Importance of the Spectral gap in Estimating Ground-State Energies

Abhinav Deshpande,[1,*,§] Alexey V. Gorshkov,[1,†] and Bill Fefferman[2,‡]

[1] *Joint Center for Quantum Information and Computer Science and Joint Quantum Institute, NIST/University of Maryland, College Park, Maryland 20742, USA*

[2] *Department of Computer Science, University of Chicago, Chicago, Illinois 60637, USA*

(Received 1 August 2020; revised 1 November 2022; accepted 15 November 2022; published 9 December 2022)

The field of quantum Hamiltonian complexity lies at the intersection of quantum many-body physics and computational complexity theory, with deep implications to both fields. The main object of study is the LOCALHAMILTONIAN problem, which is concerned with estimating the ground-state energy of a local Hamiltonian and is complete for the class QMA, a quantum generalization of the class NP. A major challenge in the field is to understand the complexity of the LOCALHAMILTONIAN problem in more physically natural parameter regimes. One crucial parameter in understanding the ground space of any Hamiltonian in many-body physics is the spectral gap, which is the difference between the smallest two eigenvalues. Despite its importance in quantum many-body physics, the role played by the spectral gap in the complexity of LOCALHAMILTONIAN is less well understood. In this work, we make progress on this issue by considering the precise regime, in which one estimates the ground-state energy to within inverse exponential precision. Computing ground-state energies precisely is a task that is important for quantum chemistry and quantum many-body physics. In the setting of inverse-exponential precision, there is a surprising result by Fefferman and Lin that the complexity of LOCALHAMILTONIAN is magnified from QMA to PSPACE, the class of problems solvable in polynomial space (but possibly exponential time). We clarify the reason behind this boost in complexity. Specifically, we show that the full complexity of the high-precision case only comes about when the spectral gap is exponentially small. As a consequence of the proof techniques developed to show our results, we uncover important implications for the representability and circuit complexity of ground states of local Hamiltonians, the theory of uniqueness of quantum witnesses, and techniques for the amplification of quantum witnesses in the presence of postselection.

DOI: 10.1103/PRXQuantum.3.040327

I. INTRODUCTION

Several exotic phenomena in our world occur only at very low temperatures, most notably, those occurring in condensed matter such as superconductivity, superfluidity, and the fractional and integer quantum Hall effects. Beyond these examples in condensed-matter physics, the low-energy physics of systems of several interacting particles is of interest in several fields such as particle physics, atomic, molecular, and optical physics, chemistry, and quantum computing. Accordingly, finding effective descriptions of ground states of many-body Hamiltonians is a very natural and important task in physics.

Given the prevalence and importance of this task, a natural question is that of the computational difficulty of solving this task in naturally occurring situations. Questions such as the hardness of solving a computational task belong to the domain of computational complexity theory. A good proxy for the difficulty of obtaining ground-state descriptions is the difficulty of solving a weaker problem, namely that of computing ground-state energies of many-body Hamiltonians. This question is studied in the domain known as "Hamiltonian complexity" (see, e.g., Ref. [1]), an area of research at the intersection of quantum many-body physics and computational complexity theory.

This area of research originated from Kitaev's result that the LOCALHAMILTONIAN problem, which is the problem of computing the ground-state energy of a local Hamiltonian, is QMA-complete [2] (we refer a reader unfamiliar with complexity-theoretic language to Sec. I A). The

*abhinavd@caltech.edu
†gorshkov@umd.edu
‡wjf@uchicago.edu
§Current affiliation: Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, California 91125, USA.

complexity class QMA is the quantum generalization of NP. Kitaev's result may be viewed as an analogue of the seminal Cook-Levin theorem [3,4] in computer science, generalized to the setting of quantum constraint satisfaction problems. Despite the tremendous amount of progress in understanding the power of local Hamiltonians, many important questions remain, such as whether the task remains hard under less-demanding notions of approximation error [5–7] and whether there exist short classical descriptions of ground states of local Hamiltonians (see, e.g., Refs. [5,8,9]), among others.

One important question about LOCALHAMILTONIAN is the role played by the spectral gap. The spectral gap is a traditionally important quantity in the context of ground-state properties of any physical system and is defined as the difference between the smallest two eigenvalues of the Hamiltonian. Many important families of Hamiltonians in physics have the "gap property," meaning that the spectral gap in the limit of large system size $n \to \infty$ is lower bounded by a constant. Important conjectures in physics are concerned with the existence of the gap property for certain Hamiltonians [10,11], a problem that is known to be undecidable in general [12]. Furthermore, the existence of a spectral gap implies various tractability results for the ground states of Hamiltonians. For instance, in one dimension, the gap property significantly restricts the entanglement structure of ground states through the area law of entanglement, implying efficient classical representations of the same [13], and further, classically efficient algorithms to compute the ground-state energy [14,15]. It is not known whether these properties hold for higher dimensions.

Despite the physical importance of the spectral gap, its role in the context of the LOCALHAMILTONIAN problem itself is much less clear. In particular, it is not known whether LOCALHAMILTONIAN is QMA-complete in the presence of nontrivial lower bounds on the spectral gaps, even when the lower bound is $\Omega(1/\mathsf{poly}(n))$ [16,17]. Meanwhile, if the spectral gap is promised to be lower bounded by a constant, there are no-go results [18,19] that rule out any QMA-hardness proof that proceeds by generalizing the clock construction technique. This technique underlies all known QMA-completeness results, in analogy with the theory of NP-completeness, where the Cook-Levin theorem plays a foundational role. Therefore, Hamiltonians with any nontrivial lower bounds on the spectral gap can be less complex than the general case.

In this work, we take an initial step towards answering the question of the role played by the spectral gap in the LOCALHAMILTONIAN problem. To do so, we study QMA in the precise setting, i.e., the class PreciseQMA. In the precise setting, the completeness (the minimum probability of accepting a correct statement) and soundness (the maximum probability of accepting an incorrect statement) of the protocol are separated by a quantity called the promise gap that scales inverse exponentially in the size of the input. For the LOCALHAMILTONIAN problem, this translates to computing the ground-state energy to within inverse-exponential precision in the system size.

Computing ground-state energies to inverse-exponential precision is not an artificial task. This task corresponds to computing polynomially many digits of the answer, which is very desirable in some cases [20]. Algorithms whose runtimes scale as $\mathsf{polylog}(1/\epsilon)$ for additive error ϵ can compute quantities to inverse-exponential precision in polynomial time, and such algorithms have been found for Hamiltonian simulation and linear systems [21–23]. There are also situations where precise knowledge of the ground-state energy of a Hamiltonian is essential. For example, in quantum chemistry, chemical reactivity rates depend on the Born-Oppenheimer potential-energy surface for the nuclei. Each point on this surface is an electronic ground-state energy for a particular arrangement of the nuclei. Small uncertainties in the ground-state energy can exponentially influence the calculated rate k via Arrhenius's law $k \propto \exp[-\beta \Delta E]$, where ΔE is an energy barrier and β the inverse temperature (see, e.g., Ref. [24]). Another example is in condensed-matter physics, where algorithms such as the density matrix renormalization group routinely compute several digits of the ground-state energy (see, e.g., Ref. [25]). Precise knowledge of the ground-state energy can enable one to identify the locations of quantum phase transitions by identifying nonanalyticities [26]. Interestingly, the class of Hamiltonians for which the energy can be precisely measured correspond to Hamiltonians that can be fast forwarded [27].

Fefferman and Lin [28] studied the complexity of the class PreciseQMA, and showed the mysterious result that it equals PSPACE. This is surprising since QMA $\subseteq$ PP [29–31] (also see Fig. 1 below for reference), and an alternative characterization of the class PP is PreciseBQP, the precise analogue of BQP (the class of problems efficiently solvable on quantum computers). Since PreciseBQP can handle inverse-exponentially small promise gaps and contains QMA, one might have expected that adding the modifier Precise- to QMA would not have changed the power of the class by much.

We provide an explanation for this seemingly unexpected boost in complexity from QMA, which is a subset of PP, to PreciseQMA, which equals PSPACE [32]. Specifically, we find that in order for the precise version of LOCALHAMILTONIAN, i.e., PRECISELOCALHAMILTONIAN, to be PSPACE-hard, the spectral gap of the Hamiltonian must necessarily shrink superpolynomially with the size of the system n (measured by the number of qudits in the system). We give strong evidence that if the spectral gap shrinks no faster than a polynomial in the system size, i.e., if the spectral gap is bounded by $\Omega(1/\mathsf{poly})$, the complexity of the problem is strictly less powerful. In particular, we show that this problem characterizes the complexity

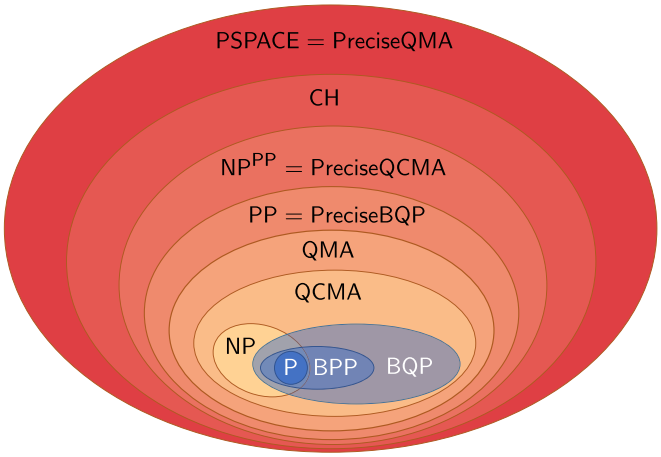

FIG. 1. Major complexity classes featuring in this work. PSPACE, $\mathsf{NP}^{\mathsf{PP}}$, and PP can be defined purely in terms of quantum computation, and are equal to PreciseQMA, PreciseQCMA, and PreciseBQP, respectively. All inclusions except $\mathsf{P} \subseteq \mathsf{BPP}$ are believed to be strict.

class PP, which is a subset of PSPACE and is widely believed to be distinct from PSPACE. If the problem were PSPACE-hard, the so-called counting hierarchy, defined as $\mathsf{CH} = \mathsf{PP} \cup \mathsf{PP}^{\mathsf{PP}} \cup \cdots$ [33], would collapse, which is considered an unlikely possibility. Our results therefore bring out the importance of the spectral gap, a quantity not well understood so far in Hamiltonian complexity.

Another main result of ours concerns the question of existence of polynomial-size quantum circuits to prepare ground states of local Hamiltonians. This is an important question that has implications in circuit complexity of ground states of natural Hamiltonians and is directly related to whether natural Hamiltonians can be efficiently cooled down to zero temperature. In complexity-theoretic language, the question may be phrased in terms of the power of classical versus quantum witnesses in Merlin-Arthur proof systems, or, more formally, the so-called QMA versus QCMA question. The (in)equivalence of these classes is an important open question in quantum complexity theory and many-body physics, which has remained unsettled despite recent progress in the oracle setting (see, e.g., Refs. [8,9]). The precise version of QCMA, or PreciseQCMA, is known to be equal to $\mathsf{NP}^{\mathsf{PP}}$ (see, e.g., Refs. [34,35]), indicating a separation between PreciseQCMA and PreciseQMA ($=$ PSPACE) unless the counting hierarchy collapses. Interestingly, we show strong *equivalence* results for the PreciseQMA versus PreciseQCMA question in the presence of spectral gaps.

Our results and proof techniques we develop here also have consequences for other areas of quantum computation, complexity theory, and many-body physics. Our second main result mentioned earlier roughly says that in the precise regime, the promise of an inverse-polynomial lower bound on the spectral gap is equivalent to the promise that there exists a polynomial-size circuit to prepare the ground state. This leads to an interesting conjecture we make in Sec. I D, which could have a bearing on the performance of near-term quantum algorithms for quantum chemistry and on the circuit complexity of various low-energy states, which is an important question in many-body physics and gravitational and high-energy physics [36,37]. We obtain some additional evidence for the conjecture in Sec. V by showing that some implications of the conjecture are correct. Furthermore, our results can shed light on an attempt to give a quantum-inspired reproof [38,39] of the celebrated $\mathsf{IP} = \mathsf{PSPACE}$ result [40] via interactive protocols for the class PreciseQMA. Our results also allow us to rule out sufficiently strong error-reduction techniques for the class postQMA.

This paper is structured as follows. In the rest of Sec. I we give an introduction to the basic notions of complexity theory used in this work (which an experienced reader may skip), state and refer to the main results, give a high-level overview of the proof techniques and their implications, and discuss the relation of our results to other work in the literature. In Sec. II, we give the definitions of some other complexity classes and define some new classes that appear in this work. We also define natural problems complete for these classes. We then formally state the results pertaining to the class PP in Sec. III and PSPACE in Sec. IV. We also consider the complexity of related classes in Sec. V, after which the Appendices have detailed proofs of our claims.

A. Preliminaries

Here, we give a very brief introduction to the complexity-theoretic definitions and terminology in this work. The reader is referred to a textbook (e.g., Refs. [41, 42]) for a more pedagogical exposition. We are generally concerned with decision problems, where the answer is either "YES" or "NO." These problems can be cast as follows: given an instance x, the task is to decide if it belongs to the class of YES instances ($x \in A_{\text{yes}}$), or to the class of NO instances ($x \in A_{\text{no}}$). In principle, there can be problems where certain instances (for example, ill-defined ones) belong neither to A_{yes} nor A_{no}. In such cases, we either allow an algorithm to answer arbitrarily, or we supplant the problem with a promise that such instances never occur. These are called *promise problems*.

In complexity theory, one is typically interested in the resources taken to solve various classes of decision problems. Furthermore, one is interested in how the resource cost scales with the size of the problem to be solved, which is quantified in terms of the length of the input, often denoted n. In this work, we use the notation $\mathsf{poly}(n)$ to denote any function that can be upper bounded by $O(n^c)$ for some constant $c = \Theta(1)$. We also denote $\mathsf{exp}(n)$ to

be any function $2^{\mathsf{poly}(n)}$. We omit the dependence on n, which in our work is taken to be the number of qudits. The results in this work are applicable generally to qudits of any dimension $d \geq 2$, but we often work with qubits in our proofs.

We first define the class BQP (bounded-error quantum polynomial time), which is the class of problems solvable in polynomial time (in n) on a quantum computer with bounded error. The error here is measured via the parameters c (minimum probability of saying "YES" if the answer is YES) and s (maximum probability of saying "YES" if the answer is NO). More formally, we have the following definition.

Definition 1 (BQP$[c,s]$)**:** The class BQP$[c,s]$ is the class of promise problems $A = (A_{\text{yes}}, A_{\text{no}})$ such that, for every instance x, there is a uniformly generated circuit U_x of size poly(n) acting on state $|0^{\otimes m}\rangle$ for $m = \mathsf{poly}(n)$, with the property that upon measuring the first bit at the output, o, also called the decision qubit, we have

(a) if $x \in A_{\text{yes}}$, then $\Pr(o = 1) \geq c$,
(b) if $x \in A_{\text{no}}$, then $\Pr(o = 1) \leq s$.

In the above, we imagine that a quantum computer applies a circuit U_x that acts on a standard initial state, measures the first bit at the output, and says YES ("accepts") or NO ("rejects"), depending on whether the bit is measured to be in state $|1\rangle$ or $|0\rangle$. The choice of the bit to measure at the output is arbitrary. The term *uniformly generated circuit* means that, given an instance x, there is a polynomial-time classical algorithm to generate a description of the circuit U_x to be applied.

Definition 2: We define $\mathsf{BQP} = \bigcup_{c-s\geq 1/\mathsf{poly}} \mathsf{BQP}[c,s]$.

The class BQP is the quantum generalization of the class BPP (bounded-error probabilistic polynomial time), the class of problems solvable in polynomial time by a randomized classical computer.

We now come to the class QMA (quantum Merlin Arthur), which is a quantum generalization of NP. We imagine two parties, Merlin (the prover) and Arthur (the verifier). The prover would like to convince the verifier that a certain problem instance x is a YES instance. The prover, who is computationally unbounded, can supply any state $|\psi\rangle$ on $w = \mathsf{poly}(n)$ qubits to the verifier as a "proof" or "witness." The verifier can apply any circuit of their choice acting on some m qubits they possess and the witness state, and accept or reject based on the outcome of a decision bit. The class QMA is the class of problems such that a YES answer can be reliably verified in this way and in case the answer is NO, no matter what state is sent by the (possibly cheating) prover, the verifier rejects with high probability. Just like with BQP, QMA is defined with respect to parameters c and s, which are called *completeness* and *soundness*, respectively.

Definition 3 (QMA$[c,s]$)**:** The class QMA$[c,s]$ is the class of problems $A = (A_{\text{yes}}, A_{\text{no}})$ with the property that, for every instance x, there exists a uniformly generated circuit U_x with the following properties: U_x is of size poly(n) and acts on an input state $|0\rangle^{\otimes m}$, together with a proof (or witness) state $|\Psi\rangle$ of size w supplied by an arbitrarily powerful prover. Both m and w are bounded by polynomials in n. Upon measuring the decision qubit o of the output register, the verifier accepts if $o = 1$, and rejects otherwise. We say that $A = (A_{\text{yes}}, A_{\text{no}})$ is a QMA$[c,s]$ problem if and only if

(a) if $x \in A_{\text{yes}}$, then there exists $|\Psi\rangle$ such that $\Pr(o = 1) \geq c$,
(b) if $x \in A_{\text{no}}$, then for all $|\Psi\rangle$, $\Pr(o = 1) \leq s$.

The class QMA is defined as $\bigcup_{c-s\geq 1/\mathsf{poly}} \mathsf{QMA}[c,s]$.

To characterize the complexity of a problem, we give "upper" and "lower" bounds on the complexity of the problem. Upper bounds are statements of the form "X $\in$ Y," which means that problem X can be solved with access to a solver for the complexity class Y. For example, Shor [43] proved that FACTORING $\in$ BQP, which means that quantum computers can factor integers in polynomial time (since quantum computers may be viewed as "solvers for the class BQP"). Lower bounds are statements of the form "X is Y-hard." This means that problem X is as hard as any problem in Y. Such statements are often shown via *reductions*. One assumes the existence of an *oracle*, a black box that can solve any instance of problem X in one timestep. A reduction is a mapping from a complexity class Y to a problem X with the property that any problem in Y can be solved by querying the oracle for X . If such a reduction exists, it implies that the problem X is at least as hard as any problem in class Y. If a problem X is both in class Y and is Y-hard then it means that the upper and lower bounds to the problem match. This means that problem X is the hardest in the class it belongs to, namely Y. In this case, we say "X is Y-complete" or "X is complete for Y." We also denote by Y^{Z} the class of problems solvable by a Y machine with access to an oracle for any problem in Z.

Lastly, we depict the known inclusions between complexity classes in Fig. 1. We also describe here the classes not mentioned so far. The class P is the class of problems efficiently solvable on classical computers, while NP is the class of problems for which a YES answer may be verified efficiently, via a protocol involving a classical prover and classical verifier. The class QCMA is analogous to QMA, except that the prover sends a classical witness instead of a quantum one. As for PP, it suffices to know that it equals PreciseBQP, a precise version of BQP. The class $\mathsf{NP}^{\mathsf{PP}}$ is a subset of $\mathsf{PP}^{\mathsf{PP}}$, since NP $\subseteq$ PP. These classes belong to the counting hierarchy (CH), which is defined as $\mathsf{CH} = \mathsf{PP} \cup \mathsf{PP}^{\mathsf{PP}} \cup \cdots$ [33]. All of these classes are

in PSPACE, the class of problems solvable on classical computers that use polynomial space (but which are free to use exponential time).

B. Results

We describe a general problem we study here, called (δ, Δ)-LOCALHAMILTONIAN. Informally, it is the problem of estimating the ground-state energy of a given k-local Hamiltonian acting on n qudits to additive error at most δ, when promised that the spectral gap is at least Δ (see precise definitions in Sec. II). In the absence of any bound on the spectral gap (i.e., $\Delta = 0$), the problem $(1/\mathsf{poly}(n), 0)$-LOCALHAMILTONIAN is, by definition, the same as k-LOCALHAMILTONIAN, which is complete for QMA for $k \geq 2$ [2,44,45]. Meanwhile, $(1/\mathsf{exp}(n), 0)$-LOCALHAMILTONIAN is, by definition, PRECISE-k-LOCALHAMILTONIAN [28], which is complete for PreciseQMA. We henceforth suppress the dependence on the number of qudits n in the notation exp and poly for the rest of the paper.

To our knowledge, Aharonov *et al.* [16] were the first to study the k-LOCALHAMILTONIAN problem in the presence of a spectral gap. Specifically, they considered $(1/\mathsf{poly}, 1/\mathsf{poly})$-LOCALHAMILTONIAN and showed it to be complete for the class PGQMA (polynomially gapped QMA). The definition of PGQMA, which is given in Sec. II, depends on a notion of a spectral gap for *proof systems*, distinct from that for Hamiltonians. For complexity classes associated with proof systems such as QMA, QCMA, and the variants we study in this work, the spectral gap corresponds to the gap in the highest and second-highest accept probabilities of the optimal witness and the next-optimal orthogonal witness. *A priori*, the two notions of a spectral gap have no relation with each other. We show that the two notions are equivalent for various cases (δ and Δ each behaving as $1/\mathsf{poly}$ or $1/\mathsf{exp}$), by showing that (δ, Δ)-LOCALHAMILTONIAN is complete for the appropriate spectral-gapped QMA class.

To understand the relation between the gapped QMA classes and the regular versions without a spectral gap, we focus on the precise regime, so that $\delta = 1/\mathsf{exp}$ henceforth for the rest of this section. By specifying the spectral gap to be $\Omega(1/\mathsf{poly})$, we get the problem $(1/\mathsf{exp}, 1/\mathsf{poly})$-LOCALHAMILTONIAN. We show in Lemma 1 below that this problem is in a class we call PrecisePGQMA (precise polynomially gapped QMA), which is the precise analogue of PGQMA. We also show (Lemma 2 below) that PrecisePGQMA $\subseteq$ PP, implying that PrecisePGQMA is likely different from PreciseQMA, which equals PSPACE. Specifically, assuming that PP $\neq$ PSPACE, there is a separation between PrecisePGQMA and PreciseQMA. The PP upper bound on PrecisePGQMA is optimal: we show that $(1/\mathsf{exp}, 1/\mathsf{poly})$-LOCALHAMILTONIAN is PP-hard (Lemma 4 below). Thus, we tightly characterize the complexity of the class by showing that PrecisePGQMA = PP and prove that $(1/\mathsf{exp}, 1/\mathsf{poly})$-LOCALHAMILTONIAN is its associated complete problem.

The results in the previous paragraph show that the PSPACE-hardness result of Ref. [28] relies on the fact that the spectral gaps of the associated Hamiltonians can decay rapidly with the system size. This raises the question of the maximum scaling of the spectral gap required in order to retain PSPACE-hardness. This is an important question since if the PSPACE-hardness results only apply when there is no promise whatsoever on the spectral gap, it would indicate that PSPACE-hardness of PRECISE-k-LOCALHAMILTONIAN is artificial. We rule out this possibility by showing that if the spectral gap is bounded below by $1/\mathsf{exp}$, i.e., if we consider the problem $(1/\mathsf{exp}, 1/\mathsf{exp})$-LOCALHAMILTONIAN, the problem remains PSPACE-hard. Specifically, we show in Theorem 2 below that this problem is complete for a class called PreciseEGQMA (precise exponentially gapped QMA). Next, we show that PreciseEGQMA equals PSPACE (Theorem 10 below), implying that instances with $\Omega(1/\mathsf{exp})$ spectral gaps are no less complex than the general case.

Lastly, we consider the analogues of these classes when the witness is classical, which gives us the classes QCMA (quantum classical Merlin Arthur), PreciseQCMA, PrecisePGQCMA (precise polynomially gapped QCMA), and PreciseEGQCMA (precise exponentially gapped QCMA). The complete problems for these classes are the appropriate versions of the LOCALHAMILTONIAN problem under the additional promise that there is an efficient classical description of a circuit to prepare a low-energy state, as we show in Theorems 3 to 6 below. We define this problem in Sec. II A and denote it (δ, Δ)-GS-DESCRIPTION-LOCALHAMILTONIAN, which is the problem of computing the ground-state energy to additive error δ, given the promise that there exists a polynomial-size circuit to prepare a low-energy state and promised that the spectral gap of the Hamiltonian is at least Δ. As stated in Corollary 1 below, we show that PrecisePGQCMA has the same complexity as PrecisePGQMA, implying that in the precise setting, once there is a $\Omega(1/\mathsf{poly})$ promise on the spectral gap, a further promise that there exists an efficient circuit to prepare a low-energy state is redundant. We comment more on this result in Sec. I D.

