseteq \{0, 1\}^m$ denote the set of query strings which correspond to correctly answering each of the m queries. Note we may have $|A| > 1$ if invalid queries are made.

---

[14]Under the assumptions that V has zero error and Π makes only valid queries, $y^* = 1^m$ can only be obtained by this procedure if all queries are for YES instances of 2-LH. If, on the other hand, query i is a NO query, then a correct proof cannot be guessed (since it does not exist), and so $y_i^* = 0$ necessarily.

- **(Incorrect query strings)** Let $B \subseteq \{0,1\}^m \setminus A$ denote the set of query strings such that for any $y \in B$, all bits of y which encode an incorrect query answer are set to 0 (whereas the correct query answer would have been 1, i.e. we failed to "guess" a good proof for this query in Step 1).

- **(Strongly incorrect query strings)** Let $C = \{0,1\}^m \setminus (A \cup B)$ denote the set of query strings such that for any $y \in C$, at least one bit corresponding to an incorrect query answer is set to 1 (whereas the correct query answer would have been 0). Such an error can only arise due to the bounded-error of our QMA verifier in Step 1(b).

Let Y be a random variable corresponding to the query string y obtained at the end of Step 3. To show correctness, we claim that it suffices to show that $\Delta := \Pr[Y \in A] - \Pr[Y \in B \cup C] > 0$. To see this, let p_1, p_2, and p_3 denote the probability that after Step 3, $y = \#$, $y \in A$, and $y \in B \cup C$, respectively. Then, $p_1 + p_2 + p_3 = 1$, and let $p_2 - p_3 = \Delta > 0$. Suppose now that the input to Π is a YES instance. Then, our protocol outputs 1 with probability at least $\frac{p_1}{2} + p_2 = \frac{1 - p_2 - p_3}{2} + p_2 = \frac{1 + \Delta}{2} > \frac{1}{2}$. If the input is a NO instance, the protocol outputs 1 with probability $\leq \frac{p_1}{2} + p_3 = \frac{1 - \Delta}{2} < \frac{1}{2}$. We hence have a PQP computation, as desired. We thus now show that $\Delta > 0$.

To ease the presentation, we begin by making two assumptions (to be removed later): (i) V is zero-error and (ii) Π makes only valid queries. In this case, assumption (i) implies $C = \emptyset$ (i.e. all incorrect query strings belong to B), and (ii) implies A is a singleton (i.e. there is a unique correct query string y^*). Thus, here $\Delta = \Pr[Y \in A] - \Pr[Y \in B]$.

To begin, note that for any $y \in \{0,1\}^m$, we have

$$\Pr[Y = y] = \Pr[y \text{ chosen in Step 2 }] \cdot \left(\frac{1}{2^q}\right)^{(n^c - 1) - |y|}. \tag{17}$$

Let $\text{HW}(x)$ denote the Hamming weight of $x \in \{0,1\}^m$. Since each query corresponds to a verifier on M proof qubits, we have for (the unique) $y^* \in A$ that

$$\Pr[y^* \text{ chosen in Step 2 }] \geq 2^{-M \cdot \text{HW}(y^*)} \geq 2^{-Mm} \tag{18}$$

(recall from Section 2 that setting $\rho = I/2^M$ simulates "guessing" a correct proof with probability at least $1/2^M$). It follows by Equations (17) and (18) that

$$
\begin{aligned}
\Delta &\geq \left(\frac{1}{2^q}\right)^{(n^c - 1) - |y^*|} \left[\frac{1}{2^{Mm}} - \sum_{y \in B} \Pr[y \text{ chosen in Step 2 }] \cdot \left(\frac{1}{2^q}\right)^{|y^*| - |y|} \right] \\
&\geq \left(\frac{1}{2^q}\right)^{(n^c - 1) - |y^*|} \left[\frac{1}{2^{Mm}} - (2^m) \left(\frac{1}{2^q}\right) \right] \geq \left(\frac{1}{2^q}\right)^{(n^c - 1)} \frac{1}{2^{Mm}} \left[1 - \frac{1}{2^m} \right],
\end{aligned} \tag{19}
$$

where the second inequality follows from the trivial bound $\Pr[y \text{ chosen in Step 2 }] \leq 1$ and since $y \in B$ if and only if $|y| < |y^*|$, and the third since $q = (M + 2)m$. Thus, $\Delta > 0$ as desired.

