Oracle Complexity Classes and Local Measurements on Physical Hamiltonians

Sevag Gharibian
Paderborn University, Paderborn Germany
sevag.gharibian@upb.de

Stephen Piddock
University of Bristol, Bristol, UK
stephen.piddock@bristol.ac.uk

Justin Yirka
The University of Texas at Austin, Austin, TX, USA
yirka@utexas.edu

Abstract

The canonical hard problems for NP and its quantum analogue, Quantum Merlin-Arthur (QMA), are MAX-k-SAT and the k-local Hamiltonian problem (k-LH), the quantum generalization of MAX-k-SAT, respectively. In recent years, however, an arguably even more physically motivated problem than k-LH has been formalized – the problem of simulating local measurements on ground states of local Hamiltonians (APX-SIM). Perhaps surprisingly, [Ambainis, CCC 2014] showed that APX-SIM is likely harder than QMA. Indeed, [Ambainis, CCC 2014] showed that APX-SIM is $\mathrm{P}^{\mathrm{QMA[log]}}$-complete, for $\mathrm{P}^{\mathrm{QMA[log]}}$ the class of languages decidable by a P machine making a logarithmic number of adaptive queries to a QMA oracle. In this work, we show that APX-SIM is $\mathrm{P}^{\mathrm{QMA[log]}}$-complete even when restricted to physically motivated Hamiltonians, obtaining as intermediate steps a variety of related complexity-theoretic results.

Specifically, we first give a sequence of results which together yield $\mathrm{P}^{\mathrm{QMA[log]}}$-hardness for APX-SIM on well-motivated Hamiltonians such as the 2D Heisenberg model:

- We show that for NP, StoqMA, and QMA oracles, a logarithmic number of adaptive queries is equivalent to polynomially many parallel queries. Formally, $\mathrm{P}^{\mathrm{NP[log]}} = \mathrm{P}^{||\mathrm{NP}}$, $\mathrm{P}^{\mathrm{StoqMA[log]}} = \mathrm{P}^{||\mathrm{StoqMA}}$, and $\mathrm{P}^{\mathrm{QMA[log]}} = \mathrm{P}^{||\mathrm{QMA}}$. (The result for NP was previously shown using a different proof technique.) These equalities simplify the proofs of our subsequent results.
- Next, we show that the hardness of APX-SIM is preserved under Hamiltonian simulations (à la [Cubitt, Montanaro, Piddock, 2017]) by studying a seemingly weaker problem, $\forall$-APX-SIM. As a byproduct, we obtain a full complexity classification of APX-SIM, showing it is complete for $\mathrm{P}, \mathrm{P}^{||\mathrm{NP}}, \mathrm{P}^{||\mathrm{StoqMA}}$, or $\mathrm{P}^{||\mathrm{QMA}}$ depending on the Hamiltonians employed.
- Leveraging the above, we show that APX-SIM is $\mathrm{P}^{\mathrm{QMA[log]}}$-complete for any family of Hamiltonians which can efficiently simulate spatially sparse Hamiltonians. This implies APX-SIM is $\mathrm{P}^{\mathrm{QMA[log]}}$-complete even on physically motivated models such as the 2D Heisenberg model.

Our second focus considers 1D systems: We show that APX-SIM remains $\mathrm{P}^{\mathrm{QMA[log]}}$-complete even for local Hamiltonians on a 1D line of 8-dimensional qudits. This uses a number of ideas from above, along with replacing the "query Hamiltonian" of [Ambainis, CCC 2014] with a new "sifter" construction.

2012 ACM Subject Classification Theory of computation → Quantum computation theory; Theory of computation → Oracles and decision trees; Theory of computation → Problems, reductions and completeness; Theory of computation → Quantum complexity theory

Keywords and phrases Quantum Merlin Arthur (QMA), simulation of local measurement, local Hamiltonian, oracle complexity class, physical Hamiltonians

Digital Object Identifier 10.4230/LIPIcs.STACS.2020.20

Related Version A full version of the paper is available at `https://arxiv.org/abs/1909.05981`.

37th International Symposium on Theoretical Aspects of Computer Science (STACS 2020).
Editors: Christophe Paul and Markus Bläser; Article No. 20; pp. 20:1–20:37

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Funding *Sevag Gharibian*: SG acknowledges support from NSF grants CCF-1526189 and CCF-1617710.
Stephen Piddock: SP was supported by EPSRC.
Justin Yirka: Part of this work was completed while JY was supported by a Virginia Commonwealth University Presidential Scholarship. JY acknowledges QIP 2019 student travel funding (NSF CCF-1840547).

Acknowledgements We are grateful to Thomas Vidick for helpful discussions which helped initiate this work. We also thank an anonymous referee for [22] (written by two of the present authors) for the suggestion to think about 1D systems.

1 Introduction

In analogy with MAX-k-SAT playing a central role in the theory of NP-completeness, the *k-local Hamiltonian problem* (denoted k-LH, and which generalizes Boolean constraint satisfaction) is the canonical complete [31] problem for the quantum analogue of NP, Quantum Merlin Arthur (QMA). Roughly, in k-LH the input is a set of $2^k \times 2^k$ Hermitian matrices $\{ H_i \}$, where each H_i is a "local quantum constraint" acting on a subset of k out of n qubits. The output is the smallest eigenvalue of $H = \sum_i H_i$, known as the *ground state energy* of H, which we denote $\lambda(H)$. (For clarity, in the sum $H = \sum_i H_i$, each H_i is implicitly in tensor product with an identity on all qubits which H_i does not act on.) In words, the ground state energy encodes the energy of the quantum system corresponding to H when cooled into its lowest energy configuration. This remarkable connection between physics and complexity theory (i.e. Kitaev's proof that k-LH is QMA-complete [31]) spawned the field of Quantum Hamiltonian Complexity (QHC) (see, e.g., [37, 6, 18]), which has since explored the complexity of computing properties of ground spaces (i.e. "solution spaces" of k-LH instances) *beyond* estimating ground state energies [9, 44, 19, 21, 24, 20, 41, 3, 14, 22, 30, 7, 13].

Approximate Simulation

Despite the role of k-LH as a "posterchild" for Quantum Hamiltonian Complexity, in 2014 Ambainis formalized [3] the arguably even more natural physical problem of simulating local measurements on low-energy states of a local Hamiltonian, denoting it Approximate Simulation (APX-SIM). Intuitively, in APX-SIM one is given a local Hamiltonian H and local measurement A, and asked to estimate the expectation value of A against the ground space of H. Formally:

▶ **Definition 1** (APX-SIM$(H, A, k, \ell, a, b, \delta)$ [3]). *Given a k-local Hamiltonian H, an ℓ-local observable A, and real numbers a, b, and δ such that $b - a \geq n^{-c}$ and $\delta \geq n^{-c'}$, for n the number of qubits*[1] *H acts on and $c, c' > 0$ some constants, decide:*

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

For clarity, any Hermitian matrix A is a valid observable representing some projective measurement; the eigenvalues of A denote the labels of the possible outcomes of the measurement, and the eigenvectors the corresponding quantum state onto which the system is projected post-measurement.

[1] We state Definition 1 using *qubits*, i.e. 2-dimensional local systems. One could also use higher dimensional qu*d*its, i.e. *d*-dimensional local systems, if desired. Indeed, in certain contexts, such as the containment result of Lemma 11, showing a result about qu*d*its is more general than just considering qubits (hence Lemma 11 explicitly uses qudits).

Motivation for APX-SIM. Given a naturally occurring quantum system with time evolution Hamiltonian H (which is typically k-local for $k \in O(1)$), we would like to learn something about the quantum state $|\psi\rangle$ the system settles into when cooled to near absolute zero. This low-temperature setting is particularly important, as it is where phenomena such as superconductivity and superfluidity manifest themselves. Thus, learning something about $|\psi\rangle$ potentially allows one to harness such phenomena for, say, materials design. The most "basic" experimental approach to learning something about $|\psi\rangle$ is to attempt to prepare a physical copy of $|\psi\rangle$, and then apply a local measurement to extract information from $|\psi\rangle$. However, given that preparing the ground state $|\psi\rangle$ of an arbitrary Hamiltonian is hard – it would allow one to solve the QMA-complete k-LH problem – we must wonder whether there is an easier approach. Formally, how hard is APX-SIM?