In Table I, we give an overview of the parameter dependence of the complexity of the two main problems studied in this work, namely (δ, Δ)-LOCALHAMILTONIAN and (δ, Δ)-GS-DESCRIPTION-LOCALHAMILTONIAN. The problems are completely characterized by the appropriately gapped versions of QMA or QCMA, or their precise variants. The complexity class in any cell in the table is a subset of all the classes below it in the same column, since these classes correspond to

TABLE I. Complexity of variants of the LOCALHAMILTONIAN problem as a function of the parameters δ, the promise gap, and Δ, the spectral gap. The problem is complete for the class mentioned in each cell. For reference, we mention in curly brackets the theorem number corresponding to the results proved in this work. The question mark corresponding to the entry EGQMA indicates that the result is a conjecture and the notation "$=_R$" denotes equivalence under randomized reductions (defined in Sec. V C).

	(δ, Δ)-GS-DESCRIPTION-LOCALHAMILTONIAN		(δ, Δ)-LOCALHAMILTONIAN	
Spectral gap (Δ)	$\delta = 1/\mathsf{poly}$	$\delta = 1/\mathsf{exp}$	$\delta = 1/\mathsf{poly}$	$\delta = 1/\mathsf{exp}$
$1/\mathsf{poly}$	PGQCMA {5} ($=_R$ QCMA {2})	PrecisePGQCMA {7} (= PP {9})	PGQMA	PrecisePGQMA {1} (= PP {8})
$1/\mathsf{exp}$	EGQCMA ($=_R$ QCMA {2})	PreciseEGQCMA {6} (= $\mathsf{NP}^{\mathsf{PP}}$ {11})	EGQMA(?)	PreciseEGQMA {2} (= PSPACE {10})
0	QCMA {3}	PreciseQCMA{4} (= $\mathsf{NP}^{\mathsf{PP}}$)	QMA	PreciseQMA (= PSPACE)

weaker promises on the spectral gap. Similarly, the complexity class associated with (δ, Δ)-GS-DESCRIPTION-LOCALHAMILTONIAN is a subset of that associated with (δ, Δ)-LOCALHAMILTONIAN, because the former problem is associated with an extra promise. While we have given evidence that PrecisePGQMA $\neq$ PreciseQMA, it is unknown whether the same holds for the question PGQMA $\stackrel{?}{=}$ QMA. Similarly, while we have proved that PreciseEGQMA $=$ PreciseQMA, it would be interesting to see if a similar result holds for EGQMA.

C. Techniques

Here, we give an overview of the primary techniques used in proving our results.

Imaginary-time evolution and the power method.—To show the containment PrecisePGQMA $\subseteq$ PP, we use a technique called the "power method" [46]. The broad idea behind the algorithm is that if a matrix A is promised to have a spectral gap between the largest two eigenvalues, the behavior of A^d for large d is dominated by the largest eigenvalue. We give a PP algorithm to compute $\mathrm{Tr}(A^d)$ for an exponentially large matrix A and $d = \mathsf{poly}(n)$ for a wide class of matrices A. This wide class includes sparse matrices and matrices representing local observables as special cases. The PP algorithm uses the Feynman sum-over-paths idea [47] to express the trace as a sum over 2^{poly} many terms, each of which is a product over quantities of the form $\langle x|R|y\rangle$ for some matrix R whose entries are efficiently computable. A PP algorithm can decide whether the sum over 2^{poly} many terms, each term computable in polynomial time, is above or below a threshold.

The power method is closely related to another technique called the "cooling algorithm," inspired by a brief discussion by Schuch *et al.* [48]. The idea is that letting a system evolve in imaginary time can produce an unnormalized state close to the ground state. Imaginary-time evolution is a linear, albeit nonunitary, operation and produces an unnormalized state ρ' in general. Schuch *et al.* relied on a quantum characterization of PP, namely postBQP. The class postBQP [49] is the class of problems solvable in polynomial time on a quantum computer with access to the resource of postselection, which is the ability to condition on exponentially unlikely events. Aaronson [49] showed that any linear operation, even nonunitary ones, may be simulated in postBQP. The algorithm of Schuch *et al.* [48] proposes to decompose the imaginary-time evolution operation $\mathsf{exp}[-\beta H]$ into a series of local operations $\mathsf{exp}[-\beta H_i]$ using Trotterization, and implementing each local operation using the resource of postselection. Unfortunately, the state-of-the-art error bounds for Trotterization of imaginary-time evolution [50] give, at best, a multiplicative error that is exponential in n (see also Refs. [51,52]), and hence this technique does not work in the precise regime. We prove a more general statement about precise computation of ground-state local observables for Hamiltonians with a spectral gap using exact imaginary time evolution as opposed to a Trotterized version. Specifically, we give a P^{PP} algorithm that provably works not just for $1/\mathsf{poly}$ precision, but also $1/\mathsf{exp}$ precision in computing local observables in addition to the Hamiltonian. Our technique is closely related to the power method, since the core of the algorithm is to compute expectation values of powers of the Hamiltonian.

Small-penalty clock construction.—Our second major technical contribution is a modification of the clock construction that we call the small-penalty clock construction. One of the ways this technique is useful is as follows. As mentioned earlier and as will be described in detail in Sec. II, it is possible to consider spectral-gapped versions of both the LOCALHAMILTONIAN problem and the class QMA and their variants. We have already discussed the (natural) notion of a spectral gap for Hamiltonians. For QMA and related classes, the spectral gap is related to the difference in accept probabilities between the optimal and next-optimal witnesses. Our technique allows us to bridge the notion of spectral gap in both cases by constructing spectral-gap-preserving reductions. In other words, the small-penalty clock construction allows us to prove that

the Hamiltonians resulting from the construction inherit a spectral gap related to the gap in accept probabilities in the circuit, for several variants of QMA. This ability is used in the proofs of Theorems 2 to 6 below. An interesting feature of the modified clock construction is that it also allows us to show that, when there is a classical witness (i.e., a QCMA computation), the resulting Hamiltonian has a classical description for a state with energy close to the ground-state energy. Another related application of the small-penalty clock construction is that it also allows us to show complexity lower bounds like in Lemmas 4 and 6 below. In these cases, we directly reduce from PP to the appropriate gapped version of the LOCALHAMILTONIAN problem instead of a reduction from the corresponding -QMA class.

We now spell out what enables the small-penalty clock construction to show the above results. As mentioned before, the clock construction and its variants encompass all current proofs of hardness for QMA and related classes. Typically, this consists of mapping a circuit to a Hamiltonian $H = H_{\text{input}} + H_{\text{prop}} + H_{\text{clock}} + H_{\text{output}}$. Roughly speaking, each term locally enforces that the computation is a valid step of a QMA protocol by adding energy penalties to undesirable states. The "witness register," where a quantum prover may input any quantum state, is left unpenalized and the Hamiltonian therefore has no terms acting on the witness register. The role of H_{output} is to ensure that witnesses and computations that lead to a *low* accept probability at the output get a *high* energy penalty. In the absence of the penalty term at the output, the ground-state space of the Hamiltonian is well known and is given by the subspace of the so-called "history states," each with the same energy. The output penalty term H_{output} is what breaks the degeneracy and helps create a promise gap, and we henceforth refer to this as simply the penalty term without qualification.

However, the addition of the penalty term makes the eigenstates of the Hamiltonian difficult to analyze, since the magnitude of the penalty can be large, i.e., $\Omega(1)$ in strength. In this work, we often choose the output penalty terms to have small strength. This might seem like a strange choice to make since one is typically interested in making the promise gap as large as possible. However, since we are dealing with instances where the promise gap is already exponentially small, our choice is not too costly. The advantage this gives us is that the ground-state energy tracks the effect of the output penalty more faithfully. More concretely, the smallness of the penalty term allows us to use tools like the Schrieffer-Wolff transformation [53,54], which can be viewed as a rigorous formulation of degenerate perturbation theory. We review the Schrieffer-Wolff transformation in Appendix A.

Spectral gap in the adjacency matrix.—For the proof of Theorem 10 below, we show a reduction [55] from a natural PSPACE-complete graph problem to an instance of a problem known as $(1/\text{exp}, 1/\text{exp})$-SPARSEHAMILTONIAN [56]. This problem is a generalization of $(1/\text{exp}, 1/\text{exp})$-$k$-LOCALHAMILTONIAN, allowing for the Hamiltonian to be any sparse Hamiltonian with a spectral gap $\geq 1/\text{exp}$. Sparse Hamiltonians are Hermitian matrices that can be exponentially large, with at most $\text{poly}(n)$ nonzero entries per row in some basis and an efficient algorithm for computing any entry of the matrix. They are a generalization of local Hamiltonians.

The PSPACE-complete graph problem may be described as SUCCINCTGRAPHREACHABILITY, which is a decision problem about whether there is a path from one vertex to another in a succinctly described graph of exponential size (also see Ref. [27]). We show that one can always construct a PSPACE-bounded Turing machine such that the resulting Hamiltonian after the reduction always has a spectral gap that is at least $1/\text{exp}(n)$. We do this through an explicit analysis of the eigenvalues of the Hamiltonian, which are related to the lengths of cycles and paths of the graph constructed from the Turing machine. Next, we give a PreciseEGQMA upper bound to $(1/\text{exp}, 1/\text{exp})$-SPARSEHAMILTONIAN, i.e., the problem in the presence of a spectral gap, establishing that PSPACE $\subseteq$ PreciseEGQMA.

D. Discussion

Our first main result was that the addition of even an inverse-polynomially small spectral gap takes the complexity of precisely estimating the ground-state energy of a local Hamiltonian from PreciseQMA = PSPACE to PrecisePGQMA = PP. Note that this result also implies a difference between the case of no spectral gap and a constant spectral gap. Therefore, we have given a *provable setting* where the difference in complexity between two problems is attributable entirely to the spectral gap.

Our second main result concerned a modification of the same problem of precisely estimating the ground-state energy of a local Hamiltonian promised to have an inverse-polynomial spectral gap. When additionally promised that there exists a classical description of a circuit to prepare a state whose energy is exponentially close to the ground-state energy, our results show that the complexity of the problem does not get weaker. Specifically, we show that the class PrecisePGQCMA is equivalent to PrecisePGQMA.

The above equivalence result is in sharp contrast with the belief PreciseQCMA $\neq$ PreciseQMA in the non-spectral-gapped case. This inequality follows from the conjecture that $\text{NP}^{\text{PP}} \neq \text{PSPACE}$, which, if false, would lead to a collapse of the counting hierarchy. The inequality PreciseQCMA $\neq$ PreciseQMA rules out the possibility of there being polynomial-size circuits to prepare ground states of local Hamiltonians to exponential precision, since otherwise the prover could simply

supply a description of such a circuit. Our equivalence result that PrecisePGQMA = PrecisePGQCMA is consistent with the following intriguing conjecture about the circuit-complexity of ground states of low-energy Hamiltonians, although it does not imply the conjecture.

Conjecture 1: *Consider any Hamiltonian H on n qubits with ground-state energy E_1 and a $1/\mathsf{poly}$ spectral gap. Then there exists a low-energy state $|\psi\rangle$ satisfying $\langle\psi|H|\psi\rangle \leq E_1 + 2^{-\mathsf{poly}(n)}$ that can be prepared by an efficient quantum circuit, namely a state of the form $|\psi\rangle = U|0\rangle^m$, where m and the size of U are both polynomials in n.*

Note that Conjecture 1 implies the following results: (i) PrecisePGQMA = PrecisePGQCMA, (ii) PGQMA = PGQCMA, and (iii) PGQCMA = QCMA. We prove (i) and give strong evidence for (iii) in Corollary 2 below by showing that PGQCMA $=_R$ QCMA. These results do not imply Conjecture 1 because the reductions do not imply anything about the classical witnesses. We also note that the quantum circuits referred to in Conjecture 1 may be hard to find—the conjecture is only concerned with the existence of such circuits, and not with whether these circuits can be obtained by an efficient algorithm. In complexity-theoretic language, these circuits may be nonuniform. This is why Conjecture 1 is not in contradiction with Ref. [57], which argues that finding efficient matrix-product-state representations of Hamiltonians with a $\Omega(1/\mathsf{poly})$ spectral gap can be hard.

If Conjecture 1 were true, it would also explain the observed success of quantum algorithms such as the variational quantum eigensolver (VQE) [58,59], which seek to solve a much simpler problem of preparing low-energy states of translation-invariant many-body Hamiltonians with energy $1/\mathsf{poly}$ close to the ground-state energy. A large class of translation-invariant Hamiltonians have a spectral gap that is either a constant, $\Theta(1)$ (gapped phases), or vanishing in the system size as $\Theta(1/n^{1/D})$ (gapless phases described by conformal field theories in D dimensions). Therefore, Conjecture 1 applies to both these cases and would imply the existence of polynomial-size circuits to prepare states with high overlap with the ground state. Such circuits are generally found in the VQE algorithm if one optimizes over sufficiently many parameters. This behavior is in line with other instances where a lower bound on the spectral gap implies tractability of the ground state in various senses [13,15,60,61].

Coming to the case of exponentially small spectral gaps, we have shown that PreciseEGQMA = PreciseQMA. This implies that PreciseEGQMA ≠ PreciseEGQCMA unless the counting hierarchy collapses. Therefore, we give a class of local Hamiltonians (in the proof of Lemma 12 below) with exponentially small spectral gaps, whose ground states have exponentially large circuit complexity. This is a result of independent interest, and it might be interesting to study whether these Hamiltonians can be classified as quantum spin glasses, which are believed to be hard to cool down to zero temperature [62].

In another intriguing line of work, Aharonov and Green [38] and Green *et al.* [39] have given interactive protocols for precise quantum complexity classes with a computationally bounded prover $\mathcal{P}$ and a computationally bounded verifier $\mathcal{V}$, denoted $\mathsf{IP}[\mathcal{P},\mathcal{V}]$. A goal of this line of work is to give a quantum-inspired proof of the result IP = PSPACE [40] by giving an interactive protocol for PreciseQMA [39] (which equals PSPACE) with a BPP verifier. This has been successful so far with PreciseBQP and PreciseQCMA (which equals $\mathsf{NP}^{\mathsf{PP}}$) but not yet with PreciseQMA. From the result of Ref. [38] and our result that PrecisePGQMA = PP, there is an IP[PreciseBQP, BPP] protocol for PrecisePGQMA. Our results indicate that the spectral gap might play an important role in extending such an interactive protocol to PSPACE. Namely, such an extension would need to be able to work with inverse-exponentially small spectral gaps.

In addition, the class postQMA [35,63] is the class where there is a quantum prover and a postBQP verifier, where one may condition (postselect) on exponentially unlikely outcomes. This class has been shown to be equal to PreciseQMA [35], so an alternative approach mentioned by Green *et al.* [39] to reprove the result IP = PSPACE is to exhibit an IP[postQMA, BPP] protocol for postQMA. To complete such a proof, it would suffice to prove a witness-preserving amplification technique like in QMA [31,64] that additionally handles postselection. Witness-preserving amplification is a technique for improving the promise gap of an interactive protocol by modifying the verifier's strategy while keeping the witness fixed. We show in Lemma 9 below that, assuming that PP ≠ PSPACE, the soundness of a postQMA protocol cannot be reduced beyond a particular point without requiring the witness to grow larger or requiring the postselection success probability to shrink. Therefore, we obtain evidence that a witness-preserving amplification technique for postQMA should differ significantly from the technique of Marriott and Watrous [31], since in the latter, repeating the verifier's circuit suffices to get any soundness parameter $s \leq 2^{-\mathsf{poly}}$.

So far, we have considered the spectral-gap promise to be applicable to both YES and NO instances of the problems defined. We can also define asymmetric problems where only the YES instances are promised to have a spectral gap. The motivation for considering such asymmetric promises is that they are related to complexity classes where the accepting witness is promised to be unique, such as the class UQMA [16]. The problems with asymmetric promises can only be harder than their

symmetric analogues, since the promise is weaker. We show that, for both $\Omega(1/\mathsf{poly})$ and $\Omega(1/\mathsf{exp})$ spectral gaps in the precise setting, there is no difference between symmetric and asymmetric promises on the spectral gaps. Specifically, we show in Theorem 11 below that the classes with asymmetric promises are of the same complexity as those with symmetric promises.

We remark here that the promise of a spectral gap above a unique ground state is distinct from assuming that we have a UQMA instance. The reason is that, for LocalHamiltonian, the presence of a spectral gap does not imply that there is a unique accepting witness, it only implies a unique ground state. In case the ground-state subspace is polynomially degenerate, the PP algorithm continues to work to produce estimates of the ground-state energy.

Lastly, we add that results shown in the precise regime do not always imply analogous results in the non-precise regime. For example, our work gives evidence that PrecisePGQCMA $\neq$ PreciseQCMA, but in the non-precise regime we can show that PGQCMA $=_R$ QCMA. In this respect, inequivalence results in the high-precision regime resemble oracle separation results in complexity theory, which is a mature area of research with several important results [65–67]. While oracle separations do not constitute strong evidence for the inequivalence of two complexity classes, they are useful in ruling out proof techniques that work relative to oracles, or "relativize." Similarly, inequivalence results in the precise regime can rule out proof techniques from extending to the precise regime. For example, a purported proof that QCMA $=$ QMA must not work in the precise regime; otherwise, we would obtain PreciseQCMA $=$ PreciseQMA, or PSPACE $=$ PP, which is believed to be unlikely.

E. Related work

The study of Hamiltonian complexity [1,44,45,68–73] has given rise to many techniques and important results applicable in quantum many-body physics, such as Refs. [12,57,74–80]. The clock construction has also been analyzed in detail recently [81–83].

The study of exponentially small promise gaps in the context of quantum classes can be traced to Watrous [84], who defined PQP and showed its equivalence with postBQP, which equals PP [49]. In the precise setting, one can sometimes give far stronger evidence for the (in)equivalence of complexity classes than in the analogous bounded error setting, as is the case for precise versions of the questions of QCMA versus QMA [28] and QMA(2) versus QMA [28,85–87]. There has been work on quantum interactive proof systems with exponentially small promise gaps, such as in the context of QMA(2) [87], or with even smaller gaps, such as in Refs. [88–90]. Fefferman and Lin [28,91] studied the precise regime of QMA, showing it to equal PSPACE, leading to other works concerning precise classes [35,92]. Gharibian *et al.* [34] considered quantum generalizations of the polynomial hierarchy, where precise classes and spectral gaps are relevant to the definitions and proof techniques.

Aharonov *et al.* [16] were the first to consider the complexity of the LocalHamiltonian problem in the presence of spectral gaps, motivated by the question of uniqueness [93] for randomized and quantum classes. They showed the equivalence of UQCMA and QCMA, and that of UQMA and PGQMA, using similar techniques as Valiant and Vazirani [93] in their proof of equivalence of UNP and NP. Jain *et al.* [17] defined the class FewQMA and showed that it is contained in $\mathsf{P}^{\mathsf{UQMA}}$, giving a technique to reduce the dimension of accepting witnesses.

More recently, González-Guillén and Cubitt [18] studied the spectral gap of a large class of Hamiltonians that encode history states in their ground state and showed that the spectral gap is upper bounded by $O(1/\mathsf{poly})$. A similar result was obtained by Crosson and Bowen [19] using different techniques. These works are mainly concerned with the existence of a $\Theta(1)$ spectral gap, whereas our results distinguish between $1/\mathsf{poly}$ and $1/\mathsf{exp}$ spectral gaps.

Finally, Ambainis [94] studied the problem of estimating spectral gaps and local observables and gave a $\mathsf{P}^{\mathsf{QMA[log]}}$ upper bound for these problems, while also giving $\mathsf{P}^{\mathsf{QMA[log]}}$-hardness results (also see Ref. [95]). The class $\mathsf{P}^{\mathsf{QMA[log]}}$ is the class of problems solvable in polynomial time by making logarithmically many (adaptive) queries to a QMA oracle. Gharibian and Yirka [95] showed that $\mathsf{P}^{\mathsf{QMA[log]}} \subseteq \mathsf{PP}$ and extended previous hardness results to more natural Hamiltonians. Gharibian *et al.* [96] also gave a very natural complete problem for the class $\mathsf{P}^{\mathsf{QMA[log]}}$ in the context of computing local observables in ground states. Novo *et al.* [97] have recently studied the closely related problem of sampling from the distribution obtained by making energy measurements and obtained various interesting hardness results, under different notions of error.

II. DEFINITIONS AND COMPLETE PROBLEMS

We have seen the definition of BQP in terms of the class BQP$[c,s]$ with general parameters c and s. The Precise-version of BQP can be defined similarly.

Definition 4: We define PreciseBQP $= \bigcup_{c-s\geq 1/\mathsf{exp}}$ BQP$[c,s]$.

This class is known to be equal to PP (see, e.g., Ref. [34]).

We now give an equivalent definition of QMA in terms of the eigenvalues of an operator called the *accept operator*. We then define a very general class called Gapped QMA, GQMA$[c,s,g_1,g_2]$, which has several

parameters. By specifying these parameters, we can define the major complexity classes in this work. The complexity classes corresponding to classical witnesses (QCMA and its derivatives) are defined analogously.

The alternative definition of QMA is in terms of the "accept operator" $Q(U_x) = \langle 0|^{\otimes m} U_x^\dagger \Pi_{\text{out}} U_x |0\rangle^{\otimes m}$ on the witness register, where Π_{out} is the projector on to the accept state ($|1\rangle_o$). For any state $|\Psi\rangle$ provided as a witness, quantity $\langle\Psi|Q_x|\Psi\rangle$ is the accept probability of the circuit. We henceforth suppress the dependence of Q on the unitary U_x and the instance x. The eigenvalues of Q, $\lambda_1(Q) \geq \lambda_2(Q) \geq \cdots$, are important quantities to consider since the accept probability of any input proof state is a convex combination of these eigenvalues. The alternative definition of QMA in terms of the operator Q is as follows.

Definition 5 (**Alternative definition of** QMA$[c,s]$)**:** We say that $A = (A_{\text{yes}}, A_{\text{no}})$ is a QMA$[c,s]$ problem if and only if, for every instance x, there exists a uniformly generated circuit U_x of size poly(n) acting on $m + w = \text{poly}(n)$ qubits, with the properties that

(a) if $x \in A_{\text{yes}}$, then $\lambda_1(Q) \geq c$,
(b) if $x \in A_{\text{no}}$, then $\lambda_1(Q) \leq s$,

where $Q = Q(U_x)$ is as above.

Note that we are typically interested in the behavior of the maximum accept probability, which equals the largest eigenvalue of Q. We are also interested in the lowest eigenvalue of a Hamiltonian H for the LocalHamiltonian problem and its variants. Therefore, we order eigenvalues in nonincreasing order for accept operators and in nondecreasing order for Hamiltonians. For the same reason, we define the spectral gap differently for accept operators and Hamiltonians. For a Hamiltonian, we define the spectral gap to be the difference in the smallest two eigenvalues $E_2 - E_1$. For accept operators, the spectral gap is the difference between the *highest two eigenvalues* $\lambda_1(Q) - \lambda_2(Q)$. This is equal to the difference in the accept probabilities of the optimal witness and the next-optimal witness orthogonal to it. It will usually be clear from the context which spectral gap we are referring to.

Now let us define the class GQMA$[c,s,g_1,g_2]$. It corresponds to a promise on the operator Q having a spectral gap of at least g_1 in the YES case, and at least g_2 in the NO case.

Definition 6 (Gapped QMA)**:** Class GQMA$[c,s,g_1,g_2]$ is the class of promise problems $A = (A_{\text{yes}}, A_{\text{no}})$ such that, for every instance x, there exists a polynomial-size verifier circuit U_x acting on poly(n) qubits and its associated accept operator Q such that

(a) if $x \in A_{\text{yes}}$, then $\lambda_1(Q) \geq c$ and $\lambda_1(Q) - \lambda_2(Q) \geq g_1$.
(b) if $x \in A_{\text{no}}$, then $\lambda_1(Q) \leq s$ and $\lambda_1(Q) - \lambda_2(Q) \geq g_2$.

This definition is a generalization of the class PGQMA defined by Aharonov *et al.* in Ref. [16].

Definition 7: We define $\text{PGQMA} = \bigcup_{c-s,g_1,g_2 \geq 1/\text{poly}} \text{GQMA}[c,s,g_1,g_2]$.

To see the relation of this class with QMA, note that, by setting $g_1 = g_2 = 0$, the promise on spectral gaps becomes vacuous, since $\lambda_1(Q) \geq \lambda_2(Q)$ by definition. Therefore, we get the equality $\text{GQMA}[c,s,0,0] = \text{QMA}[c,s]$.