Removing assumption (i). We now remove the assumption that V is zero error. In this case, A is still a singleton; let $y^* \in A$. We can now also have strongly incorrect query strings, i.e. $C \neq \emptyset$ necessarily. Assume without loss of generality that V acts on M proof qubits, and by strong error

reduction [49] has completeness $c := 1 - 2^{-p(n)}$ and soundness $s := 2^{-p(n)}$, for p a polynomial to be chosen as needed. Then, since V can err, Equation (18) becomes

$$
\begin{aligned}
\Pr[y^* \text{ chosen in Step 2 }] &\geq \left(\frac{c}{2M}\right)^{\text{HW}(y^*)} (1-s)^{m-\text{HW}(y^*)} = \frac{1}{2M}^{\text{HW}(y^*)} e^{m \ln(1 - \frac{1}{2^p})} \\
&\geq \frac{1}{2^{Mm}}\left(1 - \frac{m}{2^p - 1}\right),
\end{aligned}
\tag{20}
$$

where the equality follows by the definitions of c and s, and the second inequality by applying the Maclaurin series expansion $\ln(1+x) = \sum_{n=1}^{\infty}(-1)^{n+1}\frac{x^n}{n}$ for $|x| < 1$ and the fact that $e^t \geq 1+t$ for all $t \in \mathbb{R}$. Thus, the analysis of Equation (19) yields that

$$
\Pr[Y \in A] - \Pr[Y \in B] \geq \left(\frac{1}{2^q}\right)^{(n^c - 1)} \frac{1}{2^{Mm}}\left[1 - \frac{1}{2^m} - \frac{m}{2^p - 1}\right],
\tag{21}
$$

i.e. the additive error introduced when assumption (i) is dropped scales as $\approx 2^{-p}$. Crucially, Equation (21) holds for all $y \in B$ even with assumption (i) dropped since the analysis of Equation (19) used only the trivial bound $\Pr[y \text{ chosen in Step 2 }] \leq 1$ for any $y \in B$.

Next, we upper bound the probability of obtaining $y \in C$ in Step 2. Next, we upper bound the probability of obtaining $y \in C$ in Step 2.

Lemma 5.1. $\Pr(Y \in C) \leq \frac{2^m}{2^p}$.

Proof. Any $y \in C$ must have a bit j incorrectly set to 1, whereas the correct query answer (given bits 1 through $j-1$ of y) should have been 0. The probability of this occurring for bit j in Step 1(b) is at most 2^{-p}, by the soundness property of V. Since $|C| \leq 2^m$, the claim follows. $\square$

Lemma 5.1 now implies

$$
\Delta \geq \left(\frac{1}{2^q}\right)^{(n^c - 1)} \frac{1}{2^{Mm}}\left[1 - \frac{1}{2^m} - \frac{m}{2^p - 1}\right] - \frac{2^m}{2^p}.
\tag{22}
$$

We conclude that setting p to a sufficiently large fixed polynomial ensures $\Delta > 0$, as desired.

Removing assumption (ii). We now remove the assumption that Π only makes valid queries, which is the most involved step. Here, A is no longer necessarily a singleton. The naive approach would be to let y^* denote the *lexicographically largest* string in A, and attempt to run a similar analysis as before. Unfortunately, this no longer necessarily works for the following reason. For any invalid query i, we do not have strong bounds on the probability that V accepts in Step 1(b); in principle, this value can lie in the range $(2^{-p}, 1 - 2^{-p})$. Thus, running the previous analysis with the lexicographically largest $y^* \in A$ may cause Equation (22) to yield a negative quantity. We hence require a more delicate analysis.

We begin by showing the following lower bound.

Lemma 5.2. *Define* $\Delta' := \Pr[Y \in A] - \Pr[Y \in B]$. *Then,*

$$
\Delta' \geq \left(\frac{1}{2^q}\right)^{(n^c - 1)} \frac{1}{2^{Mm}}\left[1 - \frac{1}{2^m} - \frac{m}{2^p - 1}\right].
$$

31

Proof. For any string $y \in \{0,1\}^m$, let $I_y \subseteq \{1,\ldots,m\}$ denote the indices of all bits of y set by invalid queries. We call each such $i \in I_y$ a *divergence point*. (This name is chosen since on any invalid query, the computation can diverge into either of two correct computation paths.) Let $p_{y,i}$ denote the probability that (invalid) query i (defined given answers to queries 1 through $i-1$) outputs bit y_i, i.e. $p_{y,i}$ denotes the probability that at divergence point i, we go in the direction of bit y_i. We define the *divergence probability* of $y \in \{0,1\}^m$ as $p_y = \Pi_{i \in I_y} p_{y,i}$, i.e. p_y is the probability of answering all invalid queries as y did.