Perhaps surprisingly, it turns out that simulating a measurement on the ground state $|\psi\rangle$ is strictly *harder than* QMA. To show this, [3] proved that APX-SIM is $\mathrm{P}^{\mathrm{QMA[log]}}$-complete, for $\mathrm{P}^{\mathrm{QMA[log]}}$ the class of languages decidable by a P machine making a logarithmic number of adaptive queries to a QMA oracle. (See Section 2 and Appendix A for formal details on promise oracle classes P^{C}.) Why $\mathrm{P}^{\mathrm{QMA[log]}}$ instead of QMA? Intuitively, this is because APX-SIM does *not* include thresholds for the ground state energy as part of the input (in contrast to k-LH). This specification of APX-SIM is moreover well-motivated; typically one does not have an estimate of the ground state energy of H, since such an estimate is QMA-hard to compute to begin with.

Brief background on $\mathrm{P}^{\mathrm{QMA[log]}}$. The class $\mathrm{P}^{\mathrm{QMA[log]}}$ is likely strictly harder than QMA, since both QMA and co-QMA are contained in $\mathrm{P}^{\mathrm{QMA[log]}}$ (to put co-QMA in $\mathrm{P}^{\mathrm{QMA[log]}}$, use the QMA oracle once and flip its answer using the P machine). Thus, QMA $\neq \mathrm{P}^{\mathrm{QMA[log]}}$ unless co-QMA $\subseteq$ QMA (which appears unlikely). Just how much more difficult than QMA is $\mathrm{P}^{\mathrm{QMA[log]}}$? Intuitively, the answer is "slightly more difficult". Formally, QMA $\subseteq \mathrm{P}^{\mathrm{QMA[log]}} \subseteq$ PP [22] (where QMA $\subseteq \mathrm{A_0PP} \subseteq$ PP was known [32, 45, 34] prior to [22]; note the latter containment is strict unless the Polynomial-Time Hierarchy collapses [45]).

From a computer science perspective, there is an interesting relationship between APX-SIM and classical constraint satisfaction. Recall that k-LH is the QMA-complete generalization of MAX-k-SAT, in that the energy of a state is minimized by simultaneously satisfying as many k-local constraints as possible. Classically, one might be asked whether the solution to a MAX-k-SAT instance satisfies some easily verifiable property, such as whether the solution has even Hamming weight; such a problem is $\mathrm{P}^{\mathrm{NP[log]}}$-complete (see [46] for a survey). APX-SIM is a quantum analogue to these problems, in which we ask whether an optimal solution (the ground state) satisfies some property (expectation bounds for a specified measurement), and APX-SIM is analogously $\mathrm{P}^{\mathrm{QMA[log]}}$-complete. Beyond this connection, of course, the strong appeal of APX-SIM lies additionally in its physical motivation.

High level direction in this work. That APX-SIM is such a natural problem arguably demands that we study its hardness given natural settings. In this regard, the original $\mathrm{P}^{\mathrm{QMA[log]}}$-completeness result [3] was for simulating $O(\log n)$-local observables and $O(\log n)$-local Hamiltonians. From a physical perspective, one wishes to reduce the necessary locality, e.g. to $O(1)$-local observables and Hamiltonians. Hardness under this restriction was subsequently achieved [22], for 1-qubit observables and 5-local Hamiltonians, by combining the "query Hamiltonian" construction of Ambainis [3] with the circuit-to-Hamiltonian construction of Kitaev [31]. However, even arbitrary $O(1)$-local Hamiltonians may be rather artificial in contrast to naturally occurring systems. Ideally, one wishes to make statements

along the lines of "simulating measurements on a physical model such as the quantum Heisenberg model on a 2D lattice is harder than QMA", or "simulating measurements on a 1D local Hamiltonian is harder than QMA". This is what we achieve in the current paper. Interestingly, to attain this goal, we first take a turn into the world of parallel versus adaptive oracle queries.

Parallel versus adaptive queries

A natural question for oracle complexity classes is how the power of the class changes as access to the oracle is varied. In the early 1990's, it was shown [10, 27, 5] that a polynomial number of *parallel* or *non-adaptive* queries to an NP oracle are equivalent in power to a logarithmic number of *adaptive* queries. Formally, letting $\mathrm{P}^{||\mathrm{NP}}$ be the class of languages decidable by a P machine with access to polynomially many parallel queries to an NP oracle, it holds that $\mathrm{P}^{||\mathrm{NP}} = \mathrm{P}^{\mathrm{NP[log]}}$ [5].

We begin by considering the analogous question for $\mathrm{P}^{\mathrm{QMA[log]}}$ versus $\mathrm{P}^{||\mathrm{QMA}}$ (defined as $\mathrm{P}^{||\mathrm{NP}}$ but with a QMA oracle). For this, the direction $\mathrm{P}^{\mathrm{C[log]}} \subseteq \mathrm{P}^{||\mathrm{C}}$ was shown by [5] for all classes C. For the reverse, to show $\mathrm{P}^{||\mathrm{NP}} \subseteq \mathrm{P}^{\mathrm{NP[log]}}$, [5] first conducts binary search to determine the number of YES queries. Unfortunately, it is not clear how to carry out an analogous binary search in the setting of *promise problems*, such as for QMA. The problem is that, as explored in [22], oracles for promise classes like QMA may receive queries which violate their promise (such as an instance of k-LH with the ground state energy within the promise gap). By definition [23], in such cases the oracle can respond arbitrarily, even changing its answer given repeated queries. As a result, the number of YES queries, by which we mean queries on which the QMA oracle outputs 1, is not even well-defined. Thus, the technique of binary search fails for QMA. (Note that for a language version of QMA – i.e. the quantum analogue of MA as opposed to PromiseMA – this technique would still work.) To hence show $\mathrm{P}^{||\mathrm{QMA}} \subseteq \mathrm{P}^{\mathrm{QMA[log]}}$, we take a different approach: we show a *hardness* result. Specifically, we use a modification of the $\mathrm{P}^{\mathrm{QMA[log]}}$-hardness construction of [3], for which we require the locality improvements of [22], to show that APX-SIM is $\mathrm{P}^{||\mathrm{QMA}}$-hard. Combining with the fact that APX-SIM $\in \mathrm{P}^{\mathrm{QMA[log]}}$ [3] then yields the desired containment. This approach has two added benefits:

- First, the use of parallel, rather than adaptive, queries simplifies the "query Hamiltonian" construction of [3] significantly, which we later exploit to prove hardness results about physical Hamiltonians (Theorem 6) and 1D Hamiltonians (Theorem 10). Moreover, we give a simpler proof of Ambainis's original claim that APX-SIM is $\mathrm{P}^{\mathrm{QMA[log]}}$-hard; indeed, we generalize it to classes C beyond QMA to show:

  ▶ **Theorem 2.** *Let* C *be a class of languages or promise problems. Let $\mathcal{F}$ be a family of Hamiltonians for which k-LH is* C*-complete under poly-time many-one reductions for all $k \geq 2$. Suppose $\mathcal{F}$ is closed under positive linear combination of Hamiltonians, and that if $\{H_i\}_{i=1}^m \subset \mathcal{F}$, then $H_{cl} + \sum_{i=1}^m |1\rangle\langle 1|_i \otimes H_i \in \mathcal{F}$, where H_{cl} is any classical Hamiltonian (i.e. diagonal in the standard basis)*[2]*. Then, $\mathrm{P}^{\mathrm{C[log]}} = \mathrm{P}^{||\mathrm{C}}$, and APX-SIM is $\mathrm{P}^{\mathrm{C[log]}}$-complete when restricted to k-local Hamiltonians and observables from $\mathcal{F}$.*

  Recalling that k-LH is NP-complete, StoqMA-complete, and QMA-complete when restricted to the families of classical, stoquastic, and arbitrary k-local Hamiltonians, respectively [12], Theorem 2 thus gives the sequence of results:

[2] Briefly, the reason for the form of the expression $H_{\mathrm{cl}} + \sum_{i=1}^m |1\rangle\langle 1|_i \otimes H_i$ is that it suffices to encode our construction, while still belonging to several interesting families $\mathcal{F}$.

▶ **Corollary 3.** $\mathrm{P}^{\mathrm{NP}[\log]} = \mathrm{P}^{||\mathrm{NP}}, \mathrm{P}^{\mathrm{StoqMA}[\log]} = \mathrm{P}^{||\mathrm{StoqMA}}$, *and* $\mathrm{P}^{\mathrm{QMA}[\log]} = \mathrm{P}^{||\mathrm{QMA}}$.

- Second, we use the Cook-Levin theorem [11, 33], as opposed to Kitaev's circuit-to-Hamiltonian construction [31] as in [22]. This allows us to obtain a *constant* promise gap for the observable[3] A's threshold values (i.e. $b - a \geq \Omega(1)$, as opposed to $b - a \geq 1/\mathrm{poly}$), even when $\|A\| = O(1)$. Further, because the core of this construction is already spatially sparse, it eases proving hardness results about physical Hamiltonians (Theorem 6).