Definition 8 (**Exponentially gapped** QMA)**:** We define

$$\text{EGQMA} = \bigcup_{\substack{c-s \geq 1/\text{poly} \\ g_1,g_2 \geq 1/\text{exp}}} \text{GQMA}[c,s,g_1,g_2].$$

We now come to precise versions of these classes, where the completeness-soundness gap c-s can be exponentially small, giving us more powerful classes. The first of these is PreciseQMA, which was defined in Ref. [28] and shown to be equal to PSPACE.

Definition 9: We define $\text{PreciseQMA} = \bigcup_{c-s \geq 1/\text{exp}} \text{QMA}[c,s]$.

This definition should be compared to the precise version of GQMA, which comes in two varieties: the spectral gaps can either be polynomially small (PrecisePGQMA) or exponentially small (PreciseEGQMA).

Definition 10 (PrecisePGQMA)**:** The class Precise PGQMA, short for precise polynomially gapped QMA, is the class with exponentially small promise gaps and polynomially small spectral gaps:

$$\text{PrecisePGQMA} = \bigcup_{\substack{c-s \geq 1/\text{exp} \\ g_1,g_2 \geq 1/\text{poly}}} \text{GQMA}[c,s,g_1,g_2].$$

Definition 11 (PreciseEGQMA)**:** The class Precise EGQMA, short for precise exponentially gapped QMA, has both the promise gap and spectral gap exponentially small:

$$\text{PreciseEGQMA} = \bigcup_{\substack{c-s \geq 1/\text{exp} \\ g_1,g_2 \geq 1/\text{exp}}} \text{GQMA}[c,s,g_1,g_2].$$

We now come to complexity classes in which the prover sends a classical witness but the verifier remains quantum. The classicality of the witness can be enforced by measuring the qubits sent by the prover in the computational basis and interpreting qubits in the computational basis as classical bits. If the verifier is only allowed to make measurements at the end, we use the standard protocol for deferring measurements: we apply a "copy operation" U_c that has controlled-NOT gates from the qubits in the witness register to an ancilla register in state $|0\rangle^w$. We leave the qubits in the witness state unmeasured. This modified

circuit has the property that it preserves the accept probabilities of input witness states that are in the computational basis. Furthermore, the eigenstates of the modified accept operator acting on the register can be taken to be computational basis states. This allows us to define QCMA and its derivatives in terms of the accept operator and also allows us to consider a gapped version of QCMA.

Definition 12 (GQCMA$[c,s,g_1,g_2]$)**:** We say that $A=(A_{\text{yes}},A_{\text{no}})$ is a GQCMA$[c,s]$ problem if and only if, for every instance x, there exists a uniformly generated circuit U_x of size poly(n) acting on $m+w=\text{poly}(n)$ qubits, with the properties that

(a) if $x\in A_{\text{yes}}$, then $\lambda_1(Q)\geq c$ and $\lambda_1(Q)-\lambda_2(Q)\geq g_1$.
(b) if $x\in A_{\text{no}}$, then $\lambda_1(Q)\leq s$ and $\lambda_1(Q)-\lambda_2(Q)\geq g_2$.

where $Q=Q(U_xU_c)$ is the accept operator of the modified circuit with the copy operation U_c described above.

Definition 13: The derived classes of GQCMA are given by

(a) QCMA$[c,s]=$ GQCMA$[c,s,0,0]$,
(b) QCMA $=\bigcup_{c-s>1/\text{poly}}$ QCMA$[c,s]$,
(c) PreciseQCMA $=\bigcup_{c-s>1/\text{exp}}$ QCMA$[c,s]$,
(d) PGQCMA $=\bigcup_{c-s>1/\text{poly},g_1,g_2>1/\text{poly}}$ GQCMA$[c,s,g_1,g_2]$,
(e) PrecisePGQCMA $=\bigcup_{c-s>1/\text{exp},g_1,g_2>1/\text{poly}}$ GQCMA$[c,s,g_1,g_2]$,
(f) EGQCMA $=\bigcup_{c-s>1/\text{poly},g_1,g_2>1/\text{exp}}$ GQCMA $[c,s,g_1,g_2]$,
(g) PreciseEGQCMA $=\bigcup_{c-s>1/\text{exp},g_1,g_2>1/\text{exp}}$ GQCMA$[c,s,g_1,g_2]$.

A. Complete problems

We now come to the definitions of problems that are complete for these classes. The classic problem complete for the class QMA is the LocalHamiltonian problem [2,44,45]. We define a k-local observable to be a Hermitian operator A that can be written as a sum over operators A_i supported on k qudits at most: $A=\sum_i^{\text{poly}(n)}A_i$. We assume that each term has bounded operator norm $\|A_i\|\leq\text{poly}(n)$. The task in the LocalHamiltonian problem is to estimate the ground-state energy of a local Hamiltonian. The decision version of the problem is as follows.

k-LocalHamiltonian$[a,b]$

Input: A description of a k-local Hamiltonian $H=\sum_i h_i$ on n qubits with $h_i\geq 0$, two numbers a and b with $b>a$.
Output: YES if the ground-state energy $E_1\leq a$,
NO if $E_1\geq b$, promised that one of them is the case.

Henceforth, we omit the phrase "promised that one of them is the case" because we will be exclusively considering promise problems unless otherwise specified. Kitaev [2] showed that 5-LocalHamiltonian$[A,B]$ with $b-a=\Omega(1/\text{poly})$ is QMA-complete, which was improved to $k=3$ and then $k=2$ in Refs. [44,45]. The parameter $\delta:=b-a$, the promise gap, is a measure of the accuracy to which the solution is desired. We define the problem in terms of δ only, as follows.

Definition 14: We define δ-k-LocalHamiltonian $:=\bigcup_{b-a\geq\delta}k$-LocalHamiltonian$[A,B]$.

We now come to the gapped and precise versions of the problem, which turn out to be complete for their respective -QMA variants. We also suppress the notation k in the name of the problem, though there is formally a dependence on k. In this work, our hardness results hold for $k\geq 3$ and it may be possible to improve our results to hold for $k=2$.

LocalHamiltonian$[a,b,g_1,g_2]$

Input: Description of a k-local Hamiltonian $H=\sum_i h_i$ with $h_i\geq 0$, numbers a, b, g_1, and g_2 with $b>a$.
Output: YES if the ground-state energy $E_1\leq a$ and any state orthogonal to the ground state has energy $\geq E_1+g_1$,
NO if $E_1\geq b$ and any state orthogonal to the ground state has energy $\geq E_1+g_2$.

In both the YES and NO cases above, we see that the Hamiltonian has a unique ground state and a spectral gap of at least g_1 in the YES case and g_2 in the NO case. The above problem with promise gap $\delta=b-a$ and spectral gap $\Delta=\min[g_1,g_2]$ is defined as follows.

Definition 15: We define

$$(\delta,\Delta)-\text{LocalHamiltonian}:=\bigcup_{\substack{b-a\geq\delta\\ g_1,g_2>\Delta}}\text{LocalHamiltonian}[a,b,g_1,g_2].$$

In the nonprecise regime, the problem $(1/\text{poly},1/\text{poly})$-LocalHamiltonian was shown to be complete for PGQMA for $k\geq 2$ [16].

We now focus on the precise regime, i.e., $\delta=\Omega(1/\text{exp})$. From the results of Ref. [28], we know that $(1/\text{exp},0)$-LocalHamiltonian is PreciseQMA-complete for $k\geq 3$. We have the following results.

Theorem 1: *It holds that* $(1/\text{exp},1/\text{poly})$-Local Hamiltonian *is* PrecisePGQMA*-complete.*

Theorem 2: *It holds that* $(1/\text{exp},1/\text{exp})$-Local Hamiltonian *is* PreciseEGQMA*-complete.*

By virtue of these theorems, we can talk about the complexity of the classes PrecisePGQMA and PreciseEGQMA interchangeably with their complete problems. The proofs of these theorems are given in Appendices B and C. The hardness results rely on the small-penalty clock construction, where the size of the penalty term is either $\Theta(1/\mathsf{poly})$ or $\Theta(1/\mathsf{exp})$. The upper bounds are shown in Lemmas 15 and 16 below and rely on a modification of the standard phase-estimation protocol used to show that k-LocalHamiltonian is in QMA. Specifically, we consider the modified protocol of Ref. [28] used for Precise-k-LocalHamiltonian and observe that the spectral gaps in the energies translate to separations in the accept probabilities.

Finally, we turn to complete problems for QCMA and its derivatives. The first problem, GS-Description-LocalHamiltonian, concerns finding the ground-state energy of a k-local Hamiltonian when there is a polynomial-size circuit to prepare a state close to the ground state (which constitutes a classical description of the ground state).

GS-Description-LocalHamiltonian$[a, b, g_1, g_2]$

Input: Description of a k-local Hamiltonian $H = \sum_i h_i$, numbers a, $b \geq a + \delta$, polynomials $T(n), m(n)$, together with the promise that there exists a circuit V of size T such that $V|0^m\rangle = |\psi\rangle$ satisfies $\langle\psi|H|\psi\rangle \leq E_1 + \delta^3/f(n)^2$ for some polynomial $f(n) \geq \|H\|$.

Output: YES if the ground-state energy of H satisfies $E_1 \leq a$ and the spectral gap of H is at least g_1, NO if $E_1 \geq b$ and the spectral gap of H is at least g_2.

Definition 16: We define

$$(\delta, \Delta)\text{-GS-Description-LocalHamiltonian} := \bigcup_{\substack{b-a\geq\delta \\ g_1,g_2\geq\Delta}} \text{GS-Description-LocalHamiltonian} \times [a, b, g_1, g_2].$$

As in the case of (δ, Δ)-LocalHamiltonian, if we take $\Delta = 0$, we get a version without any promise on the spectral gap. This is a close relative of the following problem proved to be QCMA-complete for $\delta = \Omega(1/\mathsf{poly})$ [98].

δ-LowComplexity-LowEnergyStates

Input: Description of a k-local Hamiltonian $H = \sum_i h_i$, numbers a, b, and polynomials $T(n)$, $m(n)$, with $b \geq a + \delta$.

Output: Output: YES if there exists a circuit of size $\leq T(n)$ that acts on $|0^m\rangle$ to prepare a state $|\psi\rangle$ with energy $\langle\psi|H|\psi\rangle \leq a$,
NO if any state $|\psi\rangle$ obtained by applying a circuit of size $T(n)$ on $|0^m\rangle$ has energy $\langle\psi|H|\psi\rangle \geq b$.

This latter problem has a weaker promise than $(\delta, 0)$-GS-Description-LocalHamiltonian. This is because a NO instance of δ-LowComplexity-LowEnergystates is automatically a NO instance of $(\delta, 0)$-GS-Description-LocalHamiltonian, since any state necessarily has energy $\geq b$. Meanwhile, a NO instance of $(\delta, 0)$-GS-Description-LocalHamiltonian need not be a NO instance of δ-LowComplexity-LowEnergystates, since, for the latter, there is no guarantee of a circuit to prepare a state with energy close to the ground-state energy.

Despite having a stronger promise on $(\delta, 0)$-GS-Description-LocalHamiltonian (which only makes the problem less complex), our small-penalty clock construction allows us to prove the same hardness result for both $\delta = 1/\mathsf{poly}$ and $\delta = 1/\mathsf{exp}$.

Theorem 3: *It holds that* $(1/\mathsf{poly}, 0)$-GS-Description-LocalHamiltonian *is* QCMA*-complete.*

Theorem 4: *It holds that* $(1/\mathsf{exp}, 0)$-GS-Description-LocalHamiltonian *is* PreciseQCMA*-complete.*

For the latter theorem in the precise regime, we use the small-penalty clock construction with an exponentially small energy penalty. Lastly, when we add the promise of spectral gaps, we have the following results.

Theorem 5: *It holds that* $(1/\mathsf{poly}, 1/\mathsf{poly})$-GS-Description-LocalHamiltonian *is* PGQCMA*-complete.*

Theorem 6: *It holds that* $(1/\mathsf{exp}, 1/\mathsf{exp})$-GS-Description-LocalHamiltonian *is* PreciseEGQCMA*-complete.*

Theorem 7: *It holds that* $(1/\mathsf{exp}, 1/\mathsf{poly})$-GS-Description-LocalHamiltonian *is* PrecisePGQCMA*-complete.*

The upper bounds in Theorems 3 to 7 follow from a precise version of phase estimation, together with the promise that there is a classical description of a circuit to prepare a low-energy state. The lower bounds either follow directly through a small-penalty clock construction or through a reduction from a class that contains the relevant class.

III. PROBLEMS CHARACTERIZED BY PP

In this section, we discuss the complexity of the classes PrecisePGQMA and PrecisePGQCMA, both of which turn out to equal PP.

Theorem 8: *It holds that* PrecisePGQMA = PP.

Theorem 9: *It holds that* PrecisePGQCMA = PP.

We describe here the overall strategy for proving these results. First, we adapt the one-bit phase estimation circuit in Ref. [28] to show that it is possible to compute ground-state energies of sparse Hamiltonians with a spectral gap in the corresponding GQMA class. In particular, we have the following result.

Lemma 1: *It holds that* $(1/\mathsf{exp}, 1/\mathsf{poly})$-LOCAL-HAMILTONIAN $\in$ PrecisePGQMA.

Next, we use the "power method" [46] to give a PP algorithm for any problem in PrecisePGQMA.

Lemma 2 (One half of Theorem 8): *It holds that* PrecisePGQMA $\subseteq$ PP.

Proof. Suppose that we have a GQMA$[c, s, g_1, g_2]$ instance. Then we should give a PP algorithm to precisely compute the maximum eigenvalue λ_1 of the accept operator Q associated with the instance, under the promise that the spectral gap of Q is bounded below by an inverse polynomial. In particular, the spectral gap of the accept operator, given by $\lambda_1 - \lambda_2$, is at least $\min[g_1, g_2] =: \Delta$. Consider the power method to compute the maximum eigenvalue and eigenvector of a positive semidefinite operator Q. This method relies on the observation that, upon taking positive powers of the operator Q and estimating its trace, the quantity is dominated by the maximum eigenvalue of Q. In the following, we suppress the dependence of λ_i on Q:

$$\begin{aligned}\mathrm{Tr}(Q^q) &= \sum_i \lambda_i^q \\ &= \lambda_1^q \left(1 + \left(\frac{\lambda_2}{\lambda_1}\right)^q + \cdots\right) \\ &\leq \lambda_1^q + \lambda_1^q (2^w - 1)\left(1 - \frac{\Delta}{\lambda_1}\right)^q. \end{aligned} \tag{1}$$

Here w is the size of the witness register. On the other hand, we have $\mathrm{Tr}(Q^q) \geq \lambda_1^q$. Therefore, in the YES case, we have

$$\mathrm{Tr}(Q^q) \geq c^q, \tag{2}$$

while in the NO case,

$$\mathrm{Tr}(Q^q) \leq s^q + s^q (2^w - 1)\left(1 - \frac{\Delta}{\lambda_1}\right)^q. \tag{3}$$

By the promise of the spectral gap, we must have $\lambda_1 \geq \Delta$, since otherwise the second largest eigenvalue of Q would be $\lambda_2 < 0$. The difference in the two cases is

$$\begin{aligned} &c^q - s^q - s^q (2^w - 1)\left(1 - \frac{\Delta}{\lambda_1}\right)^q \\ &\quad = c^q - s^q - s^q (2^w - 1)\mathsf{exp}\left[q \log\left(1 - \frac{\Delta}{\lambda_1}\right)\right] \\ &\quad \geq c^q - s^q - s^q 2^w \mathsf{exp}\left[-\frac{q\Delta}{\lambda_1}\right] \\ &\quad = s^q \left(\left(1 + \frac{c-s}{s}\right)^q - 1\right) - s^q 2^w \mathsf{exp}\left[-\frac{q\Delta}{\lambda_1}\right] \\ &\quad \geq s^q \left(\frac{q(c-s)}{s} - 2^w \mathsf{exp}\left[-\frac{q\Delta}{\lambda_1}\right]\right) \\ &\quad \geq s^q \left(c - s - 2^w \mathsf{exp}\left[-\frac{q\Delta}{\lambda_1}\right]\right), \end{aligned} \tag{4}$$

since $q \geq 1$ and $s \leq 1$. If we pick $q = \lceil (\lambda_1/\Delta) \log (2^{w+1}/(c-s)) \rceil = O(\mathsf{poly})$, we can ensure that the term $2^w \mathsf{exp}[-q\Delta/\lambda_1]$ is at most $(c-s)/2$. Thus, the difference in $\mathrm{Tr}(Q^q)$ between the YES and NO cases is at least

$$s^q \frac{c-s}{2} = \Omega(2^{-\mathsf{poly}}). \tag{5}$$

In the last line above we assumed that $s^q \geq c^q / 2^{w+1} = \Omega(2^{-\mathsf{poly}})$ for some polynomial. In case this assumption is not true, we would nevertheless still have a difference of at least $c^q - 2^w s^q > c^q/2 \geq \Omega(2^{-\mathsf{poly}})$ in between the YES and NO cases when measuring $\mathrm{Tr}(Q^q)$.

This observation suggests that a PP algorithm can decide between the YES and NO cases by computing $\mathrm{Tr}(Q^q)$ for some large enough polynomial q. This is possible because a PP algorithm can compute a sum of 2^{poly} terms, where every term is efficiently computable in polynomial time. We prove this in Appendix E (Lemma 22). ■

The above result implies that, since $(1/\mathsf{exp}, 1/\mathsf{poly})$-LOCALHAMILTONIAN is in PrecisePGQMA, a PP algorithm can precisely compute ground-state energies of local Hamiltonians with a $\Omega(1/\mathsf{poly})$ spectral gap. A similar technique can also be used to show a slightly more general result.

Lemma 3: *Given a local Hamiltonian H and a local observable A, along with a promise that $\|A\| = O(\mathsf{poly})$ and the spectral gap of H is lower bounded by $\Omega(1/\mathsf{poly})$,*

a P^{PP} *algorithm can decide if the ground-state local observable* $\langle E_1|A|E_1\rangle$ *is either* $\leq a$ *or* $\geq b$*, for* $b-a=\Omega(2^{-\mathsf{poly}})$*, where* $|E_1\rangle$ *is the ground state of* H.

This lemma is proved in Appendix E. Note that both of these results include the case $\Delta=\Theta(1)$, the important case of constant spectral gaps.

We complete the characterization of the power of PrecisePGQMA with the following result.

Lemma 4: *It holds that* $(1/\mathsf{exp},1/\mathsf{poly})$-LOCAL-HAMILTONIAN *is* PP*-hard.*

For this proof, we use the small-penalty clock construction, albeit one for class PreciseBQP as opposed to class PreciseQMA. In this aspect, it resembles the clock construction of Aharonov *et al.* [74], where it was used to show BQP universality of the model of adiabatic quantum computing. We use the technique of applying $\Theta(1/\mathsf{poly})$ small penalties at the output so as to preserve the lower bound of $\Omega(1/\mathsf{poly})$ on the spectral gap shown in Ref. [74]. In sum, Lemmas 1, 2, and 4 together imply Theorems 1 and 8.

We now come to class PreciseQCMA and its complete problem, $(1/\mathsf{exp},0)$-GS-DESCRIPTION-LOCALHAMILTONIAN, where we are promised that there is an efficient circuit to prepare a low-energy state. We know that $\mathsf{PreciseQCMA}=\mathsf{NP}^{\mathsf{PP}}$ [35], which lies in the second level of the counting hierarchy. Since PrecisePGQMA is characterized by PP, the promise of having a spectral gap is only slightly stronger than the promise of an efficient circuit to prepare the ground state.

Consider now the gapped version of the problem, $(1/\mathsf{exp},1/\mathsf{poly})$-GS-DESCRIPTION-LOCALHAMILTONIAN, where there is a $1/\mathsf{poly}$ spectral gap in addition to the promise of an efficient circuit to prepare the ground state. This characterizes the class PrecisePGQCMA, for which the proof technique is similar to PrecisePGQMA.

We first show that the gapped version of GS-DESCRIPTION-LOCALHAMILTONIAN is in the corresponding GQCMA class, and, in particular, that the following result holds.

Lemma 5: *It holds that* $(1/\mathsf{exp},1/\mathsf{poly})$-GS-DESCRIPTION-LOCALHAMILTONIAN $\in$ PrecisePGQCMA.

The PP-hardness of the problem follows by the same argument as in the proof of Lemma 4.

Lemma 6: *It holds that* $(1/\mathsf{exp},1/\mathsf{poly})$-GS-DESCRIPTION-LOCALHAMILTONIAN *is* PP*-hard.*

We give a unified proof of Lemmas 4 and 6 in Appendix B. Since $\mathsf{PrecisePGQCMA}\subseteq\mathsf{PrecisePGQMA}=\mathsf{PP}$, this implies the following result.

Corollary 1: *It holds that* $\mathsf{PrecisePGQMA}=\mathsf{PrecisePGQCMA}=\mathsf{PP}$.

IV. PROBLEMS CHARACTERIZED BY PSPACE

In this section, we discuss the complexity of class PreciseEGQMA, which turns out to equal PSPACE. This result indicates that the complexity of the local Hamiltonian problem does not jump immediately in the presence of a tiny, nonzero spectral gap. This means that there is a notion of robustness of the complexity of the problem with respect to the spectral gap.

Theorem 10: *It holds that* $\mathsf{PreciseEGQMA}=\mathsf{PreciseQMA}\ (=\mathsf{PSPACE})$.

Proof. The containment $\mathsf{PreciseEGQMA}\subseteq\mathsf{PreciseQMA}$ follows trivially since any PreciseEGQMA instance is automatically a PreciseQMA instance. We show the other direction, $\mathsf{PreciseEGQMA}\supseteq\mathsf{PreciseQMA}$, in two steps. Our proof relies on the complexity of the following problem.

SPARSEHAMILTONIAN$[a,b,g_1,g_2]$

Input: A succinct description of a Hermitian matrix of size $2^{\mathsf{poly}(n)}\times 2^{\mathsf{poly}(n)}$, with at most $d=\mathsf{poly}(n)$ many entries in each row and two numbers a and b, with $b>a$. The magnitude of each entry is bounded by $k=\mathsf{poly}(n)$.

Output: YES if the smallest eigenvalue $E_1\leq a$ and the spectral gap of the matrix is at least g_1,
NO if $E_1\geq b$, and the spectral gap of the matrix is at least g_2.