The proof now proceeds by giving for each $1 \le i \le |A|$, which denotes the *iteration number*, a 3-tuple $(y_{i-1}^*, y_i^*, B_{y_i^*}) \in \{0,1\}^m \times \{0,1\}^m \times \mathcal{P}(B)$, where $\mathcal{P}(X)$ denotes the power set of set X. Set

$$\Delta_i' := \Pr[Y \in \{y_1^*,\ldots,y_i^*\}] - \Pr[Y \in B_{y_1^*} \cup \cdots \cup B_{y_i^*}],$$

where it will be the case that $\{B_{y_i^*}\}_{i=1}^{|A|}$ is a partition of B. Thus, we have $\Delta' \ge \Delta_{|A|}'$, implying that a lower bound on $\Delta_{|A|}'$ suffices to prove our claim. We hence prove via induction on the iteration number i that for all $1 \le i \le |A|$,

$$\Delta_i' \ge \left(\frac{1}{2^q}\right)^{(n^c-1)} \frac{1}{2^{Mm}} \left[1 - \frac{1}{2^m} - \frac{m}{2^p - 1}\right].$$

Base case (i=1). In this case y_0^* is undefined. Set y_1^* to any string in A with divergence probability at least

$$p_1^* = \prod_{i \in I_{y_1^*}} p_{y_1^*,i} \ge 2^{-\left|I_{y_1^*}\right|}. \tag{23}$$

Such a string must exist, since at each divergence point i, at least one of the outcomes in $\{0,1\}$ occurs with probability at least $1/2$. (Note that queries are not being made to a QMA oracle in hierarchical voting (since our PQP machine does not have access to a QMA oracle), but to a QMA verifier V with a maximally mixed proof as in Step 1(a). Whereas in the former case the output of the oracle on an invalid query does not have to consistently output a value with any particular probability, in the latter case, there is some fixed probability p with which V outputs 1 each time it is run on a fixed proof.) Finally, define $B_{y_1^*} := \{y \in B \mid |y| < |y_1^*|\}$ (recall $|y|$ is the non-negative integer with binary encoding y).

Let k_* denote the number of divergence points of y_1^* (i.e. $k_* = \left|I_{y_1^*}\right|$), and k_0 (k_1) the number of zeros (ones) of y_1^* arising from valid queries. Thus, $k_* + k_0 + k_1 = m$. Then, Equation (20) becomes

$$\Pr[y_1^* \text{ in Step 2}] \ge \left(\frac{c}{2^M}\right)^{k_1} (1-s)^{k_0} p_1^* \ge \left(\frac{1}{2^M}\right)^{k_1} \left(\frac{1}{2}\right)^{k_*} \left(1 - \frac{m - k_*}{2^p - 1}\right)$$
$$\ge \frac{1}{2^{Mm}} \left(1 - \frac{m}{2^p - 1}\right), \tag{24}$$

where the second inequality follows from Equation (23), and the third since $k_* \ge 0$ and $k_1 + k_* \le m$. Thus, Δ_1' is lower bounded by the expression in Equation (21) via an analogous analysis for y_1^* and $B_{y_1^*}$.

Inductive step. Assume the claim holds for $1 \le i - 1 < |A|$. We show it holds for i. Let y_{i-1}^* be the choice of y^* in the previous iteration $i-1$. Define $A_{y_i^*} := \{y \in A \mid |y| > |y_{i-1}^*|\}$. Partition

$A_{y_i^*}$ into sets S_k for $k \in [m]$, such that S_k is the subset of strings in $A_{y_i^*}$ which agrees with y_{i-1}^* on the first $k-1$ bits, but disagrees on bit k. Note that if $S_k \neq \emptyset$, then bit k of y_{i-1}^* is 0 and bit k of any string in S_k is 1. For each $S_k \neq \emptyset$, choose arbitrary representative $z_k \in S_k$, and define *bounded divergence probability*

$$q_i(k) := \prod_{t \in I_{z_k}^{\leq k}} p_{z_k,t}$$

where $I_{z_k}^{\leq k} := \{\, t \in I_{z_k} \mid t \leq k \,\}$. Note that $q_i(k) > 0$ (since $S_k \neq \emptyset$). Else if $S_k = \emptyset$, set $q_i(k) = 0$. Let q_i^* be the max such bounded divergence probability:

$$q_i^* = \max_{k \in [m]} q_i(k) \qquad \text{and} \qquad k_i^* = \arg\max_{k \in [m]} q_i(k). \tag{25}$$