The complexity of APX-SIM for physically motivated Hamiltonians

With the simplifications that moving to parallel queries affords us (i.e. working with $\mathrm{P}^{||\mathrm{QMA}}$ versus $\mathrm{P}^{\mathrm{QMA}[\log]}$), we next study $\mathrm{P}^{||\mathrm{QMA}}$-hardness for physically motivated Hamiltonians. This requires a shift of focus to *simulations*, in the sense of [13], i.e. analog Hamiltonian simulations.

Recall that Kitaev originally proved QMA-hardness of k-LH for 5-local Hamiltonians [31]; this was brought down to 2-local Hamiltonians via perturbation theory techniques [29, 28]. Since then, there has been a large body of work (e.g. [36, 8, 12, 7, 39, 40]) showing complexity theoretic hardness results for ever simpler systems, much of which uses perturbative gadgets[4] to construct Hamiltonians which have approximately the same ground state energy as a Hamiltonian of an apparently more complicated form. Here, we wish to enable a similarly large number of results for the problem APX-SIM by using the same perturbative gadget constructions and ideas of analogue simulation.

To this end, Ref. [13] defines a strong notion of simulation which approximately preserves essentially all low-temperature properties of a Hamiltonian (including the ground state energy). It then observes that the perturbative gadget constructions used in the k-LH literature are examples of this definition of simulation. Ref. [13] then shows there exist simple families of Hamiltonians (such as the 2-qubit Heisenberg interaction) which are *universal* Hamiltonians, in the sense that they can simulate all $O(1)$-local Hamiltonians efficiently. Here "efficiently" means that the important parameters of the simulator Hamiltonian, are polynomially related to the corresponding parameters of the original Hamiltonian (see Definition 14).

How do simulations affect the complexity of APX-SIM? Ideally, we would like to show that efficient simulations similarly lead to reductions between classes of Hamiltonians for the problem APX-SIM. However, this is apparently difficult, as the definition of APX-SIM is not robust to small perturbations in the eigenvalues of the system. We instead consider a closely related, seemingly easier problem which we call ∀-APX-SIM.

▶ **Definition 4** (∀-APX-SIM$(H, A, k, \ell, a, b, \delta)$)**.** *Given a k-local Hamiltonian H, an ℓ-local observable A, and real numbers a, b, and δ such that satisfy $b - a \geq n^{-c}$ and $\delta \geq n^{-c'}$, for n the number of qubits H acts on and $c, c' > 0$ some constants, decide:*

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

[3] The constant gap is only for the input thresholds a, b for the expectation value of the observable A. The required "low-energy gap" defined by δ may be inverse polynomial, i.e. $\delta \geq 1/\mathrm{poly}$, and we note that the spectral gap of the Hamiltonian H may be arbitrarily small in our constructions unless otherwise noted. Thus, it is unclear how to apply this result to resolve questions concerning Hamiltonians with improved promise gaps, e.g. the Quantum PCP Conjecture.

[4] Very roughly, perturbative gadgets allow one to "craft" a set of desired low-lying eigenvalues/eigenspaces for a local Hamiltonian by carefully penalizing certain subspaces with non-constant weights.

Above, we have a stronger promise in the YES case than in APX-SIM: namely, *all* low-energy states $|\psi\rangle$ are promised to satisfy $\langle\psi| A |\psi\rangle \leq a$, as opposed to just a single ground state. Thus, ∀-APX-SIM is easier than APX-SIM, in that ∀-APX-SIM reduces to APX-SIM. (The reduction is trivial: a valid instance of ∀-APX-SIM is already a valid instance of APX-SIM.) We conclude that ∀-APX-SIM is contained in $\mathrm{P}^{\mathrm{QMA[log]}}$. Furthermore, the proof of Theorem 2 is actually sufficient to show that ∀-APX-SIM is $\mathrm{P}^{||\mathrm{C}}$-complete (when restricted to the corresponding family of Hamiltonians for arbitrary class C).

Our second result, Lemma 15, proves that efficient simulations correspond to reductions between instances of ∀-APX-SIM. As a byproduct, we combine this result with Theorem 2 and the universality classifications from [13] (cf. Corollary 3) to obtain complexity classifications for the original APX-SIM problem restricted to several families of Hamiltonians:

▶ **Theorem 5.** *Let $\mathcal{S}$ be an arbitrary fixed subset of Hermitian matrices on at most 2 qubits. Then the APX-SIM problem, restricted to Hamiltonians H and measurements A given as a linear combination of terms from $\mathcal{S}$, is*

1. P*-complete, if every matrix in $\mathcal{S}$ is 1-local;*
2. $\mathrm{P}^{\mathrm{NP[log]}}$*-complete, if $\mathcal{S}$ does not satisfy the previous condition and there exists $U \in SU(2)$ such that U diagonalizes all 1-qubit matrices in $\mathcal{S}$ and $U^{\otimes 2}$ diagonalizes all 2-qubit matrices in $\mathcal{S}$;*
3. $\mathrm{P}^{\mathrm{StoqMA[log]}}$*-complete, if $\mathcal{S}$ does not satisfy the previous condition and there exists $U \in SU(2)$ such that, for each 2-qubit matrix $H_i \in \mathcal{S}$, $U^{\otimes 2} H_i (U^\dagger)^{\otimes 2} = \alpha_i Z^{\otimes 2} + A_i I + I B_i$, where $\alpha_i \in \mathbb{R}$ and A_i, B_i are arbitrary single-qubit Hermitian matrices;*
4. $\mathrm{P}^{\mathrm{QMA[log]}}$*-complete, otherwise.*

Hardness of simulating local measurements on lattices and spatially sparse systems. With the previous two main results in hand, we are in a position to show that ∀-APX-SIM is $\mathrm{P}^{\mathrm{QMA[log]}}$-hard even when the Hamiltonian is restricted to a spatially sparse interaction graph (in the sense of [36]). This is analogous to the equivalent result for k-LH shown in [36], which was crucial in showing that the Local Hamiltonian problem is QMA-complete for Hamiltonians on a 2D square lattice. Formally, by exploiting the previously discussed results about parallel queries (Theorem 2) and simulations (Lemma 15) and by developing a variant of the hardness construction from Theorem 2, we are able to show the following:

▶ **Theorem 6.** *Let $\mathcal{F}$ be a family of Hamiltonians which can efficiently simulate any spatially sparse Hamiltonian. Then, APX-SIM is $\mathrm{P}^{\mathrm{QMA[log]}}$-complete even when restricted to a single-qubit observable and a Hamiltonian from the family $\mathcal{F}$.*

Via Theorem 6, we now obtain many corollaries via the long line of research using perturbative gadgets to prove QMA-completeness of restricted physical Hamiltonians; for brevity, here we list a select few such corollaries. We note that the locality of the observable input to APX-SIM may increase after simulation, but only by a constant factor which can be easily calculated based on the simulation used. For example, using the perturbative gadgets constructed in [39], the following is an immediate corollary of Theorem 6:

▶ **Corollary 7.** *APX-SIM is $\mathrm{P}^{\mathrm{QMA[log]}}$-complete even when the observable A is 4-local and the Hamiltonian H is restricted to be of the form $H = \sum_{(j,k)\in E} a_{(j,k)} h_{(j,k)}$ where $h_{(j,k)} = \alpha X_j X_k + \beta Y_j Y_k + \gamma Z_j Z_k$, E is the set of edges of a 2D square lattice, $a_{(j,k)} \in \mathbb{R}$, and at least two of α, β, γ are non-zero. The case $\alpha = \beta = \gamma$ corresponds to $XX + YY + ZZ$, which is the physically motivated Heisenberg interaction.*

But, there is not always a blow-up in the locality of A, as is shown by this corollary which follows from Theorem 6 and [42]:

▶ **Corollary 8.** *APX-SIM is* $\mathrm{P}^{\mathrm{QMA[log]}}$*-complete even when the observable A is 1-local and the Hamiltonian H is restricted to be of the form: $H = \sum_{(j,k)\in E} h_{(j,k)} + \sum_j B_j$, where $h_{(j,k)} = X_jX_k + Y_jY_k + Z_jZ_k$, E is the set of edges of a 2D square lattice, and B_j is a single qubit operator (that may depend on j).*

Finally, we remark that recent work on the simulation power of families of qudit Hamiltonians [40] can be used to show the following corollary:

▶ **Corollary 9.** *Let $|\psi\rangle$ be an entangled two qudit state. Then, APX-SIM is* $\mathrm{P}^{\mathrm{QMA[log]}}$*-complete even when the Hamiltonian H is restricted to be of the form $H = \sum_{j,k} \alpha_{j,k}|\psi\rangle\langle\psi|_{j,k}$, where $\alpha_{j,k} \in \mathbb{R}$ and $|\psi\rangle\langle\psi|_{j,k}$ denotes the projector onto $|\psi\rangle$ on qudits j and k.*

Each of these corollaries follows as the corresponding references show that the described families of Hamiltonians can efficiently simulate all spatially sparse Hamiltonians.