We define (δ,Δ)-SPARSEHAMILTONIAN to be $\bigcup_{b-a\geq\delta,\,g_1,g_2\geq\Delta}$SPARSEHAMILTONIAN$[a,b,g_1,g_2]$ and consider the problem with parameters $\delta,\Delta=\Omega(1/\mathsf{exp})$. First, in Lemma 7 below, we prove that $(1/\mathsf{exp},1/\mathsf{exp})$-SPARSEHAMILTONIAN is PSPACE-hard, or, equivalently, PreciseQMA-hard. Next, we show in Lemma 8 below that $(1/\mathsf{exp},1/\mathsf{exp})$-SPARSEHAMILTONIAN may be solved in PreciseEGQMA. The theorem then follows. ■

Lemma 7: *It holds that* $(1/\mathsf{exp},1/\mathsf{exp})$-SPARSEHAMILTONIAN *is* PSPACE*-hard.*

The reduction is from any problem in PSPACE to an instance of co-$(1/\mathsf{exp},1/\mathsf{exp})$-GAPPED-SPARSEHAMILTONIAN, which is the complement of the problem, in the sense that the YES and NO instances are reversed. Since PSPACE is closed under complement, this still gives the desired hardness result. The broad idea is to represent a PSPACE computation as an exponentially large, but sparse, graph. The smallest eigenvalue of the adjacency

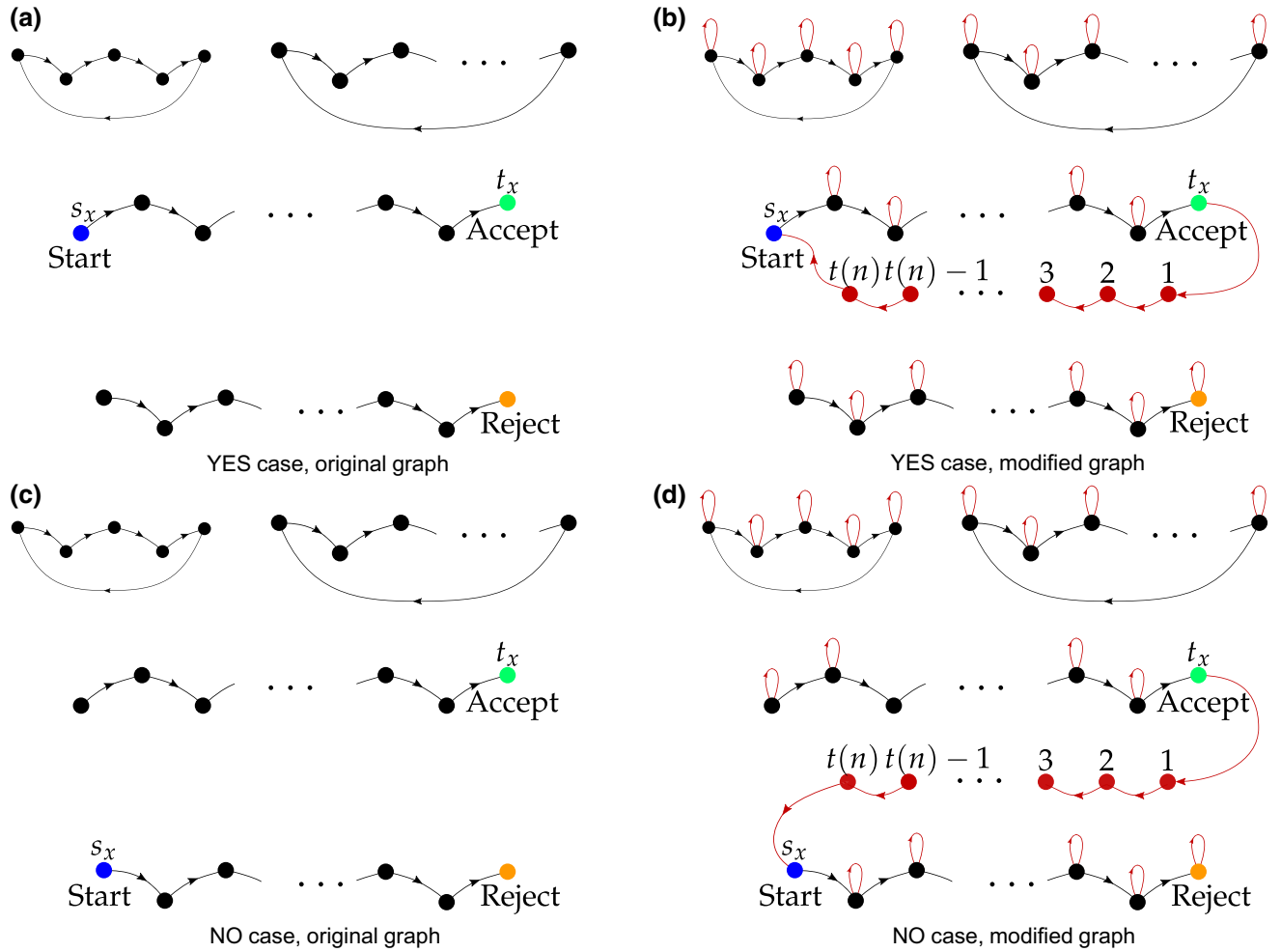

FIG. 2. Schematic of the original and modified graphs for both YES and NO cases. The original graph in both YES and NO cases consists of vertices with in degree and out degree at most 1, due to the fact that the Turing machine is reversible. The start vertex s_x is marked in blue, the accept vertex t_x in green, and the reject vertex in orange. The modified graphs have self-loops on all vertices except the start and the accept vertices. They have additional vertices $1, 2, \ldots, t(n)$ without self-loops. All modifications are in maroon. (a) YES case, original graph. (b) YES case, modified graph. (c) NO case, original graph. (d) NO case, modified graph.

matrix of this graph encodes information about whether the computation accepts or rejects.

Proof of Lemma 7. We use a proof technique adapted from an unpublished manuscript by Fefferman and Lin [91]. First, we use the fact that PSPACE with reversible operations in every step still equals PSPACE: revPSPACE = PSPACE [99]. Indeed, it is known that SPACE$[\mathfrak{s}(n)]$ = revSPACE$[\mathfrak{s}(n)]$ [100] with an overhead in time that is exponential in the space, $\mathfrak{s}(n)$. Let $t(n)$ be this upper bound on the running time of the Turing machine, so that we can restrict our attention to the class revSPACE$[\mathfrak{s}(n)] \cap$ TIME$[t(n)]$ = SPACE$[\mathfrak{s}(n)]$. Any computation on a reversible Turing machine may be viewed as traversing a directed *configuration graph*, where each vertex of the graph is determined by the state of the head and the list of symbols on the input and work tapes [Figs. 2(a) and 2(c)]. When such a Turing machine is restricted to use space polynomial in the input length n, the number of vertices in the graph is upper bounded by an exponential, $2^{\mathrm{poly}(n)}$. Consider the adjacency matrix of the graph, A_x. The description of this exponentially large matrix is succinct because it only requires specifying the input x and the rules of the Turing machine.

We modify the configuration graph $G_x \to G'_x$ so that the smallest eigenvalue of the matrix $A'^{\dagger}_x A'_x$ is 0 in the NO case and bounded away by an exponentially small amount in the YES case. We do this modification in a way that ensures that the matrix has a spectral gap lower bound of at least $\Omega(1/\mathrm{exp})$. This is done as follows. First, we modify the configuration graph of the Turing machine by adding self-loops to all vertices except for the start and accept configurations s_x and t_x. We then add a sequence of vertices $\{1, 2, \ldots, t(n)\}$ from the accept configuration t_x, with the directed edges $t_x \to 1 \to 2 \to \cdots \to t(n) \to s_x$, as shown

in Figs. 2(b) and 2(d). The adjacency matrix of this modified directed graph G'_x is A'_x, and we are interested in the eigenvalues and spectral gap of $A'^{\dagger}_x A'_x$, which is Hermitian and sparse, and also has a succinct representation.

We now analyze this construction. The proof relies on an explicit computation of the eigenvalues for the various subgraphs of the modified configuration graph. In the NO case, the graph G'_x has a path of vertices ending in the reject state [Fig. 2(d)]. This path contains the starting configuration s_x. Let ℓ be the graph distance between s_x and the reject state. Since we have added the edges $t_x \to 1 \to \cdots \to t(n) \to s_x$, these vertices and the vertices leading to the accept state are also part of the path (the Turing machine does not explore these vertices in practice). All vertices in this path except for t_x, s_x, and $i : i \in [t(n)]$ have self-loops on them. As we show in Lemma 25 below, there is a zero eigenvalue in the NO case, with a spectral gap above the zero eigenvalue. The spectral gap is lower bounded by $\Omega(1/\ell^2_{\max}) = \Omega(2^{-\mathsf{poly}})$, where $\ell_{\max}$ is the number of vertices in the longest subgraph.

In the YES case, the subgraph containing the starting vertex is a cycle, with self-loops on all vertices except for t_x, s_x, and the intermediate vertices i. In each case, the eigenvalues for any subgraph are given by $2 - 2\cos((2k-1)\pi/(2\ell+1)) = 4\sin^2((2k-1)\pi/(4\ell+2))$, $k \in [\ell]$ [91], where ℓ is the number of vertices in the subgraph. The smallest eigenvalue is therefore given by the longest subgraph and this eigenvalue is nondegenerate if no two subgraphs have the same number of vertices. This is why we have added the sequence of edges $t_x \to 1 \to \cdots \to t(n)$. The role played by these vertices is to elongate the length of the subgraph containing the start and accept configurations by $t(n)$. This ensures that no other subgraph has a length equal to the longest subgraph [since $t(n)$ is the upper bound on the total number of vertices in the graph before elongation]. Therefore, the smallest two eigenvalues are given by $4\sin^2((2k-1)\pi/(4\ell+2))$, which are separated by $\Theta(t(n)^{-2}) = \Theta(2^{-\mathsf{poly}})$.

To summarize, in the YES case we have $E_1 \geq 2^{-\mathsf{poly}}$ and $E_2 - E_1 \geq 2^{-\mathsf{poly}}$. In the NO case, we have $E_1 = 0$ and $E_2 \geq 2^{-\mathsf{poly}}$. Therefore, we have a promise gap of $2^{-\mathsf{poly}}$ and spectral gap $2^{-\mathsf{poly}}$ in both the YES and NO instances. Furthermore, the matrix $A'^{\dagger}_x A'_x$ has entries of magnitude at most 2, and is 3-sparse because of the bounded degree of the configuration graph. Since the minimum eigenvalue is small in the NO case and large in the YES case, we have a reduction to co-$(1/\mathsf{exp}, 1/\mathsf{exp})$-SPARSEHAMILTONIAN. Because of the fact that PSPACE is closed under complement, we get PSPACE-hardness of $(1/\mathsf{exp}, 1/\mathsf{exp})$-SPARSEHAMILTONIAN. ■

Lemma 8: *It holds that* $(1/\mathsf{exp}, 1/\mathsf{exp})$-SPARSEHAMILTONIAN $\in$ PreciseEGQMA.

The proof of this is mostly the same as the proof of containment of $(1/\mathsf{exp}, 1/\mathsf{exp})$-LOCALHAMILTONIAN in PreciseEGQMA and is also given in Appendix C. The only difference is that we have a sparse Hamiltonian instead of a local Hamiltonian. This distinction turns out not to matter, however, because of quantum algorithms for Hamiltonian evolution that work well with sparse Hamiltonians [21].

V. OTHER RELATED CLASSES

In this section, we discuss implications of our proof techniques for other complexity classes. The first concerns a technique for amplifying the promise gap in QMA and related classes, called in-place amplification, due to Marriott and Watrous [31]. The second is about the complexity of related classes when the spectral gap promise only applies to one kind of instance (YES instances, for example). We also complete a discussion of the results in Table I by characterizing the complexity classes PGQCMA, EGQCMA, and PreciseEGQCMA.

A. Amplification for postQMA

We first define the class postQMA.

Definition 17 (postQMA)**:** Class postQMA$[c,s]$ is the class of promise problems $A = (A_{\text{yes}}, A_{\text{no}})$ that can be decided in the following way. Apply a uniformly generated quantum circuit U of size $\mathsf{poly}(n)$ on a state $|x\rangle$ encoding the input, together with a proof state of size $w(n)$ supplied by an arbitrarily powerful prover. Postselect the first $l = \mathsf{poly}(n)$ qubits at the output onto the $|0\rangle^l$ state, and measure the first qubit of the remaining register at the output, called the decision qubit (o). The postselection probability is $\Omega(2^{-f(n)})$ for a polynomial $f(n)$.

(a) If $x \in A_{\text{yes}}$, there exists $|\psi\rangle$ such that $\Pr(o = 1) \geq c$,
(b) If $x \in A_{\text{no}}$, for all $|\psi\rangle$, $\Pr(o = 1) \leq s$.

Morimae and Nishimura [35] defined this class and showed that postQMA := postQMA$[1/3, 2/3]$ = PreciseQMA = PSPACE. This result is similar to the result postBQP = PreciseBQP(= PP). They raised the question of whether one can do a Marriott-Watrous-type in-place amplification for this class, which, for instance, means boosting the parameters c and s to be $c = 1 -$

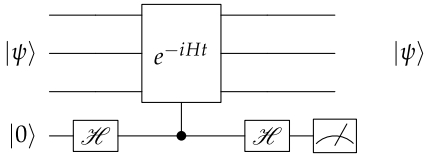

FIG. 3. One-qubit phase-estimation circuit. The symbol $\mathscr{H}$ denotes the Hadamard gate.

$2^{-\mathsf{poly}}$, $s = 2^{-\mathsf{poly}}$ without changing the size of the witness. If one is allowed to change the witness size, one can simply ask for polynomially many copies of the witness and run the verification in parallel to get the required parameters. The benefit of in-place amplification is that it allows for good completeness and soundness parameters without blowing up the witness size, which turns out to be useful in the proof of $\mathsf{QMA} \subseteq \mathsf{PP}$. In-place amplification for $\mathsf{postQMA}$ would also be useful to show that $\mathsf{IP} = \mathsf{PSPACE}$ [38,39]. Here we give a negative result for a sufficiently strong in-place amplification for $\mathsf{postQMA}$.

Lemma 9 (Upper bound for in-place amplified $\mathsf{postQMA}$): *If $f(n) = O(w(n))$ then $\mathsf{postQMA}[1 - 2^{-t(n)}, 2^{-u(n)}] \subseteq \mathsf{PP}$ for $u(n) > w(n) + 1$ and any polynomial $t(n) > 1$.*

Proof. Consider a $\mathsf{postQMA}[1 - 2^{-t(n)}, 2^{-u(n)}]$ language. Replace the witness state in the amplified protocol by a maximally mixed state $\mathbb{1}/2^w$. Now, since the overlap of any witness state with the maximally mixed state is 2^{-w}, the postselection success probability is at least $\Omega(2^{-f(n)-w(n)})$. Furthermore, in the YES case, the probability of accepting the string x (conditioned on success) is

$$\Pr(o = 1) \geq 2^{-w(n)}(1 - 2^{-t(n)}). \tag{6}$$

In the NO case, no matter what state is in the witness register, the accept probability is

$$\Pr(o = 1) \leq 2^{-u(n)}. \tag{7}$$

In $\mathsf{PreciseBQP} = \mathsf{PP}$, we can distinguish between these two cases if $2^{-w} - 2^{-t-w} > 2^{-u}$, i.e., if $1 - 2^{-t} > 2^{w-u}$, for which it suffices to have $u(n) > w(n) + 1$ and $t > 1$. ∎

This result implies that the completeness-soundness gap for $\mathsf{postQMA}$ cannot be boosted beyond a point without incurring a blowup in the size of the witness or by reducing the success probability of postselection.

B. Asymmetric promises on the spectral gap and uniqueness

Motivated by a possible connection to the study of unique witnesses for quantum complexity classes, we consider the complexity class $\mathsf{GQMA}[c, s, g_1, 0]$. Here, there is no promise on the spectral gap for NO instances. In the YES case, we have $\lambda_1(Q) \geq c$ and $\lambda_2 \leq \lambda_1 - g_1 \leq 1 - g_1$. If we choose the spectral gap g_1 to be larger than $1 - s$, we see that $\lambda_2 \leq s$, ensuring that in the YES case, there is exactly one accepting witness [101]. The existence of one accepting witness is exactly the promise that defines the class UQMA.

Definition 18 (Unique QMA [16]): The class $\mathsf{UQMA}[c, s]$ is the class of promise problems $A = (A_{\text{yes}}, A_{\text{no}})$ such that, for every instance x, there exists a polynomial-size verifier circuit U_x acting on $m = \mathsf{poly}(n)$ qubits and an input quantum proof on $w = \mathsf{poly}(n)$ qubits, and the associated accept operator Q has the properties that

(a) if $x \in A_{\text{yes}}$, then $\lambda_1(Q) \geq c$ and $\lambda_2(Q) \leq s$.
(b) if $x \in A_{\text{no}}$, then $\lambda_1(Q) \leq s$.

Definition 19: We define $\mathsf{UQMA} := \bigcup_{c-s \geq 1/\mathsf{poly}} \mathsf{UQMA}[c, s]$.

The earlier statement can be rephrased as "an instance of $\mathsf{GQMA}[c, s, 1 - s, 0]$ is a $\mathsf{UQMA}[c, s]$ instance." In the reverse direction, we can see that a $\mathsf{UQMA}[c, s]$ instance necessarily has a spectral gap $\lambda_1 - \lambda_2 \geq c - s$, and therefore is an instance of $\mathsf{GQMA}[c, s, c - s, 0]$. This hints at, but does not prove, an equivalence between the promise of uniqueness and that of an asymmetric spectral gap of $\Omega(1/\mathsf{poly})$. Aharonov *et al.* [16] proved a stronger result by showing that class UQMA is equivalent to class PGQMA under randomized reductions (defined below), where PGQMA is the class with spectral gaps for *both* the YES and the NO cases.

In the precise regime, we show the following results for the asymmetric variants of $\mathsf{PrecisePGQMA}$ and $\mathsf{PreciseEGQMA}$.

Theorem 11: *It holds that*

$$\mathsf{PrecisePGQMA} = \bigcup_{\substack{c-s \geq 1/\mathsf{exp}, \\ g_1 \geq 1/\mathsf{poly}}} \mathsf{GQMA}[c, s, g_1, 0]$$

and

$$\mathsf{PreciseEGQMA} = \bigcup_{\substack{c-s \geq 1/\mathsf{exp}, \\ g_1 \geq 1/\mathsf{exp}}} \mathsf{GQMA}[c, s, g_1, 0].$$

The proof of Theorem 11 is given in Appendix G and hinges on the problem of computing ground-state energies when there is a spectral gap only for the YES case, i.e., LocalHamiltonian$[a, b, g_1, 0]$. Since the problem with an asymmetric gap can only be more complex than the symmetric case, the nontrivial part of this theorem is to show that this problem has the same PP upper bound as the symmetric case. This is not straightforward since the power method we described before does not necessarily work for the NO case, since there is no spectral gap. We work around this by making use of the technique of Ambainis [94] of identifying spectral gaps, which is possible in PP [95].

C. Complexity of PGQCMA, EGQCMA, and PreciseEGQCMA

In this subsection we show that classes PGQCMA and EGQCMA are both equivalent to QCMA under randomized reductions, which we now define.

We say that a problem A is random reducible to problem X if every instance a of A can be mapped to a random set of polynomially instances x_i of X, such that

(a) if $a \in A_{\text{yes}}$, then $\Pr_i(x_i \in X_{\text{yes}}) \geq 1/\text{poly}$,
(b) if $a \in A_{\text{no}}$, then $\Pr_i(x_i \in X_{\text{yes}}) = 0$.

A class Y is random reducible to another class Z if every problem in Y is random reducible to some problem in Z (and vice versa), and is denoted "$=_R$."

To show that PGQCMA $=_R$ QCMA and EGQCMA $=_R$ QCMA, we make use of class UQCMA (unique QCMA), which has been defined in Ref. [16], and was shown to be equal to QCMA under randomized reductions.

Definition 20 (UQCMA$[c,s]$ [16])**:** The class UQMA$[c,s]$ is the class of promise problems $A = (A_{\text{yes}}, A_{\text{no}})$ such that, for every instance x, there exists a polynomial-size verifier circuit U_x acting on $m = \text{poly}(n)$ qubits and an input classical proof on $w = \text{poly}(n)$ qubits, whose associated accept operator Q has the properties that

(a) if $x \in A_{\text{yes}}$, then $\lambda_1(Q) \geq c$ and $\lambda_2(Q) \leq s$,
(b) if $x \in A_{\text{no}}$, then $\lambda_1(Q) \leq s$.

Definition 21: We define UQCMA $:= \bigcup_{c-s\geq 1/\text{poly}}$ UQCMA$[c,s]$.

Aharonov *et al.* [16] showed that UQCMA $=_R$ QCMA using generalizations of techniques in Ref. [93] to complexity classes with randomness. In order to show that PGQCMA $=_R$ QCMA and EGQCMA $=_R$ QCMA, we show that the following result holds.

Lemma 10: *It holds that* PGQCMA $=_R$ UQCMA.

Since PGQCMA $\subseteq$ EGQCMA $\subseteq$ QCMA, the equivalence of EGQCMA with QCMA follows.

To show Lemma 10, we observe that the proof of PGQMA $=_R$ UQMA in Ref. [16] works for classical witnesses. For completeness, we give a self-contained proof here.

Proof of Lemma 10. First, we show the direction UQCMA $\subseteq$ PGQCMA. We observe that in a YES instance of UQCMA$[c,s]$, $\lambda_1 \geq c$ and $\lambda_2 \leq s$. Thus, a YES instance already has a spectral gap of $g_1 \geq c - s$ and is a YES instance of PGQCMA. In the NO case, we modify the verifier's strategy so that it creates a spectral gap. The verifier expects an additional qubit we call the "flag qubit" from the prover, which is measured in the beginning just like the other qubits of any QCMA proof. The associated accept operator now has twice as many eigenvalues because it acts on a space with one larger qubit.

The verifier's protocol is as follows. If the state of the flag qubit is $|0\rangle$, the verifier continues with the original protocol. This gives the same eigenvalues for the accept operator as the original protocol. If the state of the flag qubit is $|1\rangle$, the verifier accepts with probability $s + (c - s)/\text{poly}$ if the state of the rest of the witness qubits is $|1\rangle^{\otimes w}$. If the state of the rest of the witness register is anything else, the verifier rejects. In the latter case (when the state of the flag qubit is $|1\rangle$), the accept operator has one eigenvalue at $s + (c - s)/\text{poly}$ and $2^w - 1$ eigenvalues with eigenvalue 0, each case corresponding to some state in the witness. The modified verifier is a PGQCMA instance with completeness c, soundness $s + (c - s)/\text{poly}$, and spectral gaps $g_1 \geq c - s$ and $g_2 \geq (c - s)(1 - 1/\text{poly})$. Therefore, UQCMA $\subseteq$ PGQCMA.

For the other direction, we give a randomized reduction PGQCMA $\subseteq_R$ UQCMA. Consider a YES instance of PGQCMA$[c,s,g_1,g_2]$. We know that $\lambda_1 \geq c$ and $\lambda_2 \leq \lambda_1 - g_1$, but we do not know if $\lambda_2 \leq s$, as is required for the instance to be a UQCMA instance. The idea in Ref. [16] is to make a query to a UQCMA$[c_j, s_j]$ oracle with completeness $c_j = c + (j+1)g_1/2$ and soundness $s_j = c + jg_1/2$, for j chosen randomly from $\{0, 1, \ldots, \lfloor 2g_1/(1-c)\rfloor\}$. In the NO case, all the queries are valid queries to a UQCMA oracle and return the correct answer (NO). In the YES case, since the completeness and soundness in each query differ by $g_1/2$, there is at least one j where $\lambda_1 \geq c_j$ and $\lambda_2 \leq s_j$ [102]. Therefore, this is a randomized reduction to UQCMA. ■

Therefore, we obtain the following result.

Corollary 2: *It holds that* PGQCMA $=_R$ QCMA *and* EGQCMA $=_R$ QCMA.

Our final result concerns the class PreciseEGQCMA. Just like we have PreciseEGQMA = PreciseQMA, we can show that exponentially small spectral gaps are no less complex in the case of classical witnesses. We show that the following result holds.

Lemma 11: *It holds that* PreciseEGQCMA = Precise QCMA.

Proof. The direction PreciseEGQCMA $\subseteq$ Precise QCMA is trivial. For the other direction, we take a PreciseQCMA$[c,s]$ instance and give a PreciseEGQCMA$[c,s,g_1,g_2]$ instance with an exponentially small spectral gap. This is done by modifying the verifier so that no two witnesses y_i and y_j are accepted with the same probability. First, we choose the verifier's gate set so that the accept probability of any witness y is given by $k_{x,y}/2^{l(n)}$, for $k_{x,y} \in [2^{l(n)}]$, where $l(n)$ is the size of

the verifier's circuit [103]. The modified verifier rejects the instance straightaway with probability $y_b/2^{\mathsf{poly}}$, where y_b is a number in $[2^w - 1]$ when interpreting the witness y in binary and the polynomial is at least $l(n) + w(n) + \log_2(1/(c-s))$. If the verifier does not reject at this step, they run the original verification protocol. The overall accept probability when given y is given by $p_y = k_{x,y}(1 - y_b/2^{\mathsf{poly}})/2^w$. Since the polynomial satisfies $\mathsf{poly} \geq l(n) + w(n) + \log_2(1/(c-s))$, the completeness and soundness are given by $c' \geq c - 2^{-w(n)}(c-s)$ and $s' = s$, which are still separated by $2^{-\mathsf{poly}}$.

We now claim that the resulting accept probabilities are distinct for distinct witnesses, and hence separated by an amount $\Omega(2^{-\mathsf{poly}})$. This is easily seen for two distinct y_i and y_j such that $k_{x,y_i} = k_{x,y_j}$. If $k_{x,y_i} \neq k_{x,y_j}$ then, for $p_{y_i} = p_{y_j}$, we need

$$k_{x,y_i} - k_{x,y_j} = \frac{2^w}{2^{l+w+\mathsf{poly}}}(y_{j_b} - y_{i_b}), \tag{8}$$

which cannot be satisfied by integers y_{j_b} and y_{i_b} in $[2^l]$. ■

The same technique also works to give a more direct proof of $\mathsf{EGQCMA} = \mathsf{QCMA}$.

ACKNOWLEDGMENTS

We thank Sevag Gharibian, Alex Grilo, Hosho Katsura, Cedric Lin, Yupan Liu, and Zachary Remscrim for useful feedback on the manuscript. We thank an anonymous referee of an earlier version of this manuscript for pointing out the similarity of imaginary-time evolution with the power method and Cedric Lin for showing that the power method can give a PP algorithm as opposed to a P^{PP} algorithm. We also thank Zachary Remscrim for suggesting the proof of Lemma 11, and Peter Love, Yuan Su, Minh Tran, and Seth Whitsitt for helpful discussions. A.D. and A.V.G. were supported in part by the DoE ASCR Accelerated Research in Quantum Computing program (Award No. DE-SC0020312), DoE ASCR Quantum Testbed Pathfinder program (Award No. DE-SC0019040), DoE QSA, NSF QLCI, U.S. Department of Energy Award No. DE-SC0019449, NSF PFCQC program, AFOSR, ARO MURI, AFOSR MURI, and DARPA SAVaNT ADVENT. A.D. also acknowledges support from NSF RAISE/TAQS 1839204. B.F. acknowledges support from AFOSR (YIP numbers FA9550-18-1-0148 and FA9550-21-1-0008). This material is based upon work partially supported by the National Science Foundation under Grant CCF- 2044923 (CAREER) and by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers. The Institute for Quantum Information and Matter is a NSF Physics Frontiers Center (PHY-1733907).