Let y_i^* be the lexicographically largest query string in $S_{k_i^*}$ with divergence probability p_i^* s.t.:

$$p_i^* \geq q_i^* \cdot 2^{-\left|I_{y_i^*}\right| + \left|I_{y_i^*}^{\leq k_i^*}\right|}. \tag{26}$$

That such a $y_i^* \in S_{k_i^*}$ exists follows from an argument similar to Equation (23): By definition, q_i^* denotes the bounded divergence probability for all invalid queries up to and including query k_i^*, and the term exponential in $\left(-\left|I_{y_i^*}\right| + \left|I_{y_i^*}^{\leq k_i^*}\right|\right)$ is obtained by greedily choosing, for all invalid queries of y_i^* *after* query k_i^*, the outcome which occurs with probability at least $1/2$. Set $B_{y_i^*} := \{\, y \in B \mid |y_{i-1}^*| < |y| < |y_i^*| \,\}$. The following lemma will be useful.

Lemma 5.3. *For any $y \in B_{y_i^*}$, $\Pr[y$ chosen in Step 2$] \leq q_i^*$, where recall q_i^* is the probability from Equation (25).*

Proof. Fix any $y \in B_{y_i^*}$. Since $|y| > |y_{i-1}^*|$, there must be an index k such that the k-th bit of y is 1 and that of y_{i-1}^* is 0. Let k denote the first such index. Since $y \notin C$ (because $B_{y_i^*} \cap C = \emptyset$), it must be that query k (defined given bits $y_1 \cdots y_{k-1}$) is invalid. Thus, bit k is a divergence point of y_{i-1}^*, and there exists a correct query string $y' \in S_k$. By Equation (25), q_i^* was chosen as the maximum over all bounded diverge probabilities. Thus, $q_i^* \geq q_i(k)$, where recall $q_i(k)$ is the bounded divergence probability for S_k, where $y' \in S_k$. But since y and y' agree on bits 1 through k inclusive, we have $\Pr[y$ chosen in Step 2$] \leq \prod_{t \in I_y^{\leq k}} p_{y,t} = q_i(k)$, from which the claim follows. $\square$

To continue with the inductive step, again consider k_*, k_0, and k_1, now corresponding to y_i^*. Then, an argument similar to Equation (24) says $\Pr[y_i^*$ chosen in Step 2$]$ is at least

$$\left(\frac{c}{2^M}\right)^{k_1}(1-s)^{k_0} p_i^* \geq \left(\frac{1}{2^M}\right)^{k_1}\left(1 - \frac{m - k_*}{2^p - 1}\right) q_i^*\left(\frac{1}{2}\right)^{\left|I_{y_i^*}\right| - \left|I_{y_i^*}^{\leq k_i^*}\right|}$$
$$\geq \frac{q_i^*}{2^{Mm}}\left(1 - \frac{m}{2^p - 1}\right), \tag{27}$$

where the first inequality follows from Equation (26), and the second since $\left|I_{y_i^*}\right| - \left|I_{y_i^*}^{\leq k_i^*}\right| \leq k_*$. Now, define $\zeta_i := \Pr[Y = y_i^*] - \Pr[Y \in B_{y_i^*}]$. Applying the argument of Equation (19) yields

$$\zeta_i \geq \left(\frac{1}{2^q}\right)^{(n^c - 1) - |y_i^*|}\left[\frac{q_i^*}{2^{Mm}}\left(1 - \frac{m}{2^p - 1}\right) - q_i^* \sum_{y \in B_{y_i^*}}\left(\frac{1}{2^q}\right)^{|y_i^*| - |y|}\right],$$

where the first q_i^* is due to Equation (27), and the second q_i^* to Lemma 5.3. Thus, similar to Equation (21),

$$\zeta_i \geq \left(\frac{1}{2^q}\right)^{(n^c-1)} \frac{q_i^*}{2^{Mm}} \left[1 - \frac{1}{2^m} - \frac{m}{2^p - 1}\right] > 0.$$