The complexity of APX-SIM on the line

We finally move to our last result, which characterizes the complexity of APX-SIM on the line. Historically, it was known that the NP-complete problem MAX-2-SAT on a line is efficiently solvable via dynamic programming or divide-and-conquer (even for large, but constant, dimension). It hence came as a surprise when [1] showed that 2-LH on a line is still QMA-complete. This result was for local dimension 13 ([1] claimed a result for 12-dimensional qudits; [26] identified and fixed an error in [1] by adding an extra dimension). [35] improved this to hardness for 12-dimensional qudits by leveraging the parity of the position of qudits. Most recently, [26] showed QMA-completeness for qudits of dimension 8 by allowing some of the clock transitions to be ambiguous. The complexity of k-LH on a 1D line remains open for local dimension $2 \leq d \leq 7$.

Returning to the setting of APX-SIM, the classical analogue of APX-SIM on a 1D line of bits is also in P; given a 2-local Boolean formula $\phi : \{0,1\}^n \mapsto \{0,1\}$, compute an optimal solution x to ϕ (which recall can be done in 1D as referenced above), and evaluate the desired efficiently computable local function on x (i.e. a "measurement" on a subset of the bits). This raises the question: is APX-SIM on a line still $\mathrm{P}^{\mathrm{QMA[log]}}$-complete? Or does its complexity in the 1D setting drop to, say, QMA? Our final result shows the former.

▶ **Theorem 10.** *APX-SIM is* $\mathrm{P}^{\mathrm{QMA[log]}}$*-complete even when restricted to Hamiltonians on a 1D line of 8-dimensional qudits and single-qudit observables.*

Thus, even in severely restricted geometries like the 1D line, simulating a measurement on a single qudit of the ground space remains harder than QMA.

Proof techniques for Theorem 10. We employ a combination of new and known ideas. We wish to simulate the idea from [22] that instead of having the P machine make m queries to a QMA oracle, it receives the answers to the queries as a "proof" $y \in \{0,1\}^m$ which it accesses whenever it needs a particular query answer. In [22], Ambainis's query Hamiltonian [3] was then used to ensure y was correctly initialized. However, it is not clear how to use Ambainis' query Hamiltonian (or variants of it) while maintaining a 1D layout.

We hence take a different approach. Instead of receiving the query *answers*, the P machine now has access to m QMA verifiers $\{V_i\}_{i=1}^m$ corresponding to the m queries, and for each of them receives a *quantum* proof $|\psi_i\rangle$ in some proof register R_i. The P machine then treats

the (probabilistic) outputs of each V_i as the "correct" answer to the query i. If a query i is a NO instance of a QMA problem, this works well – no proof can cause V_i to accept with high probability. However, if query i is a YES instance, a cheating prover may nevertheless submit a "bad" proof to verifier V_i, since flipping the output bit of V_i may cause the P machine to flip its final output bit. To prevent this, and thus ensure the P machine receives all correct answers with high probability, we use a delicate application of 1-local energy penalties, which we call "sifters", to the outputs of the V_i; just enough to penalize bad proofs for YES cases, but not enough to cause genuine NO cases to incur large energy penalties. Here, we again utilize our result that $\mathrm{P}^{\mathrm{QMA[log]}} = \mathrm{P}^{||\mathrm{QMA}}$ (Corollary 3), and choose to begin with a $\mathrm{P}^{||\mathrm{QMA}}$ instance; this allows us to apply identical, independent sifters to the output of each verifier V_i, significantly easing the subsequent analysis and transition to 1D.

We next plug this construction, where the P circuit has many sub-circuits V_i, into the 1D circuit-to-Hamiltonian construction of [26]. Similar to [22], we apply a corollary of the Projection Lemma of [28, 22] to argue that any low energy state must be close to a history state $|\psi\rangle$. Combining with our sifter Hamiltonian terms, we show in Lemma 23 that for $|\psi\rangle$ to remain in the low-energy space, it must encode V_i outputting approximately the right query answer for any query i. To then conclude that all query responses are jointly correct with high probability, and thus that the low-energy space encodes the correct final output to the $\mathrm{P}^{||\mathrm{QMA}}$ computation, we apply a known quantum non-commutative union bound. In fact, our argument immediately shows hardness for both APX-SIM and $\forall$-APX-SIM.

Open questions

Our results bring previous $\mathrm{P}^{\mathrm{QMA[log]}}$-hardness results for a remarkably natural problem, Approximate Simulation (APX-SIM), closer to the types of problems studied in the physics literature, where typically observables are $O(1)$-local, allowed interactions physically motivated, and the geometry of the interaction graph is constrained. There are many questions which remain open, of which we list a few here: (1) The coupling strengths for local Hamiltonian terms in Corollary 7,8,9 are typically non-constant, as these corollaries follow from the use of existing perturbation theory gadgets; can these coupling constants be made $O(1)$? Note this question is also open for the complexity classification of k-LH itself [12, 39]. (2) What is the complexity of $\mathrm{P}^{\mathrm{QMA[log]}}$? It is known that $\mathrm{P}^{\mathrm{QMA[log]}} \subseteq \mathrm{PP}$ [22]; can a tighter characterization be obtained? (3) Can similar hardness results for APX-SIM be shown for *translationally invariant* 1D systems? For reference, it is known that k-LH is $\mathrm{QMA}_{\mathrm{exp}}$-complete for 1D translationally invariant systems when the local dimension is roughly 40 [25, 4]. ($\mathrm{QMA}_{\mathrm{exp}}$ is roughly the quantum analogue of NEXP, in which the proof and verification circuit are exponentially large in the input size. The use of this class is necessary in [25, 4], as the only input parameter for 1D translationally invariant systems is the *length* of the chain.) If a similar hardness result holds for APX-SIM, presumably it would show $\mathrm{P}^{\mathrm{QMA}_{\mathrm{exp}}\mathrm{[log]}}$-hardness for 1D translationally invariant systems.

Organization. We prove Theorems 2, 5, 6, and 10 in Sections 3, 4, 5, and 6, respectively. Proofs omitted due to space constraints are deferred to Appendices B, D, C, and E.

2 Preliminaries

Notation. $\lambda(H)$ denotes the smallest eigenvalue of Hermitian operator H. For matrix A, define spectral norm $\|A\|_\infty := \max\{\|A|v\rangle\|_2 : \||v\rangle\|_2 = 1\}$ and trace norm $\|A\|_{\mathrm{tr}} := \mathrm{Tr}\sqrt{A^\dagger A}$. Throughout, we assume both $H = \sum_{i=1}^m H_i$ and observable $A = \sum_{i=1}^m A_i$ satisfy $m, \|H_i\|_\infty, \|A_i\|_\infty \in O(\mathrm{poly}(n))$, for n the number of qubits in the system.

Definitions. $\mathrm{P}^{\mathrm{QMA[log]}}$ [3] is the set of problems decidable by a polynomial-time deterministic Turing machine with the ability to query an oracle for a QMA-complete problem $O(\log n)$ times, where n is the size of the input. For a class C of languages or promise problems, the class $\mathrm{P}^{\mathrm{C[log]}}$ is similarly defined, except with an oracle for a C-complete problem. (See Appendix A for further formal details and discussion on promise oracle classes.) $\mathrm{P}^{||\mathrm{C}}$ is the set of problems decidable by a polynomial-time deterministic Turing machine given access to an oracle for a C-complete problem, with the restriction that all (up to $O(n^c)$ for $c \in \Theta(1)$) queries to the oracle be made in parallel. Such queries are hence *non-adaptive*, as opposed to the *adaptive* queries allowed to a $\mathrm{P}^{\mathrm{C[log]}}$ machine.