APPENDIX A: THE SCHRIEFFER-WOLFF TRANSFORMATION

In this section, we give a brief introduction to the Schrieffer-Wolff transformation [53], which is an important tool in some of our subsequent proofs. We follow the exposition in Ref. [54], specialized to our context.

In the context relevant for us, we usually have an "unperturbed" Hamiltonian H_0 and a "perturbation" H_1, together forming the full Hamiltonian $H = H_0 + H_1$. The (possibly degenerate) ground-state subspace of H_0, denoted $\mathcal{S}_0$, has energy λ_0 and is separated from the rest of the spectrum by a gap Δ. We are interested in the cases when Hamiltonian H_1 has small strength relative to the gap Δ, in the sense $\|H_1\| =: \epsilon < \Delta/2$. This ensures that all eigenvalues of H_0 are shifted by an amount smaller than $\Delta/2$ under the perturbation. Therefore, the low-energy subspace of H, given by

$$\mathcal{S} = \left\{ |\psi\rangle : \langle\psi|H|\psi\rangle \in \left[\lambda_0 - \frac{\Delta}{2}, \lambda_0 + \frac{\Delta}{2}\right] \right\}, \tag{A1}$$

has the same dimension as that of H_0. We denote the projectors on to $\mathcal{S}_0$ and $\mathcal{S}$ by P_0 and P, respectively. As long as $\epsilon < \Delta/2$, we have $\|P - P_0\| < 1$, which captures the fact that the dimensions of the two subspaces are the same.

Since the dimensions of the two subspaces are the same, there exists a unitary U that maps the subspace $\mathcal{S}_0$ to $\mathcal{S}$:

$$UPU^\dagger = P_0 \quad \text{with} \quad U = \sqrt{(2P_0 - \mathbb{1})(2P - \mathbb{1})}. \tag{A2}$$

We are interested in the effective Hamiltonian in subspace $\mathcal{S}_0$, given by

$$H_{\text{eff}} = P_0 U(H_0 + H_1)U^\dagger P_0. \tag{A3}$$

The Schrieffer-Wolff transformation allows one to express the generator $V = \log(U)$, and consequently, H_{eff}, as a convergent series in the perturbation H_1. We first write H_1 as $H_1^d + H_1^o$, where H_1^d is block diagonal in subspace $\mathcal{S}_0$ and H_1^o is block off-diagonal. Let the eigenstates of H_0 be given by $\{|i\rangle\}$, with corresponding energies $\{E_i\}$. We define $\mathcal{I}_0 = \{i : E_i = \lambda_0\}$, which is the set of indices corresponding to the ground-state space. The first few terms of the Schrieffer-Wolff expansion are given by

$$\begin{aligned} H_{\text{eff}} = {} & H_0 P_0 + P_0 H_1 P_0 \\ & + \frac{1}{2} P_0 \sum_{i \in \mathcal{I}_0, j \notin \mathcal{I}_0} \left(\frac{\langle i|H_1|j\rangle}{E_i - E_j} |i\rangle\langle j| H_1 \right. \\ & \left. + \frac{\langle j|H_1|i\rangle}{E_i - E_j} H_1 |j\rangle\langle i| \right) P_0 + O(\|H_1\|^3). \end{aligned} \tag{A4}$$

In our work, we use the first-order expansion of the Schrieffer-Wolff series. The series converges absolutely as

long as $\|H_1\| \leq \Delta/16$ [54]. We can upper bound the error caused by truncating the formal series to first order:

$$\begin{aligned}
&\|H_{\text{eff}} - H_0P_0 - P_0H_1P_0\| \\
&\quad \leq O(1)\Bigg\|P_0 \sum_{i\in\mathcal{I}_0, j\notin\mathcal{I}_0} \Bigg(\frac{\langle i|H_1|j\rangle}{E_i - E_j}\,|i\rangle\langle j\,|\,H_1 \\
&\qquad + \frac{\langle j\,|H_1|i\rangle}{E_i - E_j} H_1\,|j\,\rangle\langle i|\Bigg)P_0\Bigg\| \\
&\quad \leq O(1)\Bigg\| \sum_{i\in\mathcal{I}_0, j\notin\mathcal{I}_0, k\in\mathcal{I}_0} \frac{1}{E_i - E_j}(\langle i|H_1|j\rangle\langle j\,|H_1|k\rangle\,|i\rangle\langle k| \\
&\qquad + \langle j\,|H_1|i\rangle\langle k|H_1|j\,\rangle\,|k\rangle\langle i|)\Bigg\| \\
&\quad \leq O\left(\frac{1}{\Delta}\right)\Bigg\| \sum_{i\in\mathcal{I}_0, k\in\mathcal{I}_0} (\langle i|H_1^2|k\rangle\,|i\rangle\langle k| + \langle k|H_1^2|i\rangle\,|k\rangle\langle i|)\Bigg\| \\
&\quad = O\left(\frac{1}{\Delta}\right)\|2P_0H_1^2P_0\| \\
&\quad \leq O\left(\frac{\epsilon^2}{\Delta}\right). \qquad \text{(A5)}
\end{aligned}$$

Here we have used the fact that $|E_i - E_j| > \Delta$ for states $i \in \mathcal{I}_0, j \notin \mathcal{I}_0$.

APPENDIX B: MODIFIED CLOCK CONSTRUCTIONS WITH SPECTRAL GAPS

In this section, we present the small-penalty clock construction and use it to prove the main hardness results in this work. We first illustrate the technique by proving the following lemma.

Lemma 12: *It holds that* $(1/\mathsf{exp}, 1/\mathsf{exp})$-LOCAL-HAMILTONIAN *is* PreciseEGQMA*-hard.*

Proof. Consider a $\mathsf{GQMA}[c, s, g_1, g_2]$ instance x, where the verifier's circuit U_x acts on $m = \mathsf{poly}(n)$ qubits apart from the proof state. We assume that the circuit has $T = \mathsf{poly}(n)$ gates. The idea behind the technique is valid generally, but for concreteness, we focus on the clock construction of Kempe *et al.* [44], which proves QMA-hardness of k-LOCALHAMILTONIAN for $k \geq 3$. The clock Hamiltonian takes the form

$$H = H_{\text{input}} + H_{\text{prop}} + H_{\text{output}} + H_{\text{clock}}. \qquad \text{(B1)}$$

The first term H_{input} ensures that the ground state of H_{input} coincides with the input state to the circuit. The term on the proof register is identity, allowing for any witness state given by the prover to be input into the verifier's circuit. It is given by

$$H_{\text{input}} = \sum_{i=1}^{m} |1\rangle\langle 1|_i \otimes \mathbb{1}_{\text{proof}} \otimes H_{\text{clockinit}}. \qquad \text{(B2)}$$

In the above, the term $H_{\text{clockinit}}$ ensures that the clock is properly initialized to the $|1\rangle_{\text{clock}}$ state. Next, H_{prop} is a Hamiltonian that ensures that the ground state is "propagated" correctly with each gate applied by the verifier:

$$\begin{aligned}
H_{\text{prop}} = \sum_{i=0}^{T} &-U_{i+1} \otimes |i+1\rangle\langle i|_{\text{clock}} - U_{i+1}^{\dagger} \otimes |i\rangle\langle i+1|_{\text{clock}} \\
&+ \mathbb{1} \otimes (|i\rangle\langle i|_{\text{clock}} + |i+1\rangle\langle i+1|_{\text{clock}}). \qquad \text{(B3)}
\end{aligned}$$

The ground-state subspace of H_{prop} contains valid "partial" computations until step $i \leq T$, namely $U_i \cdots U_2U_1|\psi_0\rangle$ on any initial state $|\psi_0\rangle$ for all i. The term H_{output} penalizes states that have any nonzero probability of saying "NO" at the output qubit o of the circuit:

$$H_{\text{output}} = \epsilon\,|0\rangle\langle 0|_o \otimes |T\rangle\langle T|_{\text{clock}}\,. \qquad \text{(B4)}$$

Lastly, H_{clock} ensures that states in the clock register that do not encode a valid time step are penalized. The Hamiltonians H_{clock} and $H_{\text{clockinit}}$ both depend on the details of the particular clock construction. Our analysis does not depend on these details and is largely independent of the way the clock register encodes the time. We refer the reader to Ref. [44] for an explanation of their construction.

First consider just the Hamiltonian $H_0 = H_{\text{input}} + H_{\text{prop}} + H_{\text{clock}}$, which is the clock Hamiltonian without a penalty term at the output. The ground-state space of H_0 is exactly given by the subspace $\mathcal{S}_0$ of history states:

$$\mathcal{S}_0 = \text{span}\{|\phi_h\rangle : |\phi\rangle\text{arbitrary}\}, \text{ where}$$

$$|\phi_h\rangle := \frac{1}{\sqrt{T+1}} \sum_{i=0}^{T} U_i \cdots U_0|0^m\rangle \otimes |\phi\rangle \otimes |i\rangle_{\text{clock}}$$

with $U_0 = \mathbb{1}$. Any state having zero support on $\mathcal{S}_0$ has an energy at least $\Omega(1/T^3)$ [74], implying that the gap above the zero energy subspace is $\Delta = \Omega(1/T^3)$.

Now, let us add in the term $H_1 = H_{\text{output}}$, with $\|H_{\text{output}}\| = \epsilon$. We choose $\epsilon < \Delta/16$, unlike the regular clock construction where ϵ is usually taken to be $\Theta(1)$. As long as $\epsilon < \Delta/2$, we can restrict our attention to the zero energy space of H_0, since H_1 can change eigenvalues by at most ϵ. We use the Schrieffer-Wolff transformation tool as described in Appendix A to obtain a description of the Hamiltonian in the low-energy subspace. Subspace $\mathcal{S}_0$ is the ground-state space of states with energy 0. Since

$\|H_1\| = \epsilon$, the associated low-energy subspace of $H = H_0 + H_1$ is

$$\mathcal{S} = \text{span}\{|\Phi\rangle : \langle\Phi|H|\Phi\rangle \in [-\epsilon, \epsilon]\}, \tag{B5}$$

the subspace with energies in $[-\epsilon, \epsilon]$. In our case $H_0 P_0 = 0$ in the ground subspace spanned by history states $|\phi_h\rangle$, and the matrix elements of $P_0 H_1 P_0$ are given by

$$\begin{aligned}
&\langle\phi_h|P_0 H_1 P_0|\psi_h\rangle \\
&\quad = \langle\phi_h|H_1|\psi_h\rangle \\
&\quad = \frac{1}{T+1}\Bigg(\sum_{i=0}^{T}\langle 0|^m \otimes \langle\phi| \otimes \langle i|_{\text{clock}} U_0^\dagger \cdots U_i^\dagger\Bigg) H_1 \\
&\qquad \times \Bigg(\sum_{j=0}^{T} U_j \cdots U_0 |0^m\rangle \otimes |\psi\rangle \otimes |j\rangle_{\text{clock}}\Bigg) \\
&\quad = \frac{1}{T+1}\Bigg(\sum_{i=0}^{T}\langle 0|^m \otimes \langle\phi| \otimes \langle i|_{\text{clock}} U_0^\dagger \cdots U_i^\dagger\Bigg) \\
&\qquad \times \epsilon\, |0\rangle\langle 0|_o \otimes |T\rangle\langle T|_{\text{clock}} \\
&\qquad \times \Bigg(\sum_{j=0}^{T} U_j \cdots U_0 |0^m\rangle \otimes |\psi\rangle \otimes |j\rangle_{\text{clock}}\Bigg) \\
&\quad = \frac{1}{T+1}\langle 0|^m \otimes \langle\phi| \otimes \langle T|U^\dagger \epsilon\, |0\rangle\langle 0|_o \\
&\qquad \otimes |T\rangle\langle T|_{\text{clock}}\, U|0^m\rangle \otimes |\psi\rangle|T\rangle_{\text{clock}} \\
&\quad = \frac{\epsilon}{T+1}\langle 0|^m \otimes \langle\phi|U^\dagger\, |0\rangle\langle 0|_o\, U|0^m\rangle \otimes |\psi\rangle \\
&\quad = \frac{\epsilon}{T+1}\langle 0|^m \otimes \langle\phi|U^\dagger(\mathbb{1} - \Pi_{\text{out}})U|0^m\rangle \otimes |\psi\rangle,
\end{aligned} \tag{B6}$$

where Π_{out} is the projector onto the accepting state $|1\rangle_o$. Continuing, we have

$$\langle\phi_h|P_0 H_1 P_0|\psi_h\rangle = \frac{\epsilon}{T+1}(\langle\phi|\psi\rangle - \langle\phi|Q|\psi\rangle), \tag{B7}$$

meaning that the first-order correction $P_0 H_1 P_0$ is simply related to the accept operator Q, which was defined as $Q(U) = \langle 0|^{\otimes m} U^\dagger \Pi_{\text{out}} U|0\rangle^{\otimes m}$. Let the eigenstates of Q be $|\phi_1\rangle, |\phi_2\rangle, \ldots, |\phi_{2^w}\rangle$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{2^w}$. We use the associated history states $|\phi_{i_h}\rangle$ as a basis for subspace $\mathcal{S}_0$. In this basis, the first-order correction $P_0 H_1 P_0$ is diagonal:

$$P_0 H_1 P_0 = \frac{\epsilon}{T+1}\sum_i (1 - \lambda_i)\left|\phi_{i_h}\right\rangle\!\left\langle\phi_{i_h}\right|. \tag{B8}$$

We conclude that in the ground space of the original Hamiltonian H_0, the full Hamiltonian H has eigenvalues $\epsilon(1 - \lambda_i)/(T+1) \pm O(\epsilon^2/\Delta)$, where the quantity λ_i is the accept probability of the verifier's circuit given $|\phi_i\rangle$ as the witness. This is the same conclusion we would obtain by applying degenerate perturbation theory, except that the error bound is rigorous. We now analyze the YES and NO cases to obtain a lower bound on the promise gap. In each case, we also lower bound the spectral gaps in the resulting Hamiltonian.

In the YES case the ground-state energy is $E_1 \leq \epsilon(1 - c)/(T+1)$, as can be seen from the fact that the history state $|\phi_h\rangle$ corresponding to an accepting witness $|\phi\rangle$ would have energy $\epsilon(1 - \langle\phi|Q|\phi\rangle)/(T+1) \leq \epsilon(1-c)/(T+1)$. Our small-penalty clock construction and the Schrieffer-Wolff transformation comes in handy for the NO case. We see in the NO case that the ground-state energy is at least $E_1 \geq \epsilon(1-s)/(T+1) - O(\epsilon^2/\Delta)$. Therefore, the promise gap is at least $\epsilon(c-s)/(T+1) - O(\epsilon^2/\Delta) = \Omega(1/\mathsf{exp})$ as long as $\epsilon/\Delta = o((c-s)/(T+1))$.

In the above, if we had chosen $\epsilon = \Theta(1)$ instead of $\epsilon < \Delta/16$, the NO case would have given us a bound $E_1 \geq \Omega(1-s)/T^3$. This would mean that one would have to amplify the completeness and soundness c, s to near unity in order to get a nontrivial promise gap. However, such an amplification can, in general, shrink the spectral gap of the accept operator. Independently, a large penalty term $\epsilon = \Theta(1)$ could also reorder some eigenvalues, meaning that the spectral properties of the resulting clock Hamiltonian would not faithfully track those of the original accept operator.

The spectral gap in the YES or NO case is $E_2 - E_1 \geq \epsilon(\lambda_1(Q) - \lambda_2(Q))/(T+1) - O(\epsilon^2/\Delta)$. We take $\epsilon = o(\Delta(c-s)/(T+1)) = o((c-s)/T^4)$, which is exponentially small if $c - s$ is. As long as $\epsilon/\Delta = o(\min[g_1, g_2]/(T+1))$, both the YES and NO cases will have an exponentially small spectral gap. In summary, the choice

$$\epsilon = \frac{\min[g_1, g_2, (c-s)]}{nT^4} = \Theta(1/\mathsf{exp}) \tag{B9}$$

suffices to have a promise gap and spectral gaps bounded below by $\Omega(1/\mathsf{exp})$. This proves PreciseEGQMA-hardness of $(1/\mathsf{exp}, 1/\mathsf{exp})$-LOCALHAMILTONIAN and one half of Theorem 2. ■

We generalize the above proof technique to the case of GQCMA-hardness of GS-DESCRIPTION-LOCALHAMILTONIAN. In addition to showing a promise gap and a spectral gap, we should show that the resulting Hamiltonian has a classical description of a circuit to prepare a low-energy state. We show the following general lemma.

Lemma 13: *It holds that* (δ, Δ)-GS-DESCRIPTION-LOCALHAMILTONIAN *is* GQCMA$[c, s, g_1, g_2]$*-hard for any* δ, Δ *satisfying both of the following conditions:*

(i) $\delta = O((c-s)^2/\mathsf{poly}(n))$ *for some polynomial,*

(ii) if $c - s = o(\min[g_1, g_2])$ *then any* Δ *satisfies* $\Delta = O((c-s)\min[g_1,g_2]/\mathsf{poly}(n))$*; otherwise,* $\Delta = 0$.

Proof. To prove GQCMA-hardness, we give a reduction from GQCMA$[c, s, g_1, g_2]$ to GS-DESCRIPTION-LOCALHAMILTONIAN$[a, b, g_1', g_2']$. We are promised that the input witnesses are computational basis states (this can be assumed without loss of generality), corresponding to the classical witness sent by the prover. We would like to show that there exists a circuit V to prepare a state δ-close in energy to the ground state of the clock Hamiltonian in both the YES and NO cases.

Consider again the small-penalty clock construction, with the clock Hamiltonian (B1). Let the norm of the penalty term be $\|H_{\text{output}}\| = \epsilon$. When $\epsilon = 0$, the ground-state space is given by valid history state computations corresponding to computational basis witness states. The spectral gap above this subspace is at least $\Omega(1/T^3)$. The addition of the penalty term changes the energies to $\epsilon(1-\lambda_k)/(T+1) + O(\epsilon^2 T^3)$, where λ_k is the accept probability upon input computational basis state $|y_k\rangle$ as the witness. Consider the history state associated with witness $|y_k\rangle$:

$$|y_{k_h}\rangle := \frac{1}{\sqrt{T+1}} \sum_{i=0}^{T} U_i \cdots U_0 |0^m\rangle \otimes |y_k\rangle \otimes |i\rangle_{\text{clock}}. \tag{B10}$$

This state has energy $\langle y_{k_h}|H|y_{k_h}\rangle = \epsilon(1-\lambda_k)/(T+1)$ and is therefore $O(\epsilon^2 T^3)$-close in energy to the true ground state. Therefore, as long as $\epsilon^2 T^3 < O((b-a)^3/f(n)^2)$, a classical description of a circuit that prepares $|y_{k_h}\rangle$ is a valid ground-state description. The circuit may be described by specifying y_k and a circuit that prepares the history state $|\phi_h\rangle$ upon any quantum input $|\phi\rangle$. This latter circuit first prepares the state $[1/(\sqrt{T+1})]\sum_{i=0}^{T}|0^m\rangle|i\rangle_{\text{clock}}$ and then applies the unitaries $U_j \cdots U_0$ controlled on the clock being in time step j [98].

The same promise gap and spectral gap analyses as in the proof of Lemma 12 hold. In the YES case, the Hamiltonian has ground-state energy $\leq \epsilon(1-\lambda_1)/(T+1) \leq \epsilon(1-c)/(T+1)$. In the NO case, the ground-state energy is at least $\epsilon(1-\lambda_1)/(T+1) - O(\epsilon^2 T^3) \geq \epsilon(1-s)/(T+1) - O(\epsilon^2 T^3)$. The promise gap between the ground-state energy for YES and NO cases is $\delta \geq \epsilon(c-s)/(T+1) - O(\epsilon^2 T^3)$. We make the choice $\epsilon = \Theta((c-s)/T^4)$ to ensure that the promise gap is $\Omega((c-s)^2/T^5)$. This choice is consistent with the choice $\epsilon^2 T^3 \leq O((b-a)^3/f(n)^2)$ made above.

Let us now analyze the spectral gap of the resulting Hamiltonian. Using the Schrieffer-Wolff expansion to obtain the eigenvalues of the Hamiltonian for small ϵ, we have $\Delta \geq \epsilon(\lambda_1 - \lambda_2)/(T+1) - O(\epsilon^2 T^3)$. The spectral gap is at least $[\epsilon/(T+1)]\min[g_1, g_2]$ as long as $\epsilon^2 T^3 = o([\epsilon/(T+1)]\min[g_1,g_2])$. Using the choice of ϵ above, this means that the spectral gap is $\Omega(((c-s)/T^5)\ \min[g_1,g_2])$ as long as $c - s = o(\min[g_1,g_2])$. Otherwise, the best bound on the spectral gap is $\Delta \geq 0$. Observing that $T = \mathsf{poly}(n)$ by assumption, we obtain the lemma. ∎

The lemma allows us to show the following.

Corollary 3 (Second half of Theorems 3 to 6): *It holds that*

(a) $(1/\mathsf{poly}, 0)$-GS-DESCRIPTION-LOCALHAMILTONIAN *is* QCMA*-hard,*
(b) $(1/\mathsf{exp}, 0)$-GS-DESCRIPTION-LOCALHAMILTONIAN *is* PreciseQCMA*-hard,*
(c) $(1/\mathsf{poly}, 1/\mathsf{poly})$-GS-DESCRIPTION-LOCALHAMILTONIAN *is* PGQCMA*-hard,*
(d) $(1/\mathsf{exp}, 1/\mathsf{exp})$-GS-DESCRIPTION-LOCALHAMILTONIAN *is* PreciseEGQCMA*-hard.*

For the problem with $\delta = 1/\mathsf{exp}, \Delta = 1/\mathsf{poly}$, we do not give a direct reduction from a PrecisePGQCMA instance. Instead, we show PP-hardness through the characterization of PP in terms of the class PreciseBQP. From the PP upper bound to PrecisePGQCMA, we obtain PrecisePGQCMA-completeness of the problem $(1/\mathsf{exp}, 1/\mathsf{poly})$-GS-DESCRIPTION-LOCALHAMILTONIAN. The argument is similar for PrecisePGQMA-hardness of $(1/\mathsf{exp}, 1/\mathsf{poly})$-LOCALHAMILTONIAN.

Lemma 14 (Lemmas 4 and 6 restated): *It holds that*

(a) $(1/\mathsf{exp}, 1/\mathsf{poly})$-GS-DESCRIPTION-LOCALHAMILTONIAN *is* PP*-hard,*
(b) $(1/\mathsf{exp}, 1/\mathsf{poly})$-LOCALHAMILTONIAN *is* PP*-hard.*

Proof. We give a reduction from any problem in PreciseBQP to $(1/\mathsf{exp}, 1/\mathsf{poly})$-GS-DESCRIPTION-LOCAL HAMILTONIAN, which is also an instance of $(1/\mathsf{exp}, 1/\mathsf{poly})$-LOCALHAMILTONIAN. Since PreciseBQP is the class of problems that can be decided by quantum circuits with a promise gap $c - s = \Omega(1/\mathsf{exp})$, it can also be thought of as "PreciseQMA without an input witness." The Hamiltonian is constructed out of the PreciseBQP computation as $H = H_{\text{input}} + H_{\text{prop}} + H_{\text{output}} + H_{\text{clock}}$, where the terms are now

$$H_{\text{input}} = \sum_{i=1}^{m} |0\rangle\langle 0|_i \otimes H_{\text{clockinit}}, \tag{B11}$$

$$H_{\text{prop}} = \sum_{i=0}^{T} -U_{i+1} \otimes |i+1\rangle\langle i|_{\text{clock}} - U_{i+1}^{\dagger} \otimes |i\rangle \times \langle i+1|_{\text{clock}} + \mathbb{1} \otimes (|i\rangle\langle i|_{\text{clock}} + |i+1\rangle \times \langle i+1|_{\text{clock}}), \tag{B12}$$

$$H_{\text{output}} = \epsilon\, |0\rangle\langle 0|_o \otimes |T\rangle\langle T|_{\text{clock}}\,. \tag{B13}$$

The only difference from Eqs. (B2) to (B4) is that H_{input} does not have support on an unpenalized proof register, since PreciseBQP does not rely on a proof state given as input. This is analogous to the clock construction of Ref. [74], which was instrumental in the proof that adiabatic quantum computation is universal for BQP.