Observing the recurrence that for all i, $\Delta_i' \geq \Delta_{i-1}' + \zeta_i$, unrolling this recurrence yields $\Delta_i' \geq \Delta_1$, which by the base case yields the claim. $\square$

Finally, combining Lemmas 5.1 and 5.2 (recall Lemma 5.1 shows $\Pr(Y \in C) \leq \frac{2^m}{2^p}$; note the proof of Lemma 5.1 still applies after assumption (ii) is dropped) yields that $\Pr[Y \in A] - \Pr[Y \in B \cup C]$ is lower bounded by

$$\Pr[Y \in A] - \Pr[Y \in B] - \Pr[Y \in C] \geq \left(\frac{1}{2^q}\right)^{(n^c-1)} \frac{1}{2^{Mm}} \left[1 - \frac{1}{2^m} - \frac{m}{2^p}\right] - \frac{2^m}{2^p}.$$

For sufficiently large fixed p, this quantity is strictly positive, yielding Theorem 1.4. $\square$

Acknowledgments

References

[1] S. Aaronson. BQP and the polynomial hierarchy. In *Proceedings of the 42nd ACM Symposium on the Theory of Computing (STOC 2010)*, pages 141–150, 2010. DOI: 10.1145/1806689.1806711.

[2] D. Aharonov and T. Naveh. Quantum NP - A survey. Available at arXiv:quant-ph/0210077v1, 2002.

[3] D. Aharonov and L. Zhou. Hamiltonian sparsification and gap-simulations. In Avrim Blum, editor, *10th Innovations in Theoretical Computer Science Conference (ITCS)*, volume 124 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 2:1–2:21. Dagstuhl Publishing, 2018. DOI: 10.4230/LIPIcs.ITCS.2019.2.

[4] D. Aharonov, M. Ben-Or, F. Brandão, and O. Sattath. The pursuit for uniqueness: Extending Valiant-Vazirani theorem to the probabilistic and quantum settings. Available at arXiv:0810.4840v1 [quant-ph], 2008.

[5] A. Ambainis. On physical problems that are slightly more difficult than QMA. In *Proceedings of 29th IEEE Conference on Computational Complexity (CCC)*, pages 32–43, 2014. DOI: 10.1109/ccc.2014.12.

[6] J. Bausch and E. Crosson. Analysis and limitations of modified circuit-to-Hamiltonian constructions. *Quantum*, 2:94, 2018. DOI: 10.22331/q-2018-09-19-94.

[7] J. Bausch, T. Cubitt, and M. Ozols. The complexity of translationally invariant spin chains with low local dimension. *Annales Henri Poincaré*, 18(11):3449–3513, 2017. DOI: 10.1007/s00023-017-0609-7.

[8] J. Bausch, T. Cubitt, A. Lucia, and D. Perez-Garcia. Undecidability of the spectral gap in one dimension. Available at arXiv:1810.01858v1 [quant-ph], 2018.

[9] R. Beigel, L. A. Hemachandra, and G. Wechsung. On the power of probabilistic polynomial time: $P^{NP[\log]} \subseteq PP$. In *Proceedings of the 4th IEEE Conference on Structure in Complexity Theory*, pages 225–227, 1989. DOI: 10.1109/sct.1989.41828.

[10] A. D. Bookatz. QMA-complete problems. *Quantum Information & Computation*, 14(5–6): 361–383, 2014. DOI: 10.26421/QIC14.5-6.

[11] S. Bravyi and D. Gosset. Gapped and gapless phases of frustration-free spin-1/2 chains. *Journal of Mathematical Physics*, 56:061902, 2015. DOI: 10.1063/1.4922508.

[12] S. Bravyi and D. Gosset. Complexity of quantum impurity problems. *Communications in Mathematical Physics*, 356(2):451–500, 2017. DOI: 10.1007/s00220-017-2976-9.

[13] S. Bravyi and M. Hastings. On complexity of the quantum Ising model. *Communications in Mathematical Physics*, 349(1):1–45, 2017. DOI: 10.1007/s00220-016-2787-4.

[14] N. Breuckmann and B. Terhal. Space-time circuit-to-Hamiltonian construction and its applications. *Journal of Physics A: Mathematical and Theoretical*, 47(19):195304, 2014. DOI: 10.1088/1751-8113/47/19/195304.