For $\mathrm{P}^{\mathrm{QMA[log]}}$ we assume the P machine makes queries to an oracle for the QMA-complete [31] k-local Hamiltonian problem (k-LH), defined as follows: Given a k-local Hamiltonian H and inverse polynomial-separated thresholds $a, b \in \mathbb{R}$, decide whether $\lambda(H) \leq a$ (YES-instance) or $\lambda(H) \geq b$ (NO-instance) [28]. We say an oracle query is *valid* (*invalid*) if it satisfies (violates) the promise gap of the QMA-complete problem the oracle answers. (An invalid query hence satisfies $\lambda(H) \in (a, b)$.) For any invalid query, the oracle can accept or reject arbitrarily. A *correct* query string $y \in \{0,1\}^m$ encodes a sequence of correct answers to all of the m queries made by the P machine, and an *incorrect* query string is one which contains at least one incorrect query answer. Note that for an invalid query, any answer is considered "correct", yielding the possible existence of multiple correct query strings. Nevertheless, the P machine is required to output the same final answer (accept or reject) regardless of how such invalid queries are answered [23].

3 Parallel versus adaptive queries

This section shows Theorem 2, i.e. that $\mathrm{P}^{\mathrm{C[log]}} = \mathrm{P}^{||\mathrm{C}}$ for appropriate complexity classes C, which will follow from Lemmas 11 and 12 below.

▶ **Lemma 11.** *Let H be a k-local Hamiltonian acting on n qudits, and let A be an observable on the same system of n qudits. If k-LH for $\alpha H + \beta A$ is contained in complexity class* C *for any $0 \leq \alpha, \beta \leq \mathrm{poly}(n)$ and for all $k \geq 1$, then APX-SIM$(H, A, k, \ell, a, b, \delta) \in \mathrm{P}^{\mathrm{C[log]}}$ for all $\ell \leq O(\log n)$ and $b - a, \delta \geq O(1/\mathrm{poly}\ n)$.*

The proof of Lemma 11 generalizes the known [3] proof that APX-SIM $\in \mathrm{P}^{\mathrm{QMA[log]}}$; the basic idea is to use binary search in conjunction with the oracle for C to estimate the ground state energy λ of H. One additional oracle query is then made to to determine whether H has a ground state with energy approximately λ and satisfying the promise thresholds for observable A. This last query, in particular, is where we must be careful in our generalization to arbitrary classes C. The formal proof is in Appendix B.

▶ **Lemma 12.** *Let $\mathcal{F}$ be a family of Hamiltonians for which k-LH is* C*-hard for all $k \geq 2$. Then $\forall$-APX-SIM is $\mathrm{P}^{||\mathrm{C}}$-hard even when $b - a = \Omega(1)$, the observable A is a single Pauli Z measurement, and when restricted to Hamiltonians of the form $H = H_{cl} + \sum_i |1\rangle\langle 1|_i \otimes H_i$, where H_{cl} is a classical Hamiltonian, and the H_i are Hamiltonians from $\mathcal{F}$.*

Before discussing the proof of Lemma 12, let us remark how Lemmas 11 and 12 combine to yield Theorem 2, i.e. that $\mathrm{P}^{\mathrm{C[log]}} = \mathrm{P}^{||\mathrm{C}}$. The interesting containment here is $\mathrm{P}^{||\mathrm{C}} \subseteq \mathrm{P}^{\mathrm{C[log]}}$. To show this, Lemma 12 yields that $\forall$-APX-SIM is $\mathrm{P}^{||\mathrm{C}}$-hard. But $\forall$-APX-SIM trivially reduces to APX-SIM, which Lemma 11 says is in $\mathrm{P}^{\mathrm{C[log]}}$. Hence, we have used a hardness result to show containment. The formal argument is in Appendix B.

We develop two tools needed to show Lemma 12: How to simplify [3]'s query Hamiltonian in the context of parallel queries (which is used to enforce correct query answers), and how to employ the Cook-Levin reduction (which enforces a correct simulation of the circuit given those query answers).

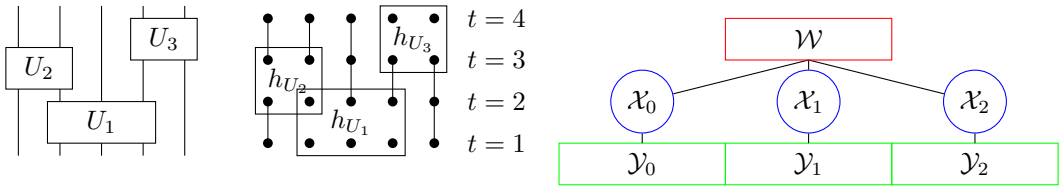

Figure 1 Left: Gates U_i in the circuit of the P machine. Middle: Hamiltonian terms h_{U_t} encoding each gate. Each straight line edge on the right represents the interaction $|01\rangle\langle 01| + |10\rangle\langle 10|$. The initialization terms H_{in} on qubits in time step $t = 0$ are omitted in the diagram. Right: The structure of the Hamiltonian $H = H_1 + H_2$ used in Lemma 12, for the case of 3 queries. H_1 acts on the space $\mathcal{W} \otimes \mathcal{X}$ and H_2 acts on $\mathcal{X} \otimes \mathcal{Y}$, where $\mathcal{X} = \bigotimes_i \mathcal{X}_i$ and $\mathcal{Y} = \bigotimes_i \mathcal{Y}_i$.

Tool 1: Simplifying Ambainis' query Hamiltonian. First, we give a simplified version of the "query Hamiltonian" of Ambainis [3], which will be useful in subsequent lemmas. Namely, given a $\mathrm{P}^{||\mathrm{C}}$ computation U for an appropriate class C, let $(H_{\mathcal{Y}_i}, a_i, b_i)$ be the instance of 2-LH corresponding to the i-th query made by U. Our "query Hamiltonian" is:

$$H = \sum_{i=1}^{m} M_i := \sum_{i=1}^{m} \left(\frac{a_i + b_i}{2} |0\rangle\langle 0|_{\mathcal{X}_i} \otimes I_{\mathcal{Y}_i} + |1\rangle\langle 1|_{\mathcal{X}_i} \otimes H_{\mathcal{Y}_i} \right), \tag{1}$$

where single qubit register $\mathcal{X}_i$ is intended to encode the answer to query i and $\mathcal{Y}_i$ encodes the ground state of $H_{\mathcal{Y}_i}$. Since each query is 2-local, H is 3-local. Notably, because U makes all of its queries in *parallel*, we are able to weight each of the m terms equally, unlike in [3, 22] which studied adaptive queries. This significantly eases our later analysis.

The key property of the query Hamiltonian H is given by the following lemma, which roughly says H encodes correct query answers in registers $\mathcal{X}_i$. This lemma is analogous to Lemma 3.1 of [22], but with an improved spectral gap; its proof (Appendix B) is similar to [22], but significantly simplified due to our use of parallel queries.

▶ **Lemma 13.** *Define for any $x \in \{0,1\}^m$ the space $\mathcal{H}_{x_1 \cdots x_m} := \bigotimes_{i=1}^m |x_i\rangle\langle x_i| \otimes \mathcal{Y}_i$. Then, there exists a correct query string $x \in \{0,1\}^m$ such that the ground state of H lies in $\mathcal{H}_{x_1 \cdots x_m}$. Moreover, if λ is the minimum eigenvalue of H restricted to this space, then for any incorrect query string $y_1 \cdots y_m$, any state in $\mathcal{H}_{y_1 \cdots y_m}$ has energy at least $\lambda + \epsilon$, where $\epsilon = \min_i(b_i - a_i)/2$.*

Tool 2: Adapting the Cook-Levin construction. We next model the Cook-Levin construction as a Hamiltonian for our setting. We may view the P machine as a circuit of classical reversible gates $U = U_m \ldots U_1$, in which time step i performs gate U_i. The evolution of the circuit is encoded into a 2D grid of qubits, where the t-th row of qubits corresponds to the state of the system at time step t; the output of the circuit is copied to a dedicated output bit in the final timestep. The overall Hamiltonian is diagonal in the computational basis with a groundspace of states corresponding to the correct evolution of the P machine.