We again let Hamiltonian H_0 be $H_{\text{input}} + H_{\text{prop}} + H_{\text{clock}}$ and $H_1 = H_{\text{output}}$. The ground state of H_0 is now nondegenerate (unique) and given by the history state

$$|0_h\rangle := \frac{1}{\sqrt{T+1}} \sum_{i=0}^{T} U_i \cdots U_0 |0^m\rangle \otimes |i\rangle_{\text{clock}}. \tag{B14}$$

Let us denote the ground-state space of H_0 and the projector onto it by Π_0. As for H_1, the ground space Π_1 is spanned by states belonging to subspaces $\mathcal{L}$ and $\mathcal{L}'$, with

$$\mathcal{L} = |1\rangle_o \otimes |T\rangle_{\text{clock}}, \tag{B15}$$

$$\mathcal{L}' = \text{span}\,\{|\psi\rangle\} \otimes \text{span}\,\{|0\rangle_{\text{clock}}, |1\rangle_{\text{clock}}, \ldots, |T-1\rangle_{\text{clock}}\}, \tag{B16}$$

with $|\psi\rangle$ arbitrary.

We observe that, when $\epsilon = 0$, the Hamiltonian exactly corresponds to H_{final} of Aharonov *et al.* [74]. Aharonov *et al.* [74] showed that this Hamiltonian H_0 has a spectral gap of $\Delta = \Omega(1/T^3)$ in the full Hilbert space. Furthermore, the ground state of H_0 corresponds to the history state of the BQP computation (PreciseBQP in this case), which starts off in a fixed, known state $|0^m\rangle$.

In the YES case, the ground-state energy of $H = H_0 + H_1$ can be bounded above by $\epsilon(1-c)/(T+1)$. For the NO case, we again use the expression for the perturbed energies in the ground-state space coming from the Schrieffer-Wolff transformation. Specifically, in the NO case, we have $E_1 \geq \epsilon(1-s)/(T+1) - O(\epsilon^2/\Delta)$, where Δ is the spectral gap above the ground state, just as in the proof of Lemma 12. The promise gap is lower bounded by

$$\epsilon\frac{1-s}{T+1} - \epsilon\frac{1-c}{T+1} - O\left(\frac{\epsilon^2}{\Delta}\right). \tag{B17}$$

Therefore, as long as $\epsilon/\Delta = o((c-s)/(T+1))$ and $\epsilon = \Omega(2^{-\text{poly}})$, the promise gap is at least $\Omega(\epsilon(c-s)/(T+1)) = \Omega(2^{-\text{poly}})$. The spectral gap for the unperturbed Hamiltonian H_0, which is the same as the final Hamiltonian in Ref. [16], is at least $\Omega(1/T^3)$. Therefore, we pick $\epsilon = (c-s)/(nT^4)$, which ensures that the conditions above are satisfied.

Coming to the spectral gap of the full Hamiltonian, we observe that, since the original Hamiltonian had a spectral gap of $\Omega(1/T^3)$ and perturbation H_1 is exponentially small, the eigenvalues can change at most by $\|H_1\| = \epsilon$, preserving the spectral gap. So far, we have a reduction from any PreciseBQP instance to an instance of $(1/\text{exp}, 1/\text{poly})$-LocalHamiltonian.

It remains for us to see that there is an efficient circuit that can prepare a state close in energy to the ground state. By the justification in the proof of Lemma 13, we know that choosing the output penalty term to be exponentially small causes the history state of computation $|0_h\rangle$ to be exponentially close to the ground state in energy. We have also seen the existence of a polynomial-size circuit that prepares the history state given a description of the input (which here is $|0^m\rangle$ for PreciseBQP). Note that, when $\epsilon = 0$, the ground state is unique and has a $\Omega(1/\text{poly})$ spectral gap above and therefore taking ϵ exponentially small does not pose a problem with spectral gaps. ■

The difference between the proof of Lemma 14 and the proof of Lemma 12 is that it is the perturbation ϵ that creates the spectral gap in the proof of Lemma 14, while in the proof of Lemma 12, the spectral gap already exists in the unperturbed Hamiltonian. This is why we can afford to take ϵ exponentially small here, which is needed to obtain an instance with a promise gap.

Thus, we have seen PP-hardness of $(1/\text{exp}, 1/\text{poly})$-LocalHamiltonian. The PrecisePGQMA-hardness of the problem follows from the fact that PrecisePGQMA $\subseteq$ PP (Lemma 2).

Corollary 4: *It holds that* $(1/\text{exp}, 1/\text{poly})$-LocalHamiltonian *is* PrecisePGQMA*-hard.*

Similarly, the PP-hardness of $(1/\text{exp}, 1/\text{poly})$-GS-Description-LocalHamiltonian and the result PrecisePGQCMA $\subseteq$ PrecisePGQMA $=$ PP together imply the following result.

Corollary 5: *It holds that* $(1/\text{exp}, 1/\text{poly})$-GS-Description-LocalHamiltonian *is* PrecisePGQCMA*-hard.*

Lastly, the remaining case is $(1/\text{poly}, 1/\text{exp})$-GS-Description-LocalHamiltonian with $\delta = 1/\text{poly}$, $\Delta = 1/\text{exp}$, for which we argue that an instance with spectral gap $\Delta = \Omega(1/\text{poly})$ is also an instance with $\Delta = \Omega(1/\text{exp})$. Therefore, $(1/\text{poly}, 1/\text{exp})$-GS-Description-LocalHamiltonian is PGQCMA-hard, and, since PGQCMA $=_R$ EGQCMA, EGQCMA-hard under randomized reductions. For the case of $(1/\text{poly}, 1/\text{exp})$-LocalHamiltonian, we do not currently have a hardness result. This is because, in performing a reduction

from EGQMA, we get an instance of $(1/\mathsf{poly}, 0)$-LOCALHAMILTONIAN and do not get any promise on the spectral gap that results.

APPENDIX C: PRECISE PHASE ESTIMATION OF GAPPED HAMILTONIANS

In this section, we show that the $(1/\mathsf{exp}, \Delta)$-LOCALHAMILTONIAN problems with either $1/\mathsf{poly}$ or $1/\mathsf{exp}$ spectral gaps defined in Sec. II A are in the corresponding PreciseGQMA class. Together with the results of the previous section, this proves Theorems 1 and 2.

Lemma 15: *It holds that* $(1/\mathsf{exp}, 1/\mathsf{poly})$-LOCALHAMILTONIAN $\in$ PrecisePGQMA.

Lemma 16: *It holds that* $(1/\mathsf{exp}, 1/\mathsf{exp})$-LOCALHAMILTONIAN $\in$ PreciseEGQMA.

The proof relies on phase estimation to infer energies of a local Hamiltonian. The standard phase estimation circuit requires $\mathsf{exp}(n)$ many gates in order to infer the eigenvalues to $1/\mathsf{exp}$ precision. However, since we want to show containment in a Precise- class, we can use the power of being able to distinguish between two cases with exponentially close accept probabilities. It turns out that phase estimation with a single ancillary qubit is enough to distinguish between the YES and NO cases, as shown in Ref. [28]. Moreover, we show that the circuit preserves spectral gaps of the Hamiltonian: if two eigenstates have energies separated by some amount then the phase estimation circuit also has a gap in the accept probabilities corresponding to these input states.

Below we give a unified proof of Lemmas 8, 15, and 16. Specifically, we show PrecisePGQMA (PreciseEGQMA) containment of the problem $(1/\mathsf{exp}, \Delta)$-GAPPEDSPARSEHAMILTONIAN with $\Delta = 1/\mathsf{poly}$ ($\Delta = 1/\mathsf{exp}$).

Lemma 17: *Problem* GAPPEDSPARSEHAMILTONIAN$[a, b, g_1, g_2]$ *has a* GQMA$[c, s, g_1', g_2']$ *protocol with spectral gaps* $g_1' = \Omega(g_1^2/\mathsf{poly})$ *and* $g_2' = \Omega(g_2^2/\mathsf{poly})$ *and promise gap* $c - s = (b-a)^2/\mathsf{poly}$.

Proof. The strategy is to ask the prover for the ground state of the sparse Hamiltonian. The verifier then performs phase estimation on the witness state with a single ancillary qubit, as illustrated in Fig. 3, and uses the power to decide between two cases with exponentially close accept probabilities. This power effectively enables computation of the phase of e^{-iHt} to exponential precision, despite having a single ancilla qubit in the phase estimation circuit (see Ref. [28] for more details). If $t \le \pi/(2\|H\|)$, all eigenstates of H would correspond to a unique phase and a unique accept probability for the circuit. We know an upper bound dk on $\|H\|$ through the Gershgorin circle theorem because we are assured that the magnitude of the entries is $\le k$ and the sparsity is d. Therefore, it suffices to choose $t \le \pi/(2dk)$.

In order to perform phase estimation to exponentially small error, we need to apply a controlled-e^{-iHt} rotation to error $\epsilon = 1/\mathsf{exp}$. This is possible due to Hamiltonian simulation algorithms for sparse Hamiltonians, whose circuit size scales as $\mathsf{poly}(n)\log(1/\epsilon)$ [21], which is polynomial in n, as desired. The accept probability of the circuit upon input an eigenstate $|E_i\rangle$ of the Hamiltonian is $(1+\cos(E_i t))/2$. The promise gap can be lower bounded by an inverse exponential, as has been analyzed previously [28].

We can also show a spectral gap in the accept operator, or, equivalently, a gap in the accept probabilities of the circuit for the optimal state and any state orthogonal to it. Since the phase estimation circuit does not apply the exact controlled-e^{-iHt} unitary but a unitary U_x exponentially close to it, the eigenstates of $Q = \langle 0|\Pi_{\text{in}} U_x^\dagger \Pi_{\text{out}} U_x \Pi_{\text{in}}|0\rangle$ are not exactly the eigenstates of e^{-iHt} (or of H). However, since $\|e^{-iHt} - U_x\| \le \epsilon$, the eigenvalues of Q are exponentially close to the accept probabilities of the eigenstates $|E_i\rangle$ of H. The difference in accept probabilities can be bounded by ϵ.

The difference in the ideal accept probabilities of the ground state and the first excited state is $(\cos(E_0 t) - \cos(E_1 t))/2$. Applying Taylor's theorem to $\cos(E_1 t)$ around the point $E_0 t$, we get

$$\cos(E_1 t) = \cos(E_0 t) - \sin(E_0 t) t(E_1 - E_0) - \cos(E_0 t)\frac{t^2(E_1-E_0)^2}{2} + \sin(E_0 t)\frac{h^3}{6} \quad \text{(C1)}$$

for some $h \in [0, (E_1 - E_0)t]$. Therefore,

$$\begin{aligned}\cos(E_0 t) - \cos(E_1 t) &= \sin(E_0 t) t(E_1 - E_0) + \cos(E_0 t)\frac{t^2(E_1-E_0)^2}{2} - \sin(E_0 t)\frac{h^3}{6} \\ &\ge \frac{t^2(E_1-E_0)^2}{2} - \frac{t^3(E_1-E_0)^3}{6} \\ &\ge \Omega(t^2(E_1-E_0)^2), \end{aligned} \quad \text{(C2)}$$

where in the second line we have used the fact that $E_0 t, E_1 t < \pi/2$, and in the third line we have used the fact that $(E_1 - E_0)^3 t^3 = O(t(E_1 - E_0))$. Therefore, the ideal accept probabilities also have a gap of $\Omega((E_1 - E_0)^2/\|H\|^2) = \Omega(\Delta^2/\mathsf{poly})$ as long as $\epsilon \le O(t^2\Delta^2/n) = O(\Delta^2/\mathsf{poly})$. Now, when the applied unitary differs from the ideal one by ϵ in operator norm distance, the gap in

the accept probabilities differs from the ideal accept probabilities by 2ϵ. We therefore choose ϵ sufficiently small, i.e., we choose, say, $\epsilon = \Theta(t^2(E_1 - E_0)^2/2^n)$, which is still $\Omega(2^{-\mathsf{poly}})$, as needed.

To see the existence of a promise gap, note that $E_0 \leq a$ in the YES case and $E_1 \geq b$ in the NO case, giving $c - s = \Omega(t^2(b-a)^2 - 2\epsilon) = \Omega((b-a)^2/\mathsf{poly})$. This proves the lemma. ■

As corollaries, we obtain Lemmas 8, 15, and 16, since a local Hamiltonian is also a sparse Hamiltonian.

APPENDIX D: PHASE ESTIMATION IN THE PRESENCE OF EFFICIENT CIRCUIT DESCRIPTIONS

In this section, we show that the problem (δ, Δ)-GS-DESCRIPTION-LOCALHAMILTONIAN is in GQCMA with appropriate bounds on the promise and spectral gaps (Theorems 3 to 7).

We first deal with the case of zero spectral gap.

Lemma 18 (One half of Theorems 3 and 4): *It holds that*

(a) $(1/\mathsf{poly}, 0)$-GS-DESCRIPTION-LOCALHAMILTONIAN $\in$ QCMA,
(b) $(1/\mathsf{exp}, 0)$-GS-DESCRIPTION-LOCALHAMILTONIAN $\in$ PreciseQCMA.

Proof. For the upper bound, we describe a QCMA or PreciseQCMA protocol. We are promised that in both the YES and NO cases there exists a classical description of a circuit V of polynomial size that will create a state with energy close to the ground-state energy. Specifically, the energy of this state is ϵ-close to the ground-state energy, for $\epsilon < (b-a)^3/f(n)^2$ for a polynomial $f(n) \geq \|H\|$. For QCMA, we have $b - a \geq \Omega(1/\mathsf{poly})$, while for PreciseQCMA, $b - a \geq \Omega(1/\mathsf{exp})$. The verifier asks the prover to give this description (which is promised to exist). The verifier then creates a state $|\psi\rangle$ with low energy by applying V to $|0^m\rangle$. The verifier measures the energy of this state via the one-bit phase-estimation protocol outlined in Appendix C, which involves applying a controlled-e^{-iHt} for time $t \leq \pi/(2\|H\|)$.

The proof that this verification protocol works is slightly more involved than the QMA$[c,s]$ case. This is because, in the case of QMA a verifier can assume without loss of generality that the prover sends the optimal eigenstate as a witness. However, in the case of GS-DESCRIPTION-LOCALHAMILTONIAN, we are only promised the existence of an efficient circuit to prepare a state close in energy to the ground state, and not the ground state itself [104]. Despite this complication, we can still show that a state close in energy to the ground state behaves similarly with respect to the accept probabilities of the QCMA$[c,s]$ verifier.

In the YES case, there is a description V that produces a state $|\psi\rangle$ with energy close to the ground-state energy [i.e., with energy $\leq E_1 + \epsilon < a + (b-a)^3/\mathsf{poly}(n)$]. We show in Lemma 19 below that the accept probability of the verifier upon performing one-bit phase estimation on state $|\psi\rangle$ is at least $\cos^2(bt/2) + \Omega((b-a)^2/\mathsf{poly})$. In the NO case, the optimal strategy for the prover is to send the description of a circuit that makes a state as close as possible to the ground state, since the accept probabilities are monotonic in energy and there exists no other state with smaller energy, by definition. Even if the prover sends the verifier a circuit that exactly prepares the ground state $|E_1\rangle$, its energy in the NO case is already $\geq b$. This means that the verifier will accept with probability at most $(1 + \cos E_1 t)/2 \leq (1 + \cos bt)/2$. Therefore, there is a separation in the accept probabilities in the YES and NO cases of $c - s = \Omega((b-a)^2/\mathsf{poly})$, which is $\Omega(1/\mathsf{poly})$ for $b - a = \Omega(1/\mathsf{poly})$ and $\Omega(1/\mathsf{exp})$ for $b - a = \Omega(1/\mathsf{exp})$. ■

Lemma 19: *If a state $|\psi\rangle$ has energy $\langle\psi|H|\psi\rangle = \langle E\rangle \leq E_1 + 5(b-a)^3/(24f(n)^2)$ for some polynomial $f(n) \geq \|H\|$ then in the YES case, the accept probability of the state upon phase estimation with one bit of precision is $\langle p\rangle \geq \cos^2(bt/2) + \delta$, where $\delta = \Omega(5(b-a)^2/24f(n)^2)$.*

Proof. We are given a state $|\psi\rangle$ with energy $\langle E\rangle$. Let $p_j = |\langle E_j|\psi\rangle|^2$ be the weight of the energy eigenstate E_j. Then we know that $p_1E_1 + p_2E_2 + \cdots + p_{2^n}E_{2^n} = \langle E\rangle$. The probability of accepting $|\psi\rangle$ in the one-bit phase estimation circuit is given by $\langle p\rangle = p_1\cos^2(E_1t/2) + p_2\cos^2(E_2t/2) + \cdots + p_{2^n}\cos^2(\|H\|t/2)$, where $\|H\| = E_{2^n}$. Given the constraint on the energy $\langle E\rangle$, we show in Lemma 20 below that $\langle p\rangle \geq \cos^2(E_1t/2)(1-x) + \cos^2(\|H\|t/2)x$, where $x := (\langle E\rangle - E_1)/(\|H\| - E_1)$. Now in order to have $\langle p\rangle \geq \cos^2(bt/2) + \delta$, it suffices to have

$$x \leq \frac{\cos^2(E_1t/2) - \cos^2(bt/2) - \delta}{\cos^2(E_1t/2) - \cos^2(\|H\|t/2)} = \frac{\cos(E_1t) - \cos(bt) - 2\delta}{\cos(E_1t) - \cos(\|H\|t)}. \tag{D1}$$

It is therefore sufficient if

$$x \leq \frac{(b-a)t}{2}\left(bt - \frac{b^3t^3}{6}\right) - \delta, \tag{D2}$$

since

$$\frac{(b-a)t}{2}\left(bt-\frac{b^3t^3}{6}\right)-\delta \le \frac{(b-a)t\sin(bt)-2\delta}{2} \le \frac{(b-a)t\sin(bt)-2\delta}{\cos(at)-\cos(\|H\|t)} \le \frac{\cos(at)-\cos(bt)-2\delta}{\cos(at)-\cos(\|H\|t)} \le \frac{\cos(at)-\cos(bt)-2\delta}{\cos(E_1t)-\cos(\|H\|t)}, \tag{D3}$$

where we have used the inequalities $\sin(bt) \ge bt - b^3t^3/6$, $E_1 \le a$, $\cos(at)-\cos(bt) \ge (b-a)t\sin(bt)$, and $2 \le \cos(at)-\cos(\|H\|t)$. We now require $\delta \ge (b-a)t\sin(bt)/4$, so that condition (D2) translates to $x \le (b-a)t(bt-b^3t^3/6)/4$.

Let us choose $t = \min[1/f(n), 1/b] = 1/f(n)$, since otherwise $b \ge f(n) \ge \|H\|$ and the instance is trivial. We thus know that $\|H\|t \le 1 \le \pi/2$, $t \ge 1/f(n)$, and $bt < 1$. We also assume that in the YES case $\|H\| - E_1 \ge b-a$. This is because otherwise a verifier can compute $\mathrm{Tr}(H)/2^n$ efficiently given the Hamiltonian and accept straightaway if $\mathrm{Tr}(H)/2^n \le b$. This works since $E_1 \le \mathrm{Tr}(H)/2^n$, and by the promise, $E_1 \le b \implies E_1 \le a$. Therefore, without loss of generality, one can assume that the nontrivial instances satisfy $b \le \mathrm{Tr}(H)/2^n \le \|H\|$, or $\|H\| - E_1 \ge b-a$.

Therefore, since $\langle E\rangle \le E_1 + 5(b-a)^3/(24f(n)^2)$, we have

$$\langle E\rangle \le E_1 + \frac{5(b-a)^2(\|H\|-E_1)}{24f(n)^2} \tag{D4}$$

$$\implies x \le \frac{(b-a)^2}{4f(n)^2}\frac{5}{6} = \frac{(b-a)^2}{4f(n)^2}\left(1-\frac{1}{6}\right) \le \frac{(b-a)^2}{4f(n)^2}\left(1-\frac{b^2t^2}{6}\right) \le \frac{(b-a)}{4f(n)}\frac{b}{f(n)}\left(1-\frac{b^2t^2}{6}\right) \le \frac{(b-a)t}{4}\left(bt-\frac{b^3t^3}{6}\right), \tag{D5}$$

as required. To sum up, we have shown that $\langle E\rangle \le E_1 + 5(b-a)^3/(24f(n)^2)$ implies that $\delta \ge (b-a)t\sin(bt)/4 \ge (b-a)^2(1-b^2t^2/6)/(4f(n)^2) \ge 5(b-a)^2/(24f(n)^2)$. ■

Lemma 20: *For probabilities $p_j, j \in [2^n]$, satisfying $\sum_j p_j E_j \le \langle E\rangle$ and numbers $E_1 \le E_2 \le \cdots \le E_{2^n}$ satisfying $E_j t \in [0, \pi/2]$, the quantity $\sum_j p_j \cos^2(E_j t/2)$ is bounded below by $\cos^2(E_1t/2)(1-x) + \cos^2(E_{2^n}t/2)x$, where x is given by $(\langle E\rangle - E_1)/(E_{2^n}-E_1)$.*

Proof. Since the function $f(x) = -\cos^2(xt/2)$ is convex for $xt/2 \in [0, \pi/2)$, we have

$$\frac{f(E_1)(E_{2^n}-E_j)+f(E_{2^n})(E_j-E_1)}{E_{2^n}-E_1} \ge f(E_j). \tag{D6}$$

Therefore,

$$p_j\frac{f(E_1)(E_{2^n}-E_j)+f(E_{2^n})(E_j-E_1)}{E_{2^n}-E_1} \ge p_j f(E_j) \tag{D7}$$

$$\implies \frac{f(E_1)(E_{2^n}-\langle E\rangle)+f(E_{2^n})(\langle E\rangle-E_1)}{E_{2^n}-E_1} \ge \sum_j p_j f(E_j) \tag{D8}$$

$$\implies \sum_j p_j \cos^2\left(\frac{E_j t}{2}\right) \ge \cos^2\left(\frac{E_1 t}{2}\right)\frac{E_{2^n}-\langle E\rangle}{E_{2^n}-E_1} + \cos^2\left(\frac{E_{2^n}t}{2}\right)\frac{\langle E\rangle - E_1}{E_{2^n}-E_1}, \tag{D9}$$

which completes the proof. ■

We now turn to the cases where in addition to the promise of an efficient circuit to prepare a low-energy state, the Hamiltonian is promised to have a spectral gap Δ. For this case, we can show the following result.

Lemma 21: *Problem* GS-DESCRIPTION-LOCALHAMILTONIAN$[a,b,g_1,g_2] \in$ GQCMA$[c,s,g_1',g_2']$ *for* $c-s = \Omega((b-a)^2/f(n)^2)$ *and* $\min[g_1',g_2'] \ge 5\Delta^2/(36f(n))$, *where $f(n)$ is a polynomial upper bound to $\|H\|$, and* $\Delta = \min[g_1,g_2] \ge (b-a)^3/f(n)^2$.

Proof. We analyze the same algorithm as the nongapped case and show that the verification protocol, with slight modifications, preserves the spectral gap. In particular, in the first step of the original protocol, the verifier straightaway accepts if $\mathrm{Tr}(H)/2^n \le b$ or if the upper bound to the norm of the Hamiltonian, $f(n)$, satisfies $f(n) \le b$. We modify this to requiring the verifier to accept only if, in addition to the previous conditions, measurement of the witness register yields the all-zeroes string 0^w (where w is the size of the witness register). This has the effect of creating a spectral gap, since in this case only the all-zeroes state is accepted and all other computational basis states are rejected.

If the first step does not cause the verifier to accept, the verifier assumes that the witness state is a description of the circuit V to prepare a low-energy state $|\psi\rangle$. The verifier then proceeds to prepare this state and measure its energy using the one-bit phase-estimation protocol. As shown in

the proof of Lemma 18, the protocol has a promise gap $c - s = \Omega\left((b-a)^2/f(n)^2\right)$.