[15] B. Brown, S. Flammia, and N. Schuch. Computational difficulty of computing the density of states. *Physical Review Letters*, 107(4):040501, 2011. DOI: 10.1103/physrevlett.107.040501.

[16] S. Buss and L. Hay. On truth-table reducibility to SAT. *Information and Computation*, 91(1): 86 – 102, 1991. DOI: 10.1016/0890-5401(91)90075-D.

[17] L. Caha, Z. Landau, and D. Nagaj. Clocks in Feynman's computer and Kitaev's local Hamiltonian: Bias, gaps, idling, and pulse tuning. *Phys. Rev. A*, 97(6):062306, 2018. DOI: 10.1103/PhysRevA.97.062306.

[18] J. Castro and C. Seara. Characterizations of some complexity classes between θ_2^p and Δ_2^p. In *Proceedings of the 9th Annual Symposium on Theoretical Aspects of Computer Science (STACS)*, Lecture Notes in Computer Science, pages 303–317, 1992.

[19] A. Chailloux and O. Sattath. The complexity of the separable Hamiltonian problem. In *Proceedings of 27th IEEE Conference on Computational Complexity (CCC)*, pages 32–41, 2012. DOI: 10.1109/ccc.2012.42.

[20] L. Chen. A note on oracle separations for BQP. Available at arXiv:1605.00619v1 [quant-ph], 2016.

[21] S. Cook. The complexity of theorem proving procedures. In *Proceedings of the 3rd ACM Symposium on Theory of Computing (STOC)*, pages 151–158, 1972. DOI: 10.1145/800157.805047.

[22] T. Cubitt and A. Montanaro. Complexity classification of local Hamiltonian problems. *SIAM J. Comput.*, 45(2):268–316, 2016. DOI: 10.1137/140998287.

[23] T. Cubitt, D. Perez-Garcia, and M. M. Wolf. Undecidability of the spectral gap. *Nature*, 528: 207–211, 2015. DOI: 10.1038/nature16059.

[24] T. Cubitt, A. Montanaro, and S. Piddock. Universal quantum Hamiltonians. *Proceedings of the National Academy of Sciences*, 115(38):9497–9502, 2018. DOI: 10.1073/pnas.1804949115.

[25] B. Fefferman and C. Lin. A complete characterization of unitary quantum space. In A. R. Karlin, editor, *9th Innovations in Theoretical Computer Science Conference (ITCS)*, volume 94 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 4:1–4:21. Dagstuhl Publishing, 2018. DOI: 10.4230/LIPIcs.ITCS.2018.4.

[26] B. Fefferman, R. Shaltiel, C. Umans, and E. Viola. On beating the hybrid argument. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, ITCS '12, pages 468–483, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1115-1. DOI: 10.1145/2090236.2090273. URL http://doi.acm.org/10.1145/2090236.2090273.

[27] R. Feynman. Quantum mechanical computers. *Optics News*, 11(2):11–20, 1985. DOI: 10.1364/on.11.2.000011.

[28] S. Gharibian. *Approximation, proof systems, and correlations in a quantum world*. PhD thesis, University of Waterloo, 2013. Available at arXiv:1301.2632 [quant-ph].

[29] S. Gharibian and J. Kempe. Hardness of approximation for quantum problems. In *Automata, Languages and Programming*, volume 7319 of *Lecture Notes in Computer Science*, pages 387–398, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-31594-7_33.

[30] S. Gharibian and J. Sikora. Ground state connectivity of local Hamiltonians. *ACM Transactions on Computation Theory*, 10(2), 2018. DOI: 10.1145/3186587.

[31] S. Gharibian, Y. Huang, Z. Landau, and S. Woo Shin. Quantum Hamiltonian complexity. *Foundations and Trends in Theoretical Computer Science*, 10(3):159–282, 2015. DOI: 10.1561/0400000066.

[32] S. Gharibian, M. Santha, J. Sikora, A. Sundaram, and J. Yirka. Quantum generalizations of the polynomial hierarchy with applications to QMA(2). In I. Potapov, P. Spirakis, and J. Worrell, editors, *43rd International Symposium on Mathematical Foundations of Computer Science (MFCS 2018)*, volume 117 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 58:1–58:16. Dagstuhl Publishing, 2018. DOI: 10.4230/LIPIcs.MFCS.2018.58.