Let I_t be the set of qubits which U_t acts non-trivially on. If a qubit $i \notin I_t$ (i.e. it is not acted on by the circuit at time step t), then there is an interaction $|01\rangle\langle 01| + |10\rangle\langle 10|$ on qubits (i,t) and $(i,t+1)$, to penalize states which encode a change on qubit i. To encode a classical reversible gate $U_t : x \mapsto U_t(x)$ acting at time t, we define an interaction $h_{U_t} = I - \sum_x |x\rangle\langle x|_t \otimes |U_t(x)\rangle\langle U_t(x)|_{t+1}$ acting non-trivially only on qubits (i,t') for $i \in I_t$ and t' equal to t or $t+1$. Figure 1 (middle) gives an illustration of this Hamiltonian. This

yields (positive semi-definite) propagation Hamiltonian

$$H_{\text{prop}} = \sum_{t=1}^{m} \left(h_{U_t} + \sum_{i \notin I_t} |0\rangle\langle 0|_{(i,t)} |1\rangle\langle 1|_{(i,t+1)} + |1\rangle\langle 1|_{(i,t)} |0\rangle\langle 0|_{(i,t+1)} \right), \tag{2}$$

whose ground space is spanned by states of the form: $|w(x)\rangle = |x\rangle_{t=1} \otimes |U_1 x\rangle_{t=2} \otimes \cdots \otimes |U_m \dots U_1 x\rangle_{t=m+1}$. We also add an H_{in} term consisting of 1-local $|1\rangle\langle 1|$ terms acting on the first ($t = 1$) row to ensure the start configuration is correct. One can show that the resulting Hamiltonian $H_{\text{prop}} + H_{\text{in}}$ has (1) unique ground state $|w(0^n)\rangle$ encoding the action of the circuit on the 0^n string, (2) ground state energy 0, and (3) spectral gap at least 1.

Proof sketch of Lemma 12. With our two tools in hand, we can finally sketch the proof of Lemma 12. Split the Hilbert space into three parts $\mathcal{W}$, $\mathcal{X} = \bigotimes_i \mathcal{X}_i$, $\mathcal{Y} = \bigotimes_i \mathcal{Y}_i$ and consider Hamiltonian $H = H_1 + H_2$, where H_1 acts on $\mathcal{W}$ and $\mathcal{X}$, and H_2 acts on $\mathcal{X}$ and $\mathcal{Y}$ (Figure 1). H_2 is the query Hamiltonian of Equation (1) (Tool 1); by Lemma 13, the low energy space of H_2 encodes in register in $\mathcal{X}$ a correct string of query answers for oracle C. $H_1 = H_{\text{prop}} + H_{\text{in}}$ is the classical Hamiltonian encoding the evolution of a classical P circuit, using the Cook-Levin construction (Tool 2), where H_{prop} is defined in Equation (2). Intuitively, the low energy space of H_1 simulates "reading" the correct query answers from $\mathcal{X}$ and using these to simulate the underlying P circuit in register $\mathcal{W}$. (Thus, $\mathcal{X}$ plays the role of a "message" register passing information between H_1 and H_2.) Details are in Appendix B.

4 Simulations and APX-SIM for physical classes of Hamiltonians

To study the complexity of APX-SIM for physically motivated Hamiltonians in Section 5, we require two tools: (1) hardness results for parallel query classes $\text{P}^{||\text{C}}$ (Section 3), and (2) an understanding of how *simulations* affect the hardness of the problem APX-SIM, which we now focus on. Roughly speaking, a simulation allows us to "reproduce" the low-energy physics of a desired physical model H, by instead using Hamiltonians H' of a different form. Formally, below we consider a simplified notion of simulation (a special case of the full definition given in [13]). This simpler case includes all of the important details necessary for the general case. Proofs with regard to the general definition of simulation are in Appendix F.

▶ **Definition 14** (Special case of definition in [13]; variant of definition in [7]). *We say that H' is a (Δ, η, ϵ)-simulation of H if there exists a local isometry $V = \bigotimes_i V_i$ such that*

1. *There exists an isometry $\widetilde{V}$ such that $\widetilde{V}\widetilde{V}^\dagger = P_{\leq \Delta(H')}$, where $P_{\leq \Delta(H')}$ is the projector onto the space of eigenvectors of H' with eigenvalues less than Δ, and $\|\widetilde{V} - V\| \leq \eta$;*
2. *$\|H'_{\leq \Delta} - \widetilde{V} H \widetilde{V}^\dagger\| \leq \epsilon$, where $H'_{\leq \Delta} = P_{\leq \Delta(H')} H' P_{\leq \Delta(H')}$.*

We say that a family $\mathcal{F}'$ of Hamiltonians can simulate a family $\mathcal{F}$ of Hamiltonians if, for any $H \in \mathcal{F}$ and any $\eta, \epsilon > 0$, and $\Delta \geq \Delta_0$ for some $\Delta_0 > 0$, there exists $H' \in \mathcal{F}'$ such that H' is a (Δ, η, ϵ)-simulation of H. We say that the simulation is efficient if, for H acting on n qudits, $\|H'\| = \text{poly}(n, 1/\eta, 1/\epsilon, \Delta)$; H' and $\{V_i\}$ are computable in polynomial-time given H, Δ, η and ϵ and provided that $\Delta, 1/\eta, 1/\epsilon$ are $O(\text{poly } n)$; and each isometry V_i maps from at most one qudit to $O(1)$ qudits.

Unlike in [13], here we have the additional requirement that the local isometry V is efficiently computable. This ensures that given some input Hamiltonian H and local observable A, we can use the notion of simulation to efficiently produce a simulating Hamiltonian H' *and* a simulating observable A'. As far as we are aware, all known constructions satisfying the notion of efficient simulation from [13] fulfill this additional requirement.

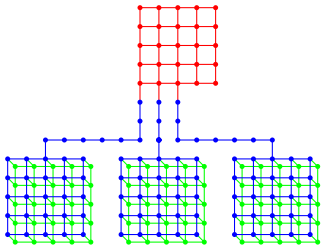

Figure 2 (Color figure) Geometric structure of Hamiltonian $H = H_1 + H_2 + H_3$ for the case of 3 queries. In words, H_1 is the top square, H_3 is the set of connecting wires and bottom three squares to which they are connected. H_2 is the remaining set of three squares at the bottom of the diagram.

One of the key properties of simulation is that it preserves eigenvalues of the target Hamiltonian up to a small additive factor ϵ. Unfortunately, even such small perturbations in eigenvalues can change the answer to APX-SIM if the Hamiltonian H has a small spectral gap. We hence instead work with the "more robust" problem $\forall$-APX-SIM, which recall trivially reduces to APX-SIM. Let $\mathcal{F}$-$\forall$-APXSIM denote the problem $\forall$-APX-SIM restricted to Hamiltonians taken from the family $\mathcal{F}$. The main result of this section is:

▶ **Lemma 15** (Simulations preserve hardness of $\forall$-APX-SIM)**.** *Let $\mathcal{F}$ be a family of Hamiltonians which can be efficiently simulated by another family $\mathcal{F}'$. Then, $\mathcal{F}$-$\forall$-APXSIM reduces to $\mathcal{F}'$-$\forall$-APXSIM via polynomial-time many-one reductions.*

The proof is rather technical, and hence deferred to Appendix C. As a corollary of Lemma 15, we also show Theorem 5, which recall gives a classification of the complexity of APX-SIM when restricted to families of Hamiltonians and measurements built up from a set of interactions $\mathcal{S}$: APX-SIM is either in P, $\mathrm{P}^{\mathrm{NP[log]}}$-complete, $\mathrm{P}^{\mathrm{StoqMA[log]}}$-complete, or $\mathrm{P}^{\mathrm{QMA[log]}}$-complete. Intuitively, this follows by combining the complexity characterizations and simulation results for k-LH of [12, 7, 13] with Lemma 15 and Theorem 2. However, some work is required to satisfy the preconditions of Theorem 2; details are in Appendix C.

5 Spatially sparse construction

We now combine the tools developed in the previous sections to study the complexity of APX-SIM for physical Hamiltonians. Our approach is to show that $\forall$-APX-SIM is $\mathrm{P}^{||\mathrm{QMA}}$-hard even for Hamiltonians on a *spatially sparse interaction graph*, defined below:

▶ **Definition 16** (Spatial sparsity [36])**.** *A spatially sparse interaction (hyper)graph G on n vertices is defined as a (hyper)graph in which*

1. *every vertex participates in $O(1)$ hyper-edges, and*
2. *there is a straight-line drawing in the plane such that every hyper-edge overlaps with $O(1)$ other hyper-edges and the surface covered by every hyper-edge is $O(1)$.*

▶ **Lemma 17.** *$\forall$-APX-SIM is $\mathrm{P}^{||\mathrm{QMA}}$-hard even for $b - a = \Omega(1)$, 1-local (single-qubit) observable A, and 4-local Hamiltonian H with a spatially sparse interaction graph.*

By combining Lemma 15, Lemma 17 and Corollary 3, we obtain Theorem 6, which recall shows APX-SIM is hard not only for families of Hamiltonians which are universal (i.e. families that can efficiently simulate any k-local Hamiltonian), but also for restricted families of Hamiltonians which can only efficiently simulate spatially sparse Hamiltonians. As stated in Section 1, this then yields the desired hardness results for APX-SIM on physical Hamiltonians such as the Heisenberg interaction on a 2D lattice (see, e.g., Corollary 7).