We now analyze the spectral gap. Let us denote by y the quantity $(\langle E\rangle - E_1)/(E_2 - E_1)$ and by x the quantity $(\langle E\rangle - E_1)/(E_{2^n} - E_1) \leq y$. Any state with energy $\langle E\rangle := \langle H\rangle\psi \leq E_1 + (b-a)^3/f(n)^2 \leq E_1 + \Delta$ has a large overlap with the ground state:

$$|\langle\psi|E_1\rangle|^2 \geq 1 - \frac{\langle E\rangle - E_1}{E_2 - E_1} = 1 - y. \tag{D10}$$

Therefore, any state $|\phi\rangle$ orthogonal to $|\psi\rangle$ must have an overlap with the ground state that satisfies $|\langle\phi|E_1\rangle|^2 \leq y$. This means that the accept probability for any witness orthogonal to that corresponding to the ground-state description is

$$\begin{aligned}\langle p_\phi\rangle &= \sum_j p_j \cos^2\left(\frac{E_j t}{2}\right)\\ &\leq y\cos^2\left(\frac{E_1 t}{2}\right) + (1-y)\cos^2\left(\frac{E_2 t}{2}\right).\end{aligned} \tag{D11}$$

On the other hand, the accept probability of the optimal witness is at least (Lemma 19)

$$\langle p_\psi\rangle \geq (1-x)\cos^2\left(\frac{E_1 t}{2}\right) + x\cos^2\left(\frac{E_2 t}{2}\right). \tag{D12}$$

The difference in these two is a lower bound for the spectral gap of the accept operator:

$$\begin{aligned}g_1, g_2 \geq& \langle p_\psi\rangle - \langle p_\phi\rangle\\ &\geq \cos^2\left(\frac{E_1 t}{2}\right)(1-x-y)\\ &\quad + \cos^2\left(\frac{E_2 t}{2}\right)(x+y-1)\\ &= \frac{1-x-y}{2}(\cos(E_1 t) - \cos(E_2 t))\\ &\geq \frac{1-2y}{2}(E_2 - E_1)\sin(E_2 t).\end{aligned} \tag{D13}$$

Now, we know from the promise that $y = (\langle E\rangle - E_1)/(E_2 - E_1) \leq (b-a)^3/f(n)^2\Delta \leq 1/3$, and $E_2 \geq E_1 + \Delta \geq \Delta$. Also, we have chosen $t \geq 1/f(n)$ for a polynomial $f(n) \geq \|H\|$. Therefore,

$$\begin{aligned}\min[g_1', g_2'] &\geq \frac{\Delta}{6}\sin(\Delta t)\\ &\geq \frac{\Delta}{6}\left(\Delta t - \frac{\Delta^3 t^3}{6}\right)\\ &\geq \frac{\Delta}{6}\left(\frac{\Delta}{f(n)} - \frac{\Delta^3}{6f(n)^3}\right)\\ &= \frac{\Delta^2}{6f(n)}\left(1 - \frac{\Delta^2}{6f(n)^2}\right)\\ &\geq \frac{5\Delta^2}{36f(n)},\end{aligned} \tag{D14}$$

since $\Delta \leq \|H\| \leq f(n)$. ■

This proves the following results.

Corollary 6 (One half of Theorems 5 to 7): *It holds that*

(a) $(1/\mathsf{poly}, 1/\mathsf{poly})$-GS-DESCRIPTION-LOCALHAMILTONIAN $\in$ PGQCMA,
(b) $(1/\mathsf{exp}, 1/\mathsf{exp})$-GS-DESCRIPTION-LOCALHAMILTONIAN $\in$ PreciseEGQCMA,
(c) $(1/\mathsf{exp}, 1/\mathsf{poly})$-GS-DESCRIPTION-LOCALHAMILTONIAN $\in$ PrecisePGQCMA.

APPENDIX E: DETAILS OF THE PP ALGORITHM

In this section we complete the proof of Lemma 2 by expanding upon the PP algorithm. We also prove Lemma 3 by giving a P^{PP} algorithm to precisely compute ground-state local observables of $\Omega(1/\mathsf{poly})$-spectral-gapped Hamiltonians.

Lemma 22: *A* PP *algorithm can decide whether* $Tr[Q^q A] \leq a'$ *or* $\geq b'$ *when input thresholds* a' *and* b' *for matrices* Q *and* A *of size* $2^{\mathsf{poly}(n)} \times 2^{\mathsf{poly}(n)}$ *satisfy the following properties (we use the symbol* R *to denote both matrices* Q *and* A *in the following).*

1. *The norm of matrix* R *is upper bounded by a polynomial in* n.
2. *Matrix* R *may be written as a polynomial of degree* $d = \mathsf{poly}(n)$ *in terms of matrices* R_i, $i \in [m]$, *in the computational basis for* $m = \mathsf{poly}(n)$, *such that*

   (a) the matrix elements of each matrix R_i *are computable to precision* δ *in time polynomial in* n *and* $\log(1/\delta)$.

Proof. The quantity $\mathrm{Tr}(Q^q A)$ may be expressed as

$$\sum_x \langle Q^q A\rangle x = \sum_x \sum_{x_1,x_2,\ldots,x_q} \langle x|Q\,|x_1\rangle\langle x_1|\,Q|x_2\rangle \times \cdots \times \langle x_{q-1}|Q\,|x_q\rangle\langle x_q|\,A|x\rangle. \tag{E1}$$

If Q is a polynomial of degree d in terms of matrices $R_1,\ldots,R_m$ for $m = \mathsf{poly}(n)$, we can write it as

$$Q = \sum_{\substack{i_1,i_2,\ldots,i_m\in[d]\\ i_1+i_2+\cdots+i_m\le d}} p_{i_1 i_2\cdots i_m} R_1^{i_1} R_2^{i_2}\cdots R_m^{i_m}, \tag{E2}$$

where each tuple $(i_1,\ldots,i_m)$ specifies a monomial. The number of terms in the polynomial is bounded above by $(d+1)^m = \mathsf{exp}[m\ \ \log(d+1)] = O(\mathsf{exp}[\mathsf{poly}(n)])$. We write a term of Eq. (E2) below, $\langle x_j|Q|x_{j+1}\rangle$, as

$$\langle x_j|Q|x_{j+1}\rangle = \sum_{\substack{i_1,i_2,\ldots,i_m\in[d]\\ i_1+i_2+\cdots+i_m\le d}} p_{i_1 i_2\cdots i_m}\langle x_j|R_1^{i_1}|z_{j,1}\rangle \times \langle z_{j,1}|R_2^{i_2}\,|z_{j,2}\rangle\langle z_{j,2}|\cdots\langle z_{j,m-1}|R_m^{i_m}|x_{j+1}\rangle. \tag{E3}$$

We can further insert resolutions of the identity in Eq. (E3) to get a sum over yet more terms. Each term in the resulting sum is a product over polynomially many quantities of the form $\langle w_1|R_s|w_2\rangle$ for some computational basis states $|w_1\rangle, |w_2\rangle$ and an index $s\in[m]$. Each of these can be computed in polynomial time. The number of terms in the final sum of the form in Eq. (E2) below is still bounded above by 2^{poly}.

From the assumption, the matrix elements of matrices R_i can be computed to additive error $2^{-g(n)}$ in time scaling as $O(g(n))$. We therefore choose $g(n)$ to be such that the total additive error resulting from the 2^{poly} many paths in Eq. (E2) below is negligible compared to $(b'-a')\times 2^{\mathsf{poly}}$, where the second term (2^{poly}) corresponds to the number of terms in the sum. This can be ensured by taking $g(n)$ to be a sufficiently large polynomial.

Equation (E2) below is a sum over $T = O(2^{\mathsf{poly}})$ many terms f_i, each of which may be computed in polynomial time. Each term of Eq. (E2) may be interpreted as a path in a Turing machine. Therefore, a PP machine can decide whether $\sum_{i=1}^T f_i$ is $\le a'$ or $\ge b'$ for some thresholds $a', b' \ge a' + \Omega(2^{-\mathsf{poly}})$ input to the PP machine. This is seen as follows. Each term f_i is an efficiently computable real-valued function of the trajectory $x_0^i, x_1^i, \ldots, x_K^i$. Let $a_{\max}$ be an upper bound to the norm of A. The PP machine selects a uniformly random trajectory and computes f_i. It accepts with probability $\frac{1}{2} - f_i/(2^{n+1}a_{\max}) > 0$ and rejects otherwise. The overall acceptance probability is $[1/T]\sum_i(\frac{1}{2} - f_i/(2^{n+1}a_{\max}))$. In the YES case, this is at least $\frac{1}{2} - a'/(2^{n+1}Ta_{\max})$, while in the NO case, it is at most $\frac{1}{2} - b'/(2^{n+1}Ta_{\max})$. Since we at least have a separation of $2^{-n-1}/T\times\Omega(b'-a') = \Omega(2^{-\mathsf{poly}(n)})$ between the YES and NO instances, this is a valid PP algorithm. ■

Lemma 22 applies to the proof of Lemma 2 because the accept operator Q in that proof is a degree-$(2T+3)$ polynomial in matrices with efficiently computable entries.

For the proof of Lemma 3, we show in Lemma 23 below that beginning from the maximally mixed initial state, imaginary time evolution for "time" $-i\beta$ produces a thermal state with high enough overlap with the ground state for a suitable β. Computing local observables in the obtained thermal state then suffices to get exponentially good estimates of ground-state local observables for gapped systems. We make the choice of a maximally mixed initial state in the above because it is guaranteed to have at least overlap 2^{-n} with the ground state.

Lemma 23: *For a Hamiltonian H with spectral gap at least Δ, let ρ_β be the thermal state at temperature $1/(2\beta)$. Also, let $|E_1\rangle$ be the ground state of H, and let A be any local observable satisfying $\|A\|\le\mathsf{poly}(n)$. Then, for $\beta=\Omega(n\Delta^{-1})$, the thermal expectation value satisfies $|Tr[\rho_\beta A] - \langle E_1|A|E_1\rangle| \le 2^{-\mathsf{poly}}$.*

Proof. Let the eigenstates of the Hamiltonian be given by $|E_i\rangle$, $i\in[2^n]$, with the eigenvalues E_i arranged in nondecreasing order. Consider the initial state $\rho = \mathbb{1}/2^n$ and apply the linear operation $\mathsf{exp}(-\beta H)$, which performs imaginary time evolution for "time" $-i\beta$:

$$\rho \to \rho' = \mathsf{exp}(-\beta H)\rho\ \mathsf{exp}(-\beta H) \tag{E4}$$

up to normalization. The maximally mixed initial state $\rho = \mathbb{1}/2^n = \sum_i |E_i\rangle\langle E_i|/2^n$ transforms to the state ρ_β, given by

$$\begin{aligned}\rho_\beta &= \frac{\rho'}{\mathcal{N}}\\ &= \frac{1}{\mathcal{N}} e^{-\beta H}\sum_i \frac{1}{2^n}|E_i\rangle\langle E_i|\,e^{-\beta H}\\ &= \frac{1}{2^n\mathcal{N}}\sum_i e^{-2\beta E_i}|E_i\rangle\langle E_i|\,.\end{aligned} \tag{E5}$$

This state is the same as the thermal state $e^{-2\beta H}$ at temperature $1/(2\beta)$ up to normalization. The normalization factor $\mathcal{N} = \mathrm{Tr}\rho'$ is given by $\sum_i e^{-2\beta E_i}/2^n$. The overlap of the

normalized state with the ground state is thus

$$\begin{aligned}\mathrm{Tr}[\rho_\beta\,|E_1\rangle\langle E_1|] &= \frac{e^{-2\beta E_1}}{2^n\mathcal{N}}\\ &= \frac{e^{-2\beta E_1}}{\sum_i e^{-2\beta E_i}}\\ &= \left(1+\sum_{i\neq 1} e^{-2\beta(E_i-E_1)}\right)^{-1}. \end{aligned} \tag{E6}$$

Since $E_i - E_1 = \Delta = \Omega(1/n^c)$, if β is taken to be $\Omega(n^d)$ with $d \geq c+1$, we have $e^{-2\beta(E_i-E_1)} \leq \mathsf{exp}[-2n^{d-c}]$. This means that the overlap is at least $1/(1+\mathsf{exp}[n\log 2 - 2n^{d-c}]) \geq 1 - \mathsf{exp}[n\log 2 - 2n^{d-c}]$, which means that the trace distance between the normalized states is $\varepsilon = O(\mathsf{exp}[-n^{d-c}])$. Therefore, the choice $\beta = \Theta(n\Delta^{-1})$ suffices to ensure that the resulting (normalized) state ρ_β is exponentially close to the ground state. Therefore, the thermal expectation value of any local observable A with polynomially bounded spectral norm is also exponentially close to the ground-state expectation value $\langle E_1|A|E_1\rangle$. ■

We now show the following lemma about computing unnormalized thermal expectation values, which is the core subroutine of our P^{PP} algorithm.

Lemma 24: *A* PP *algorithm can decide whether* $\mathrm{Tr}[e^{-2\beta H}A] \leq a'$ *or* $\geq b'$ *when given as input a sparse Hamiltonian* H*, a number* $\beta \geq 0$*, a local observable* A*, and thresholds* a' *and* b'*.*

Proof. We express the unnormalized thermal expectation value as a sum over several paths as follows:

$$\begin{aligned}\mathcal{A}_\beta &= \mathrm{Tr}[e^{-2\beta H}A] \qquad\qquad (\text{E7})\\ &= \sum_{x,y}\langle x|e^{-2\beta H}\,|y\rangle\langle y|\,A|x\rangle\\ &\approx \sum_{x,y}\langle x|\left(\mathbb{1} - 2\beta H + 2(\beta H)^2 + \cdots + \frac{(-2\beta H)^K}{K!}\right)\\ &\quad\times |y\rangle\langle y|A|x\rangle =: \mathcal{A}'_\beta \qquad\qquad (\text{E8})\\ &= \sum_{k=0}^{K}\frac{1}{k!}\sum_{x_0,x_1,\ldots,x_k}\langle x_0| - 2\beta H\,|x_1\rangle\langle x_1| - 2\beta H|x_2\rangle\\ &\quad\times\cdots\times\langle x_{k-1}| - 2\beta H\,|x_k\rangle\langle x_k|\,A|x_0\rangle. \qquad\qquad (\text{E9})\\ &= \sum_{i=1}^{T} f_i. \end{aligned}$$

This expression is reminiscent of a Euclidean path integral, although there are some differences. In a Euclidean path integral, one Trotterizes the map $\mathsf{exp}(-\beta H) \approx (\prod_i \mathsf{exp}(-\beta H_i/r))^r$ and uses the fact that each term of Hamiltonian H_i is local in order to compute terms in the series. In contrast, here we have used the Taylor expansion for $\mathsf{exp}(-\beta H)$ and have inserted resolutions of the identity in order to compute the terms $\langle x|H^k|y\rangle$. Using the Taylor series allows us to get exponentially small additive error, which is not guaranteed by Trotterization.

Before we move on, let us analyze the additive error in Eq. (E8). It is given by

$$\epsilon \leq \frac{(2\beta\|H\|)^{K+1}}{(K+1)!}\|A\| \times O(1). \tag{E10}$$

By choosing $K > 2\beta e\|H\| + f(n)$ for some polynomial $f(n) = O(\beta\|H\|/n)$ and $f(n) = \Omega(n)$, we can ensure that the error is bounded above by $\|A\|\mathsf{exp}[-f(n)]$:

$$\begin{aligned} & K+1 \geq 2\beta e\|H\| + f(n) \qquad\qquad (\text{E11})\\ \implies\quad & (K+1)\log(K+1)\\ & \geq (K+1)\log(2\beta e\|H\|)\\ & \quad + (K+1)\log\left(1+\frac{f(n)}{2\beta e\|H\|}\right)\\ & \geq (K+1)\log(2\beta e\|H\|) + (K+1)\frac{f(n)}{2\beta e\|H\|}\\ & \quad - \frac{K+1}{2}\left(\frac{f(n)}{2\beta e\|H\|}\right)^2\\ & \geq (K+1)\log(2\beta e\|H\|) + \Omega(f(n)). \end{aligned}$$

Here we have used the facts that $\log(1+x) \geq x - x^2/2$ for small x and that $f(n) = o(\beta\|H\|)$. Therefore,

$$\log\left(\frac{(2\beta\|H\|)^{k+1}}{(K+1)!}\right) \leq -\Omega(f(n)), \tag{E12}$$

$$\text{giving}\quad \epsilon \leq O(\|A\|\mathsf{exp}[-f(n)]). \tag{E13}$$

This completes the proof. ■

Proof of Lemma 3. From Lemma 23, we know that the normalized state is exponentially close to the true ground state. Therefore, deciding whether the ground state has $\mathrm{Tr}[|\Psi\rangle\langle\Psi|\,A] \leq a$ or $\geq b$ is equivalent to deciding whether the unnormalized state has expectation value $\mathrm{Tr}[\rho' A] \leq a' = \mathcal{N}_{\mathrm{est}}(a + \|A\|\varepsilon)$ or $\geq b' = \mathcal{N}_{\mathrm{est}}(b - \|A\|\varepsilon)$, where $\mathcal{N}_{\mathrm{est}}$ is an estimate of the normalization of the state and ε the trace distance between the ground state and the thermal state. To maintain a gap between the YES and NO cases, we need $\varepsilon < 2^{-u(n)}/\|A\|$ for some polynomial u, which can be satisfied by taking n^{d-c} in Lemma 23 to be $\geq u(n) + \log\|A\|$. The norm of A is bounded above by a polynomial in n and therefore is a subleading term.

Since the thresholds a' and b' depend on the normalization, we should compute the normalization $\mathcal{N}$ beforehand. Since the normalization is a special case of Eq. (E7) with $A = \mathbb{1}$, we can use the PP procedure to decide if $\mathcal{N} \leq a_1$ or $\mathcal{N} \geq a_2$ for some a_1, a_2 with $a_2 - a_1 = \Omega(1/\mathrm{exp})$. Performing binary search over the interval $(0, 1]$ with polynomially many queries to the PP oracle, we can estimate the normalization to exponentially small additive error, giving an estimate $\mathcal{N}_{\mathrm{est}}$.

Therefore, we have shown that a P^{PP} machine can do all the above: compute the normalization and then compute the thermal expectation value for a low-temperature state. Since we have also shown that setting $\beta = (n/\Delta)$ suffices to get exponentially small error, we have shown that the problem is in P^{PP}. ■

This technique is also applicable to Hamiltonians or Hermitian operators that are not necessarily local, or even sparse. For example, it can apply to Hermitian operators of the kind in Lemma 22.

APPENDIX F: TURING MACHINE CONSTRUCTION FOR PSPACE-HARDNESS

In this section, we complete the proof of Lemma 7.

Lemma 25 (Lower bound on spectral gap for the PSPACE-hard construction): *In the NO case, the construction in the proof of Lemma 7 has a spectral gap of* $\Theta(\ell_{\max}^{-2})$*, where* $\ell_{\max}$ *is the number of vertices in the largest subgraph of* G'_x.

Proof. Recall the form of the graph G'_x in the NO case, reproduced here in Fig. 4(a). We first restrict our attention to the subgraph of G'_x containing the start and accept configurations. The matrix $A_x^{\dagger\prime} A'_x$, when restricted to this subspace, is further composed of three subspaces, each corresponding to a subgraph, as shown in Fig. 4(b). We write $A_x^{\dagger\prime} A'_x = \mathcal{G}_1 \oplus \mathcal{G}_2 \oplus \mathcal{G}_3$. The block $\mathcal{G}_1$ corresponds to the vertices leading to t_x (not including t_x). The block $\mathcal{G}_2$ corresponds to the vertices $\{t_x\} \cup \{1, \ldots, t(n)\}$. Lastly, $\mathcal{G}_3$ is the block with the vertices starting from s_x and leading to the reject state, which are the configurations visited by the Turing machine. We have

$$\mathcal{G}_1 = \begin{pmatrix} 2 & 1 & & & & \\ 1 & 2 & 1 & & & \\ & 1 & 2 & \ddots & & \\ & & \ddots & \ddots & 1 & \\ & & & 1 & 2 & \end{pmatrix}_{\ell_1 \times \ell_1},$$

$$\mathcal{G}_2 = \mathbb{1}_{\ell_2 \times \ell_2},$$

$$\text{and} \quad \mathcal{G}_3 = \begin{pmatrix} 1 & 1 & & & & \\ 1 & 2 & 1 & & & \\ & 1 & 2 & \ddots & & \\ & & \ddots & \ddots & 1 & \\ & & & 1 & 2 & 1 \\ & & & & 1 & 1 \end{pmatrix}_{\ell_3 \times \ell_3}.$$

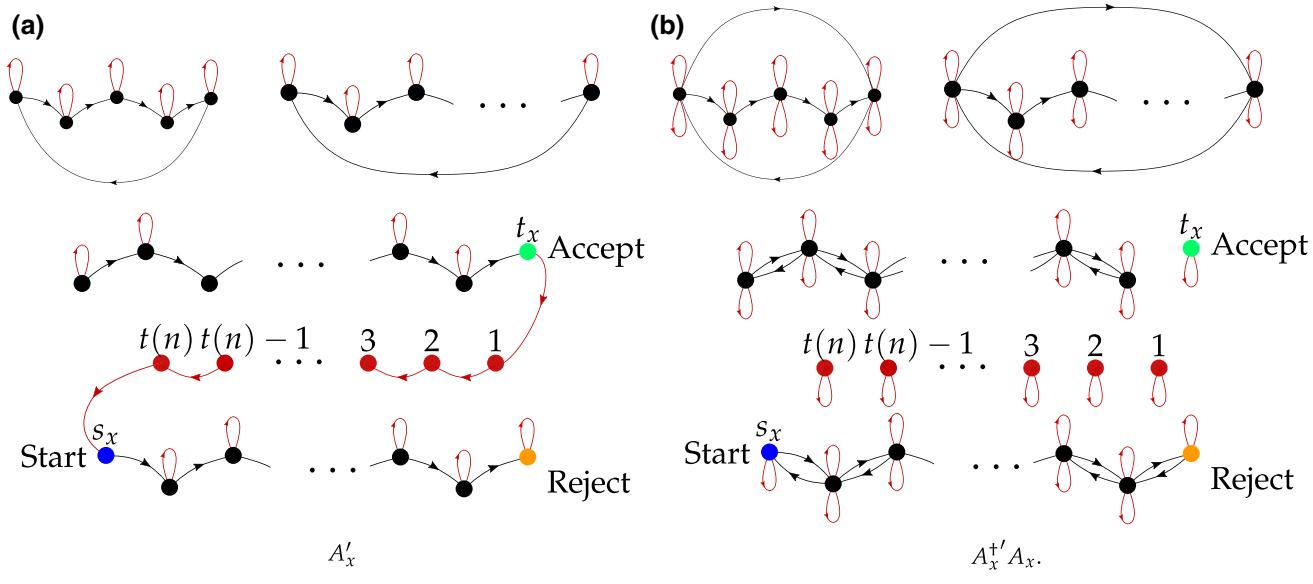

FIG. 4. (a) Graph G'_x with adjacency matrix A'_x, adapted from Fig. 2(d). (b) Graph with (weighted, directed) adjacency matrix $A_x^{\dagger\prime} A_x$. Vertices with two self-loops can be thought of as a single self-loop with weight 2.

It may be seen that there is a zero eigenvector $(0,0,\ldots,0,1,-1,1,\ldots,(-1)^{\ell_3})^T$, with the zeros corresponding to subspaces $\mathcal{G}_1$ and $\mathcal{G}_2$. We now lower bound the next-smallest eigenvalue. Let

$$r_n(\lambda) := \det[\mathcal{G}_3 - \lambda \mathbb{1}_n] = \det \begin{pmatrix} 1-\lambda & 1 & & & & & \\ 1 & 2-\lambda & 1 & & & & \\ & 1 & 2-\lambda & \ddots & & & \\ & & \ddots & \ddots & 1 & & \\ & & & 1 & 2-\lambda & 1 & \\ & & & & 1 & 1-\lambda \end{pmatrix}_{n\times n}, \tag{F1}$$

$$p_n(\lambda) := \det \begin{pmatrix} 2-\lambda & 1 & & & & & \\ 1 & 2-\lambda & 1 & & & & \\ & 1 & 2-\lambda & \ddots & & & \\ & & \ddots & \ddots & 1 & & \\ & & & 1 & 2-\lambda & 1 & \\ & & & & 1 & 1-\lambda \end{pmatrix}_{n\times n}. \tag{F2}$$

The polynomial $p_n(\lambda)$ can be computed exactly [91], and is given by $p_n(2-2\cos\theta) = (\sin((n+1)\theta) - \sin(n\theta))/\sin\theta = \cos((n+\frac{1}{2})\theta)/\cos(\theta/2)$. We can obtain $r_n(\lambda)$ in terms of $p_n(\lambda)$, $r_n(\lambda) = (1-\lambda)p_{n-1}(\lambda) - p_{n-2}(\lambda)$, giving us

$$r_n(\lambda) = f_n(\theta) = (2\cos\theta - 1)\frac{\cos((n-1/2)\theta)}{\cos(\theta/2)} - \frac{\cos((n-3/2)\theta)}{\cos(\theta/2)}, \tag{F3}$$

where $\theta = \cos^{-1}(1-\lambda/2)$, or $\lambda = 2-2\cos\theta$. The eigenvalues of $\mathcal{G}_3$ are related to the roots of the characteristic polynomial $f_n(\theta) = 0$. We can see that $\theta = 0$ is always a root of the polynomial, giving us the zero eigenvalue ($\lambda = 2-2\cos\theta = 0$) for the NO case.