[33] S. Gharibian, S. Piddock, and J. Yirka. Oracle complexity classes and local measurements on physical Hamiltonians. Available at arXiv:1909.05981 [quant-ph], 2019.

[34] J. Gill. Computational complexity of probabilistic Turing machines. *SIAM Journal on Computing*, 6(4):675–695, 1977. DOI: 10.1137/0206049.

[35] A. P. Gilyén and O. Sattath. On preparing ground states of gapped Hamiltonians: An efficient quantum Lovász local lemma. In *IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 439–450, 2017. DOI: 10.1109/FOCS.2017.47.

[36] O. Goldreich. On promise problems: A survey. In O. Goldreich, A.L. Rosenberg, and A.L. Selman, editors, *Theoretical Computer Science*, volume 3895 of *Lecture Notes in Computer Science*, pages 254–290. Springer, Berlin, Heidelberg, 2006. DOI: 10.1007/11685654_12.

[37] D. Gosset and E. Mozgunov. Local gap threshold for frustration-free spin systems. *Journal of Mathematical Physics*, 57:091901, 2016. DOI: 10.1063/1.4962337.

[38] D. Gosset, J. C. Mehta, and T. Vidick. QCMA hardness of ground space connectivity for commuting Hamiltonians. *Quantum*, 1:16, 2017. DOI: 10.22331/q-2017-07-14-16.

[39] J. Hartmanis. Sparse complete sets for NP and the optimal collapse of the polynomial hierarchy. In G. Rozenberg and A. Salomaa, editors, *Current Trends in Theoretical Computer Science*, pages 403–411. World Scientific, 1993. DOI: 10.1142/9789812794499_0029.

[40] L. Hemachandra. The strong exponential hierarchy collapses. *Journal of Computer and System Sciences*, 39(3):299 – 322, 1989. DOI: 10.1016/0022-0000(89)90025-1.

[41] Edith Hemaspaandra, Lane A. Hemaspaandra, and Jörg Rothe. Exact analysis of Dodgson elections: Lewis Carroll's 1876 voting system is complete for parallel access to NP. *Journal of the ACM*, 44(6):806–825, 1997. DOI: 10.1145/268999.269002.

[42] R. Karp. Reducibility among combinatorial problems. In R.E. Miller, J.W. Thatcher, and J.D. Bohlinger, editors, *Complexity of Computer Computations*, The IBM Research Symposia Series, pages 85–103. Springer, Boston, MA, 1972. DOI: 10.1007/978-1-4684-2001-2_9.

[43] J. Kempe, A. Kitaev, and O. Regev. The complexity of the local Hamiltonian problem. *SIAM Journal on Computing*, 35(5):1070–1097, 2006. DOI: 10.1137/s0097539704445226.

[44] A. Kitaev and J. Watrous. Parallelization, amplification, and exponential time simulation of quantum interactive proof systems. In *Proceedings of the 32nd ACM Symposium on Theory of Computing (STOC)*, pages 608–617, 2000. DOI: 10.1145/335305.335387.

[45] A. Kitaev, A. Shen, and M. Vyalyi. *Classical and Quantum Computation*. American Mathematical Society, 2002.

[46] L. Levin. Universal sequential search problems. *Problems of Information Transmission*, 9(3): 265–266, 1973.

[47] Y. K. Liu. Consistency of local density matrices is QMA-complete. In J. Díaz, K. Jansen, J.D.P. Rolim, and U. Zwick, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, volume 4110 of *Lecture Notes in Computer Science*, pages 438–449. Springer, Berlin, Heidelberg, 2006. DOI: 10.1007/11830924_40.

[48] J. Lockhart and C. E. González-Guillén. Quantum state isomorphism. Available at arXiv:1709.09622v1 [quant-ph], 2017.

[49] C. Marriott and J. Watrous. Quantum Arthur-Merlin games. *Computational Complexity*, 14 (2):122–152, 2005. DOI: 10.1007/s00037-005-0194-x.

[50] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 10th edition, 2011.

[51] R. Oliveira and B. M. Terhal. The complexity of quantum spin systems on a two-dimensional square lattice. *Quantum Information & Computation*, 8(10):900–924, 2008. DOI: 10.26421/QIC8.10.

[52] T. J. Osborne. Hamiltonian complexity. *Reports on Progress in Physics*, 75(2):022001, 2012. DOI: 10.1088/0034-4885/75/2/022001.