Proof sketch of Lemma 17. We adapt the proof of Lemma 12. Recall that the Hamiltonian H in Lemma 12 is $H = H_1 + H_2$, where H_2 uses a simplification of Ambainis's query Hamiltonian on each of the registers $\mathcal{X}_i \otimes \mathcal{Y}_i$ to encode the answer to that query into the state of $\mathcal{X}_i$, and H_1 encodes the evolution of the P circuit using the Cook-Levin construction on the $\mathcal{W}$ register (controlling on the states of the $\mathcal{X}_i$ registers). We arrange the qubits of the $\mathcal{W}$ register on a square lattice and note H_1 is already spatially sparse – this is an advantage of using the Cook-Levin construction over the Kitaev [31] history state construction. Furthermore, the Hamiltonian $H_{\mathcal{Y}_i}$, corresponding to the i-th QMA query, without loss of generality acts on a 2D square lattice [36], and hence is also spatially sparse.

The problem is that our version of Ambainis's query Hamiltonian H_2 is far from spatially sparse, since every qubit of $\mathcal{Y}_i$ interacts with $\mathcal{X}_i$. We resolve this by replacing each 1-qubit $\mathcal{X}_i$ register with a multi-qubit register of n_i qubits labeled $\{\mathcal{X}_i(j)\}_{j=1}^{n_i}$ (n_i the number of qubits of $\mathcal{Y}_i$). We spread out the qubits of the $\mathcal{X}_i$ register in space around the $\mathcal{Y}_i$ register, and modify H_2 so that each term is controlled only on a nearby qubit in $\mathcal{X}_i$. We also need to introduce a third term H_3 to ensure that all qubits in each $\mathcal{X}_i$ register are either all $|0\rangle$ or all $|1\rangle$. The construction is illustrated in Figure 2; the full proof is in Appendix D.

Finally, a brief comment about the claim that $b - a = \Omega(1)$ – as in Lemma 17, this is the promise gap for $\langle\psi| A |\psi\rangle$, and is $\Omega(1)$ due to our use of the Cook-Levin construction, which does not utilize history states (cf. [31]) to encode the action of the P machine.

6 Simulating measurements on a 1D line

We now show Theorem 10, i.e. that APX-SIM remains $\mathrm{P}^{\mathrm{QMA[log]}}$-complete even on a 1D line of 8-dimensional qudits with single-qudit observables. As the construction is rather involved, here we provide a sketch. Full details and a correctness proof are in Appendix E.

Sketch of 1D hardness construction. We give a reduction from $\mathrm{P}^{||\mathrm{QMA}}$ to $\forall$-APX-SIM, which by Theorem 2 yields the claim. Let Π be a $\mathrm{P}^{||\mathrm{QMA}}$ computation which takes in an input of size n and which consists of a uniformly generated polynomial-size classical circuit C making $m = O(\log n)$ 2-LH queries $\pi_i := (H_i, a_i, b_i)$ to a QMA oracle. As in Lemma 12, we treat the "answer register" in which C receives answers to its m queries as a proof register.

Our high-level approach consists of three steps: (1) construct a "master" circuit V composed of the verification circuits V_i corresponding to each query π_i and of the circuit C; (2) run V through the 1D circuit-to-Hamiltonian construction of [26] to obtain a 1D Hamiltonian G with local dimension 8, such that the low-energy space $\mathcal{S}$ of G consists of history states (of the form described in [26]); and (3) carefully add additional 1-local "sifter" penalty terms acting on the output qubits corresponding to each verification circuit V_i. Together, this yields a Hamiltonian H, whose low-energy space we show encodes satisfying proofs to each V_i (when possible). Specifically, the final step of "fine-grained splitting" of $\mathcal{S}$ (Step (3)) forces the output qubits of the circuits V_i to encode correct answers to query π_i, and thus the final circuit C receives a correct proof, hence leading the history states of step (2) to encode a correct simulation of Π. The answer to the computation Π can then be read off the ground state of H via an appropriate single qudit measurement. Note that Step (3) allows us to bypass the use of Ambainis' query Hamiltonian, which is a first for the study of $\mathrm{P}^{\mathrm{QMA[log]}}$.

1. Construction of V. Suppose each query π_i has corresponding QMA verification circuit V_i. We view the "master circuit" V as consisting of two phases:

1. (Verification phase) Given supposed proofs for each query, V runs all verification circuits V_i in parallel, where V_i acts on space $\mathcal{Y}_i \otimes \mathcal{W}_i \otimes \mathcal{X}_i$, for proof register $\mathcal{Y}_i$, ancilla register $\mathcal{W}_i$, and single-qubit output register $\mathcal{X}_i$.
2. (Simulated classical phase) The simulated P circuit C now receives the query answers $\mathcal{X} := \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_m$ as its proof register as well as an ancilla register $\mathcal{W}_0$. It outputs a single qubit to an output register $\mathcal{X}_0$.

Crucially, note that given a set of proofs in register $\mathcal{Y} = \bigotimes_{i=1} \mathcal{Y}_i$, V does *not* necessarily yield the same answer as Π, since a malicious prover could intentionally send a "bad" proof to a YES query, flipping the final answer of V.

2. Construction of G. We now plug V into the circuit-to-Hamiltonian construction of Hallgren, Nagaj, and Narayanaswami [26] and make a small modification (see Appendix E) to obtain a nearest-neighbor 1D Hamiltonian G on 8-dimensional qudits.

3. Adding 1-local "sifters". We now add 1-local Hamiltonian terms which serve to "sift" through bad proofs, or more accurately split the ground space of G, so as to force low-energy states to encode correct query answers. Namely, even a correct simulation of circuit V may not output the correct answer for Π if a malicious prover gives a "bad" proof to query register $\mathcal{Y}_i$, even though π_i is a YES-query. We hence wish to penalize proofs $|\psi_i\rangle$ which lead verifier V_i to reject, whenever there exists a proof $|\phi_i\rangle$ V_i would have accepted (i.e. π_i is a YES query). To do so, we (roughly) add a "sifter" penalty term $\epsilon\,|0\rangle\langle 0|_{\mathcal{X}_i}$ to each answer register $\mathcal{X}_i$, for ϵ a carefully chosen inverse polynomial. In particular, ϵ must simultaneously (1) penalize NO answers enough to ensure the ground space encodes YES answers for YES-queries, but (2) be small enough that genuine NO query proofs are not accidentally rejected (i.e. when π_i is a NO-query). We collectively give the sifter terms the label H_{out}.

Final construction. The final Hamiltonian is $H := G + H_{\text{out}}$, with 1-qudit observable A penalizing output 0 on the designated output qudit of G. Proving correctness requires a number of additional lemmas (Appendix E); as a sample, two tools used are a corollary of the Projection Lemma of [28, 22] (low energy states must be close to a history state $|\psi\rangle$), and a (known) quantum union bound (all query answers are jointly correct with high probability).

References

1 D. Aharonov, D. Gottesman, S. Irani, and J. Kempe. The power of quantum systems on a line. *Communications in Mathematical Physics*, 287(1):41–65, 2009. doi:10.1007/s00220-008-0710-3.

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

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

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

5 R. Beigel. Bounded queries to SAT and the Boolean hierarchy. *Theoretical computer science*, 84(2):199–223, 1991. doi:10.1016/0304-3975(91)90160-4.

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

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

8 Sergey Bravyi, David P. DiVincenzo, and Daniel Loss. Schrieffer–Wolff transformation for quantum many-body systems. *Annals of Physics*, 326(10):2793–2826, 2011. doi:10.1016/j.aop.2011.06.004.

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

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

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

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

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

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

15 G. De las Cuevas and T. Cubitt. Simple universal models capture all classical spin physics. *Science*, 351(6278):1180–1183, 2016. doi:10.1126/science.aab3326.

16 C. A. Fuchs and J. van de Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Transactions on Information Theory*, 45(4):1216–1227, 1999. doi:10.1109/18.761271.

17 J. Gao. Quantum union bounds for sequential projective measurements. *Phys. Rev. A*, 92(5):052331, 2015. doi:10.1103/PhysRevA.92.052331.

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

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

20 S. Gharibian, Z. Landau, S. W. Shin, and G. Wang. Tensor network non-zero testing. *Quantum Information & Computation*, 15(9–10):885–899, 2015. doi:10.26421/QIC15.9-10.