Now, it remains to be shown that the next smallest eigenvalue is bounded away from zero. First consider $\mathcal{G}_1$, whose eigenvalues are the roots of the characteristic equation $\det[\mathcal{G}_1 - \lambda\mathbb{1}_{\ell_1}]$. The eigenvalues of $\mathcal{G}_1$ can be computed in a similar fashion to those of $\mathcal{G}_3$ and are given by $4\sin^2(k\pi/2(\ell_1+1))$, $k \in [n]$. The smallest eigenvalue of $\mathcal{G}_1$ is therefore at least $\Omega(1/{\ell_1}^2)$. It is also easily seen that $\mathcal{G}_2 \succ 0$.

We now come to $\mathcal{G}_3$. As we have seen, $\mathcal{G}_3$ has a zero eigenvalue. In order to show a spectral gap for $\mathcal{G}_3$, we show that the next root of polynomial $f_{\ell_3}(\theta)$ must occur at least a distance $\Omega({\ell_3}^{-2})$ away. The roots of $\mathcal{G}_3$ are given by [105]

$$\lambda_j = 2 + 2\cos\left(\frac{\pi j}{\ell_3}\right), \qquad j \in [\ell_3]. \tag{F4}$$

Setting $j = \ell_3$ gives the zero eigenvalue and $j = \ell_3 - 1$ the first nonzero eigenvalue. The spectral gap of $\mathcal{G}_3$ is therefore

$$\begin{aligned} \lambda_{\ell_3-1} &= 2 - 2\cos\left(\frac{\pi}{\ell_3}\right) \\ &= 4\sin^2\left(\frac{\pi}{2\ell_3}\right) \\ &\geq \frac{\pi^2}{\ell_3^2} - O\left(\frac{\pi^4}{\ell_3^4}\right) \\ &= \Omega(1/\ell_3^2). \end{aligned} \tag{F5}$$

Finally, we consider other subgraphs that do not contain the start vertex. Just like the analysis of the YES case, the eigenvalues for these are bounded away from 0 by ℓ^{-2}, where ℓ is the number of vertices in the subgraph. We have therefore lower bounded the value of the nonzero eigenvalue in each case, showing that the spectral gap is $\Omega(\ell_{\max}^{-2}) = \Omega(2^{-\mathsf{poly}})$. ■

APPENDIX G: COMPLEXITY OF PrecisePGQMA AND PreciseEGQMA WITH ASYMMETRIC SPECTRAL GAPS

We show here that the promise of asymmetric spectral gaps does not change the complexity class for both PrecisePGQMA and PreciseEGQMA, proving Theorem 11.

Proof of Theorem 11. It is easy to see that $\mathsf{GQMA}[c,s,g_1,g_2] \subseteq \mathsf{GQMA}[c,s,g_1,0]$ simply by ignoring the promise on the NO instance. It remains to show that the same upper bounds as the symmetric case hold for the asymmetric case too. For the case of $c-s = \Omega(1/\mathsf{exp})$, $g_2 = \Omega(1/\mathsf{exp})$, we observe that one can also ignore the promise on the YES instance and obtain containment in $\mathsf{PreciseQMA} = \mathsf{PSPACE}$, which equals PreciseEGQMA.

It remains to give an upper bound for the class $\bigcup_{c-s\geq\Omega(1/\mathsf{exp}),g_1\geq\Omega(1/\mathsf{poly})} \mathsf{GQMA}[c,s,g_1,0]$. We give a PP algorithm for any instance from this class, which implies equivalence of the two classes.

We are given a description of a circuit, with the promise that the YES case has $\Omega(1/\mathsf{poly})$ spectral gap for the accept operator Q. We want to decide if $\lambda_1(Q)$ is $\geq c$ (YES) or $\leq s$ (NO). The overall PP algorithm is as follows.

1. Use the $\mathsf{P}^{\mathsf{QMA[log]}}$ algorithm of Ambainis [94] to determine whether an instance has spectral gap $\Delta \geq g_1$ (YES) or $\leq g_1/2$ (NO) for $g_1 = \Omega(1/\mathsf{poly})$.
2. If the spectral gap is g_1 or larger, run the algorithm in Lemma 22 with Hamiltonian $\mathbb{1} - Q$ and accept or reject according to the answer returned by the algorithm.
3. Otherwise, reject.

We claim that the algorithm of Ambainis works not just for local Hamiltonians, but also for accept operators like Q. This is because the QMA queries in Ambainis's algorithm pertain to whether the ground-state energy (or the minimum eigenvalue $1-\lambda_1$ in this case) is smaller or larger than a threshold. A QMA verifier can compute the eigenvalue of the accept operator given an eigenstate, using phase estimation. Therefore, all queries to the oracle about $1-\lambda_1$ are still valid QMA queries. Also, the final query in Ambainis's algorithm is for the operator $(\mathbb{1}-Q)\otimes\mathbb{1}+\mathbb{1}\otimes(\mathbb{1}-Q)$ on two registers, restricted to the antisymmetric subspace. Since a QMA verifier can also perform a projection onto the antisymmetric subspace, Ambainis's algorithm (i.e., the first step) works to estimate the spectral gap of Q in $\mathsf{P}^{\mathsf{QMA}[\log]}$.

Now, since $\mathsf{P}^{\mathsf{QMA}[\log]}\subseteq\mathsf{PP}$ [95], the overall algorithm is a valid PP algorithm, since the two queries can be made in parallel. To see the correctness, we see that if the instance has a YES answer then it has a spectral gap of at least g_1 by virtue of the promise. In this case the spectral gap algorithm would return YES. This ensures that the PP algorithm in Lemma 22 works correctly and returns the correct answer $E_1\le a$ (YES) or $E_1\ge b$ (NO). The algorithm outputs YES since the instance has low energy.

In the NO case, there may or may not be a spectral gap. If the spectral gap $\Delta\le g_1/2$ is not large enough, the spectral gap algorithm returns NO. We reject in this case. If the spectral gap algorithm returns YES then the spectral gap is at least $\Delta\ge g_1/2$ (this includes the cases when the spectral gap is in the window $[g_1/2,g_1]$, which is outside of the promise in the spectral gap algorithm). This means that the algorithm in Appendix E will work, and return the correct output (NO). Therefore, we see that $\bigcup_{c-s\ge\Omega(1/\mathsf{exp}),\,g_1\ge\Omega(1/\mathsf{poly})}\mathsf{GQMA}[c,s,g_1,0]=\mathsf{PrecisePGQMA}$. ■

We remark that it can be seen that LOCALHAMILTONIAN $[a,b,g_1,0]$ with $b-a=\Theta(1/\mathsf{exp})$ is PrecisePGQMA-complete when the spectral gap g_1 is $1/\mathsf{poly}$ and PreciseEGQMA-complete when g_1 is $1/\mathsf{exp}$.

[1] S. Gharibian, Y. Huang, Z. Landau, and S. W. Shin, Quantum Hamiltonian complexity, Found. Trends® Theor. Comput. Sci. **10**, 159 (2015).

[2] A. Kitaev, A. Shen, and M. Vyalyi, *Classical and Quantum Computation*, Graduate Studies in Mathematics (American Mathematical Society, Providence, Rhode Island, 2002), Vol. 47.

[3] S. A. Cook, in *Proceedings of the Third Annual ACM Symposium on Theory of Computing - STOC '71* (ACM Press, Shaker Heights, Ohio, United States, 1971), p. 151.

[4] B. Trakhtenbrot, A survey of Russian approaches to perebor (brute-force searches) algorithms, IEEE Annals Hist. Comput. **6**, 384 (1984).

[5] D. Aharonov and T. Naveh, Quantum NP - A survey, arXiv:quant-ph/0210077 (2002).

[6] D. Aharonov, I. Arad, Z. Landau, and U. Vazirani, in *Proceedings of the 41st Annual ACM Symposium on Symposium on Theory of Computing - STOC '09* (ACM Press, Bethesda, MD, USA, 2009), p. 417.

[7] D. Aharonov, I. Arad, and T. Vidick, Guest column: The quantum PCP conjecture, SIGACT News **44**, 47 (2013).

[8] S. Aaronson and G. Kuperberg, Quantum versus classical proofs and advice, Theory Comput. **3**, 129 (2007).

[9] B. Fefferman and S. Kimmel, in *43rd International Symposium on Mathematical Foundations of Computer Science, MFCS 2018, August 27-31, 2018, Liverpool, UK*, LIPIcs, Vol. 117, edited by I. Potapov, P. G. Spirakis, and J. Worrell (Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018), p. 22:1.

[10] F. D. M. Haldane, Nonlinear Field Theory of Large-Spin Heisenberg Antiferromagnets: Semiclassically Quantized Solitons of the One-Dimensional Easy-Axis Néel State, Phys. Rev. Lett. **50**, 1153 (1983).

[11] E. Witten, Physical law and the quest for mathematical understanding, Bull. Am. Math. Soc. **40**, 21 (2002).

[12] T. S. Cubitt, D. Perez-Garcia, and M. M. Wolf, Undecidability of the spectral gap, Nature **528**, 207 (2015).

[13] M. B. Hastings, An area law for one-dimensional quantum systems, J. Stat. Mech. **2007**, P08024 (2007).

[14] Z. Landau, U. Vazirani, and T. Vidick, A polynomial time algorithm for the ground state of one-dimensional gapped local Hamiltonians, Nat. Phys. **11**, 566 (2015).

[15] I. Arad, Z. Landau, U. Vazirani, and T. Vidick, Rigorous RG algorithms and area laws for low energy eigenstates in 1D, Commun. Math. Phys. **356**, 65 (2017).

[16] D. Aharonov, M. Ben-Or, F. G. S. L. Brandao, and O. Sattath, The pursuit for uniqueness: Extending Valiant-Vazirani theorem to the probabilistic and quantum settings, Quantum **6**, 668 (2022).

[17] R. Jain, I. Kerenidis, G. Kuperberg, M. Santha, O. Sattath, and S. Zhang, On the power of a unique quantum witness, Theory Comput. **8**, 375 (2012).

[18] C. E. González-Guillén and T. S. Cubitt, History-state Hamiltonians are critical, arXiv:1810.06528 (2018).

[19] E. Crosson and J. Bowen, Quantum ground state isoperimetric inequalities for the energy spectrum of local Hamiltonians, arXiv:1703.10133 (2018).

[20] A. M. Dalzell and F. G. S. L. Brandao, Locally accurate MPS approximations for ground states of one-dimensional gapped local Hamiltonians, Quantum **3**, 187 (2019).

[21] D. W. Berry, A. M. Childs, R. Cleve, R. Kothari, and R. D. Somma, Exponential improvement in precision for simulating sparse Hamiltonians, Proc. 46th Annu. ACM Symp. Theory Comput. - STOC 14, 283 (2014).

[22] R. Babbush, D. W. Berry, I. D. Kivlichan, A. Y. Wei, P. J. Love, and A. Aspuru-Guzik, Exponentially more precise quantum simulation of fermions in second quantization, New J. Phys. **18**, 033032 (2016).

[23] A. M. Childs, R. Kothari, and R. D. Somma, Quantum algorithm for systems of linear equations with exponentially improved dependence on precision, SIAM J. Comput. **46**, 1920 (2017).

[24] M. Aguilera Sammaritano, M. González Vera, P. Marcelo Cometto, T. Nicola Tejero, G. F. Bauerfeldt, and A. Mellouki, Temperature dependence of rate coefficients for the gas phase reaction of OH with 3-chloropropene. A theoretical and experimental study, Chem. Phys. Lett. **755**, 137757 (2020).

[25] C. Hubig, J. Haegeman, and U. Schollwöck, Error estimates for extrapolations with matrix-product states, Phys. Rev. B **97**, 045125 (2018).

[26] S. Sachdev, *Quantum Phase Transitions* (Cambridge University Press, Cambridge, 2011), 2nd ed.

[27] Y. Atia and D. Aharonov, Fast-forwarding of Hamiltonians and exponentially precise measurements, Nat. Commun. **8**, 1572 (2017).

[28] B. Fefferman and C. Y.-Y. Lin, in *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*, Leibniz International Proceedings in Informatics (LIPIcs), Vol. 94, edited by A. R. Karlin (Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2018), p. 4:1.

[29] A. Kitaev and J. Watrous, in *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing - STOC '00* (ACM Press, Portland, Oregon, United States, 2000), p. 608.

[30] M. N. Vyalyi, in *ECCCTR: Electronic Colloquium on Computational Complexity, Technical Reports* (2003).

[31] C. Marriott and J. Watrous, Quantum Arthur–Merlin games, Comput. Complex. **14**, 122 (2005).

[32] The class PSPACE contains PP, and is believed to be unequal to and much larger than PP; see Fig. 1.

[33] E. W. Allender and K. W. Wagner, in *Current Trends in Theoretical Computer Science* (World Scientific, 1993), p. 469.

[34] S. Gharibian, M. Santha, J. Sikora, A. Sundaram, and J. Yirka, in *43rd International Symposium on Mathematical Foundations of Computer Science (MFCS 2018)*, Leibniz International Proceedings in Informatics (LIPIcs), Vol. 117 (Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2018), p. 58:1.

[35] T. Morimae and H. Nishimura, Merlinization of complexity classes above BQP, Quantum Inf. Comput. **17**, 959 (2017).

[36] R. Jefferson and R. C. Myers, Circuit complexity in quantum field theory, J. High Energ. Phys. **2017**, 107 (2017).

[37] C. D. White, C. Cao, and B. Swingle, Conformal field theories are magical, Phys. Rev. B **103**, 075145 (2022).

[38] D. Aharonov and A. Green, A quantum inspired proof of $\mathsf{P}^{\#\mathsf{P}} \subseteq \mathsf{IP}$, arXiv:1710.09078 (2017).

[39] A. Green, G. Kindler, and Y. Liu, Towards a quantum-inspired proof for IP = PSPACE, Quantum Inf. Comput. **21**, 377 (2021).

[40] A. Shamir, IP = PSPACE, J. ACM **39**, 869 (1992).

[41] S. Arora and B. Barak, *Computational Complexity: A Modern Approach* (Cambridge University Press, Cambridge; New York, 2009).

[42] M. Sipser, *Introduction to the Theory of Computation* (Cengage Learning, Boston, MA, 2012), 3rd ed.

[43] P. W. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, SIAM J. Comput. **26**, 1484 (1997).

[44] J. Kempe and O. Regev, 3-local Hamitonian is QMA-complete, Quantum Inf. Comput. **3**, 258 (2003).

[45] J. Kempe, A. Kitaev, and O. Regev, The complexity of the local Hamiltonian problem, SIAM J. Comput. **35**, 1070 (2006).

[46] R. V. Mises and H. Pollaczek-Geiringer, Praktische Verfahren der Gleichungsauflösung, Z. Für Angew. Math. Mech. **9**, 152 (1929).

[47] L. M. Adleman, J. DeMarrais, and M.-D. A. Huang, Quantum computability, SIAM J. Comput. **26**, 1524 (1997).

[48] N. Schuch, M. M. Wolf, F. Verstraete, and J. I. Cirac, Computational Complexity of Projected Entangled Pair States, Phys. Rev. Lett. **98**, 140506 (2007).

[49] S. Aaronson, Quantum computing, postselection, and probabilistic polynomial-time, Proc. R. Soc. Math. Phys. Eng. Sci. **461**, 3473 (2005).

[50] A. M. Childs, Y. Su, M. C. Tran, N. Wiebe, and S. Zhu, Theory of Trotter Error with Commutator Scaling, Phys. Rev. X **11**, 011020 (2021).

[51] J. Haah, M. B. Hastings, R. Kothari, and G. H. Low, in *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, Paris, 2018), p. 350.

[52] M. C. Tran, A. Y. Guo, Y. Su, J. R. Garrison, Z. Eldredge, M. Foss-Feig, A. M. Childs, and A. V. Gorshkov, Locality and Digital Quantum Simulation of Power-Law Interactions, Phys. Rev. X **9**, 031006 (2019).

[53] J. R. Schrieffer and P. A. Wolff, Relation between the Anderson and Kondo Hamiltonians, Phys. Rev. **149**, 491 (1966).

[54] S. Bravyi, D. DiVincenzo, and D. Loss, Schrieffer-Wolff transformation for quantum many-body systems, Ann. Phys. **326**, 2793 (2011).

[55] This reduction is inspired by unpublished work by one of us and Cedric Lin [91], and we supplement it with a technique to create spectral gaps.

[56] We actually show a reduction to the complement of the problem (where YES and NO instances are reversed), but this turns out not to matter because PSPACE is closed under complement.

[57] N. Schuch, I. Cirac, and F. Verstraete, Computational Difficulty of Finding Matrix Product Ground States, Phys. Rev. Lett. **100**, 250501 (2008).

[58] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, Nat. Commun. **5**, 4213 (2014).

[59] T. Jones, S. Endo, S. McArdle, X. Yuan, and S. C. Benjamin, Variational quantum algorithms for discovering Hamiltonian spectra, Phys. Rev. A **99**, 062304 (2019).

[60] D. Aharonov, I. Arad, and S. Irani, Efficient algorithm for approximating one-dimensional ground states, Phys. Rev. A **82**, 012315 (2010).

[61] A. Molnar, N. Schuch, F. Verstraete, and J. I. Cirac, Approximating Gibbs states of local Hamiltonians efficiently with projected entangled pair states, Phys. Rev. B **91**, 045138 (2015).

[62] S. Knysh, Zero-temperature quantum annealing bottlenecks in the spin-glass phase, Nat. Commun. **7**, 12370 (2016).

[63] N. Usher, M. J. Hoban, and D. E. Browne, Non-unitary quantum computation in the ground space of local Hamiltonians, Phys. Rev. A **96**, 032321 (2017).

[64] D. Nagaj, P. Wocjan, and Y. Zhang, Fast amplification of QMA, Quantum Inf. Comput. **9**, 1053 (2009).

[65] S. Aaronson, in *Proceedings of the Thiry-Fourth Annual ACM Symposium on Theory of Computing*, STOC '02 (Association for Computing Machinery, New York, NY, USA, 2002), p. 635.

[66] S. Aaronson, in *Proceedings of the Forty-Second ACM Symposium on Theory of Computing*, STOC '10 (Association for Computing Machinery, New York, NY, USA, 2010), p. 141.

[67] R. Raz and A. Tal, *Oracle Separation of BQP and PH*, Tech. Rep. 107 (2018).

[68] S. Bravyi, D. P. DiVincenzo, D. Loss, and B. M. Terhal, Simulation of Many-Body Hamiltonians using Perturbation Theory with Bounded-Strength Interactions, Phys. Rev. Lett. **101**, 070503 (2008).

[69] R. Oliveira and B. M. Terhal, The complexity of quantum spin systems on a two-dimensional square lattice, Quantum Inf. Comput. **8**, 0900 (2008).

[70] S. Bravyi, D. P. DiVincenzo, R. I. Oliveira, and B. M. Terhal, The complexity of stoquastic local Hamiltonian problems, Quantum Inf. Comput. **8**, 0361 (2008).

[71] D. Gottesman and S. Irani, The quantum and classical complexity of translationally invariant tiling and Hamiltonian problems, arXiv:0905.2419(2009).

[72] D. Aharonov, D. Gottesman, S. Irani, and J. Kempe, The power of quantum systems on a line, Commun. Math. Phys. **287**, 41 (2009).

[73] T. Cubitt and A. Montanaro, Complexity classification of local Hamiltonian problems, SIAM J. Comput. **45**, 268 (2013).

[74] D. Aharonov, W. van Dam, J. Kempe, Z. Landau, S. Lloyd, and O. Regev, Adiabatic quantum computation is equivalent to standard quantum computation, SIAM J. Comput. **37**, 166 (2007).

[75] N. Schuch and F. Verstraete, Computational complexity of interacting electrons and fundamental limitations of density functional theory, Nat. Phys. **5**, 732 (2007).

[76] T. S. Cubitt, A. Montanaro, and S. Piddock, Universal quantum Hamiltonians, Proc. Natl. Acad. Sci. USA **115**, 9497 (2018).

[77] J. Bausch, T. S. Cubitt, A. Lucia, D. Perez-Garcia, and M. M. Wolf, Size-driven quantum phase transitions, Proc. Natl. Acad. Sci. **115**, 19 (2018).

[78] S. Piddock and A. Montanaro, Universal qudit Hamiltonians, arXiv:1802.07130 (2018).

[79] T. Kohler and T. Cubitt, Toy models of holographic duality between local Hamiltonians, J. High Energ. Phys. **2019**, 17 (2019).

[80] T. Kohler, S. Piddock, J. Bausch, and T. Cubitt, Translationally-invariant universal quantum Hamiltonians in 1D, Annales Henri Poincaré **23**, 223 (2022).

[81] J. Bausch and E. Crosson, Analysis and limitations of modified circuit-to-Hamiltonian constructions, Quantum **2**, 94 (2018).

[82] 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**, 062306 (2018).

[83] J. D. Watson, Detailed analysis of circuit-to-Hamiltonian mappings, arXiv:1910.01481 (2019).

[84] J. Watrous, in *Encyclopedia of Complexity and Systems Science*, edited by R. A. Meyers (Springer New York, New York, NY, 2009), p. 7174.

[85] F. G. S. L. Brandao, Entanglement theory and the quantum simulation of many-body physics, (2008), arXiv:0810.0026.

[86] H. Blier and A. Tapp, in *Proceedings of the 2009 Third International Conference on Quantum, Nano and Micro Technologies*, ICQNM '09 (IEEE Computer Society, USA, 2009), p. 34.

[87] A. Pereszlényi, Multi-prover quantum Merlin-Arthur proof systems with small gap, arXiv:1205.2761 (2012).

[88] T. Ito, H. Kobayashi, and J. Watrous, in *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, ITCS '12 (Association for Computing Machinery, Cambridge, Massachusetts, 2012), p. 266.

[89] Z. Ji, in *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing - STOC 2017* (ACM Press, Montreal, Canada, 2017), p. 289.

[90] J. Fitzsimons, Z. Ji, T. Vidick, and H. Yuen, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing - STOC 2019* (ACM Press, Phoenix, AZ, USA, 2019), p. 473.

[91] B. Fefferman and C. Y.-Y. Lin, Quantum Merlin Arthur with exponentially small gap, arXiv:1601.01975 (2016).

[92] B. Fefferman, H. Kobayashi, C. Y.-Y. Lin, T. Morimae, and H. Nishimura, in *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*, Vol. 55 (Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016), p. 14:1.

[93] L. Valiant and V. Vazirani, NP is as easy as detecting unique solutions, Theor. Comput. Sci. **47**, 85 (1986).

[94] A. Ambainis, in *Proceedings of the 2014 IEEE 29th Conference on Computational Complexity*, CCC '14 (IEEE Computer Society, Washington, DC, USA, 2014), p. 32.

[95] S. Gharibian and J. Yirka, The complexity of simulating local measurements on quantum systems, Quantum **3**, 189 (2019).

[96] S. Gharibian, S. Piddock, and J. Yirka, in *Proceedings of 37th International Symposium on Theoretical Aspects of Computer Science (STACS 2020)* (Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2020), p. 20:1.

[97] L. Novo, J. Bermejo-Vega, and R. García-Patrón, Quantum advantage from energy measurements of many-body quantum systems, Quantum **5**, 465 (2021).

[98] P. Wocjan, D. Janzing, and T. Beth, Two QCMA-complete problems, Quantum Inf. Comput. **3**, 635 (2003).

[99] C. H. Bennett, Time/space trade-offs for reversible computation, SIAM J. Comput. **18**, 766 (1989).

[100] K.-J. Lange, P. McKenzie, and A. Tapp, Reversible space equals deterministic space, J. Comput. Syst. Sci. **60**, 354 (2000).

[101] In the sense that any witness orthogonal to the accepting witness rejects with probability at least $1 - s$.

[102] In the YES case, there could be some queries that are not valid UQCMA instances, and the oracle can answer arbitrarily for such ill-formed queries. This does not, however, hamper the proof, since a single valid query is enough to give a nonzero probability of saying YES.

[103] S. P. Jordan, H. Kobayashi, D. Nagaj, and H. Nishimura, Achieving perfect completeness in classical-witness quantum Merlin-Arthur proof systems, Quantum Inf. Comput. **12**, 461 (2012).

[104] The weaker promise is more natural since it is more robust.

[105] S. Dooley, G. Kells, H. Katsura, and T. C. Dorlas, Simulating quantum circuits by adiabatic computation: Improved spectral gap bounds, Phys. Rev. A **101**, 042302 (2020).