[53] S. Piddock and A. Montanaro. The complexity of antiferromagnetic interactions and 2D lattices. *Quantum Information & Computation*, 17(7&8):636–672, 2017. DOI: 10.26421/QIC17.7-8.

[54] S. Piddock and A. Montanaro. Universal qudit Hamiltonians. Available at arXiv:1802.07130v1 [quant-ph], 2018.

[55] R. Raz and A. Tal. Oracle separation of BQP and PH. Available at Electronic Colloquium on Computational Complexity (ECCC), 2018.

[56] Z. Remscrim. The Hilbert function, algebraic extractors, and recursive Fourier sampling. In *IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 197–208, Oct 2016. DOI: 10.1109/FOCS.2016.29.

[57] P. Schnoebelen. Oracle circuits for branching-time model checking. In J. C.M. Baeten, J. K. Lenstra, J. Parrow, and G. J. Woeginger, editors, *Automata, Languages and Programming*, volume 2719 of *Lecture Notes in Computer Science*, pages 790–801, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg. DOI: 10.1007/3-540-45061-0_62.

[58] N. Schuch and F. Verstraete. Computational complexity of interacting electrons and fun-

damental limitations of Density Functional Theory. *Nature Physics*, 5:732–735, 2009. DOI: 10.1038/nphys1370.

[59] Y. Shi and S. Zhang. Note on quantum counting classes. Available at `http://www.cse.cuhk.edu.hk/syzhang/papers/SharpBQP.pdf`.

[60] L.G. Valiant and V.V. Vazirani. NP is as easy as detecting unique solutions. *Theoretical Computer Science*, 47:85–93, 1986. DOI: 10.1016/0304-3975(86)90135-0.

[61] M. Vyalyi. QMA=PP implies that PP contains PH. Available at Electronic Colloquium on Computational Complexity (ECCC), 2003.

[62] J. Watrous. Quantum computational complexity. In R. Meyers, editor, *Encyclopedia of Complexity and System Science*. Springer, 2013. DOI: 10.1007/978-3-642-27737-5_428-3.

[63] T. Yamakami. Quantum NP and a quantum hierarchy. In Baeza-Yates R., Montanari U., and Santoro N., editors, *Foundations of Information Technology in the Era of Network and Mobile Computing*, volume 96 of *IFIP — The International Federation for Information Processing*, pages 323–336. Springer, Boston, MA, 2002. DOI: 10.1007/978-0-387-35608-2_27.

A   Additional proofs

Lemma A.1. *Assume the terminology of Lemma 4.3. Then, there exists a valid history state $|\psi'\rangle$ on W, C, Q, and Q' such that $|\langle\psi|\psi'\rangle|^2 \geq 1 - O(\gamma^{-2})$.*

Proof. Recall that the smallest non-zero eigenvalue of $H_1 = \Delta(H_{\text{in}} + H_{\text{prop}} + H_{\text{stab}})$ is at least $J := \pi^2\Delta/(64L^3)$ (by Lemma 2.2), $\delta = 1/\Delta$, and $\eta \geq \max(\|H_2\|_\infty, 1) =: K$. Lemma 4.1 now implies there exists a valid history state $|\psi'\rangle$ satisfying

$$
\begin{aligned}
|\langle\psi|\psi'\rangle|^2 &\geq 1 - \left(\frac{K + \sqrt{K^2 + \delta(J - 2K)}}{J - 2K}\right)^2 \\
&\geq 1 - \left(\frac{\eta + \sqrt{\eta^2 + \frac{\pi^2}{64L^3}}}{\frac{\pi^2\Delta}{64L^3} - 2\eta}\right)^2 \\
&\geq 1 - \left(\frac{(1 + \sqrt{2})\eta}{\frac{\pi^2\Delta}{64L^3} - 2\eta}\right)^2 \\
&= 1 - \left(\frac{(1 + \sqrt{2})64L^3\eta}{\pi^2\Delta - 128L^3\eta}\right)^2 \\
&\geq 1 - \Theta\left(\frac{1}{\gamma^2}\right),
\end{aligned}
$$

where the second inequality follows since $0 \leq K \leq \eta$, the third since $\pi^2/(64L^3) \leq 1$ for $L \geq 1$ and $1 \leq \eta^2$, and the last since $\Delta = 128L^3\eta\gamma$ and $\gamma \geq 1$. $\qquad\square$