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

22 S. Gharibian and J. Yirka. The complexity of simulating local measurements on quantum systems. In M. M. Wilde, editor, *12th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2017)*, volume 73 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 2:1–2:17, Dagstuhl, Germany, 2018. Schloss Dagstuhl. doi:10.4230/LIPIcs.TQC.2017.2.

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

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

25 D. Gottesman and S. Irani. The quantum and classical complexity of translationally invariant tiling and Hamiltonian problems. *Theory of Computing*, 9(2):31–116, 2013. doi:10.4086/toc.2013.v009a002.

26 S. Hallgren, D. Nagaj, and S. Narayanaswami. The local Hamiltonian problem on a line with eight states is QMA-complete. *Quantum Information & Computation*, 13(9–10):721–750, 2013. doi:10.26421/QIC13.9-10.

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

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

29 J. Kempe and O. Regev. 3-local Hamiltonian is QMA-complete. *Quantum Information & Computation*, 3(3):258–264, 2003. doi:10.26421/QIC3.3.

30 I. H. Kim. Markovian matrix product density operators : Efficient computation of global entropy. Available at arXiv:1709.07828v2 [quant-ph], 2017.

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

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

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

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

35 D. Nagaj. *Local Hamiltonians in quantum computation*. PhD thesis, Massachusetts Institute of Technology, 2008. Available at arXiv:0808.2117v1 [quant-ph].

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

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

38 S. K. Oskouei, S. Mancini, and M. M. Wilde. Union bound for quantum information processing. *Proceedings of the Royal Society A*, 475(2221), 2019. doi:10.1098/rspa.2018.0612.

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

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

41 G. Scarpa, A. Molnar, Y. Ge, J. J. Garcia-Ripoll, N. Schuch, D. Perez-Garcia, and S. Iblisdir. Computational complexity of PEPS zero testing. Available at arXiv1802.08214 [quant-ph], 2018.

42 N. Schuch and F. Verstraete. Computational complexity of interacting electrons and fundamental limitations of Density Functional Theory. *Nature Physics*, 5:732–735, 2009. doi:10.1038/nphys1370.

43 P. Sen. Achieving the Han-Kobayashi inner bound for the quantum interference channel. In *Proceedings of the 2012 IEEE International Symposium on Information Theory*, pages 736–740, 2012. doi:10.1109/ISIT.2012.6284656.

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

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

46 K.W. Wagner. Bounded query computations. In *Proceedings of Structure in Complexity Theory Third Annual Conference*, pages 260–277, 1988. doi:10.1109/SCT.1988.5286.

A Additional notation and definitions

Notation. For a subspace S, $S^{\perp}$ denotes the orthogonal complement of S. We denote the restriction of an operator H to subspace S as $H|_S$. The null space of H is denoted $\mathrm{Null}(H)$.

Definitions. We define the notion of containment of a promise problem $\Pi = (\Pi_{\mathrm{yes}}, \Pi_{\mathrm{no}})$ in complexity class P^{C} for promise class C following Definition 1.3 of Goldreich [23].

▶ **Definition 18** (Cook reduction among promise problems [23]). *A promise problem $\Pi = (\Pi_{\mathrm{yes}}, \Pi_{\mathrm{no}})$ is Cook-reducible to promise problem $\Pi' = (\Pi'_{\mathrm{yes}}, \Pi'_{\mathrm{no}})$ if there exists a polynomial-time oracle Turing machine M such that:*

- *For every $x \in \Pi_{\mathrm{yes}}$, $M^{\Pi'}(x) = 1$,*
- *for every $x \in \Pi_{\mathrm{no}}$, $M^{\Pi'}(x) = 0$,*

and any query q to Π' is answered by 1 if $q \in \Pi'_{\mathrm{yes}}$, by 0 if $q \in \Pi'_{\mathrm{no}}$, and arbitrarily otherwise.

Remarks: (1) Quantum complexity classes, such as QMA, are typically promise classes, despite the lack of the prefix "Promise" in their name. (In contrast, in the classical complexity theory community, one typically distinguishes between, say, MA and PromiseMA.) Thus, as stated in Section 2, a QMA oracle can (in line with Definition 18) answer an invalid QMA query (i.e. violating the promise of the oracle) arbitrarily. Regardless of how such invalid queries are answered, the P machine must output the same final answer [23]. (2) If according to Definition 18, Π Cook-reduces to Π' for Π' in promise class C, then we say $\Pi \in \mathrm{P}^{\mathrm{C}}$. However, it is crucial to note that this does *not necessarily* imply that $\Pi \in C$. A notorious example [23] of this is the promise problem xSAT, for which Π_{yes} is the set of two-tuples (ϕ_1, ϕ_2) where ϕ_1 and ϕ_2 are satisfiable and unsatisfiable Boolean formulae, respectively, and Π_{no} is analogous but with ϕ_2 and ϕ_1 being satisfiable and unsatisfiable, respectively. Then, one can show (see Theorem 5.1 of [23]) that NP is Cook-reducible to xSAT, and that xSAT is in NP $\cap$ co-NP, and yet this does not necessarily imply that NP $\in$ NP $\cap$ co-NP.

B Proofs for Section 3

Proof of Lemma 11. We need to show the existence of a poly(n) time classical algorithm to decide APX-SIM while making at most $O(\log n)$ queries to an oracle for C. As with the proof in [3], the idea is to use $O(\log n)$ oracle queries to determine the ground space energy $\lambda(H)$ of H by binary search, and then use one final query to determine the answer. In [3] the final query is a QMA query; here we show how this final query can be performed differently so that only an oracle for C is required.

First calculate a lower bound μ for $\lambda(A)$, the lowest eigenvalue of A. If A acts only on $O(1)$ qudits, then $\lambda(A)$ can be calculated via brute force (up to, say, inverse exponential additive error) in $O(1)$ time. If A acts on many qudits, then $\lambda(A)$ can alternatively be approximated to within inverse polynomial additive error by binary search (as in [3]) by querying the C oracle $O(\log \|A\|) = O(\log n)$ times. Note that without loss of generality, we may assume $0 \leq b - \mu \leq q(n)$ for some efficiently computable polynomial q. The lower bound holds since if $b < \mu \leq \lambda(A)$, we conclude our APX-SIM instance is a NO instance, and we reject. For the upper bound, it holds that $\mu \leq \|A\|_{\infty}$, and we may assume $b \leq \|A\|_{\infty}$, as otherwise our APX-SIM instance is either a YES or invalid instance, and in both cases we can accept. By assumption, $\|A\|_{\infty} \leq q(n)$ for appropriate polynomial q which can be computed efficiently by applying the triangle inequality to the local terms of A; note $\|A\|_{\infty}$ may hence be replaced by q in the bounds above.

Perform binary search with the oracle for C (an example of how to perform binary search with an oracle for a promise problem is given in [3]) to find λ^* such that $\lambda(H) \in [\lambda^*, \lambda^* + \epsilon]$ where

$$\epsilon = \frac{\delta(b-a)}{2(b-\mu)} \geq 1/\mathrm{poly}(n)$$

since $0 \leq b - \mu \leq \mathrm{poly}(n)$. This requires $O(\log 1/\epsilon) = O(\log n)$ queries to the oracle for C. Next perform one final query to the C oracle to solve k-LH with Hamiltonian H' with thresholds a' and b', where

$$H' = (b-\mu)H + \delta A \quad \text{and} \quad \begin{array}{l} a' = (\lambda^* + \epsilon)(b-\mu) + \delta a \\ b' = \lambda^*(b-\mu) + \delta b \end{array}$$

and accept if and only if this final query accepts. Observe this is an allowed query for the C oracle because H' is of the form required in the statement of the lemma (recall $b - \mu \geq 0$), and also

$$b' - a' = \delta(b-a) - \epsilon(b-\mu) = \delta(b-a)/2 \geq 1/\mathrm{poly}(n).$$

Now, if APX-SIM$(H, A, k, l, a, b, \delta)$ is a YES instance, then there exists $|\psi\rangle$ such that $\langle\psi| H |\psi\rangle = \lambda(H)$ and $\langle\psi| A |\psi\rangle \leq a$. Then

$$\langle\psi| (b-\mu)H + \delta A |\psi\rangle \leq \lambda(H)(b-\mu) + \delta a \leq (\lambda^* + \epsilon)(b-\mu) + \delta a = a'$$

and the algorithm accepts as required.

Now suppose the input is a NO instance. We will show that $\langle\psi| H' |\psi\rangle \geq b'$ for any $|\psi\rangle$ and so the algorithm rejects as required. First, if $|\psi\rangle$ is low-energy with $\langle\psi| H |\psi\rangle \leq \lambda(H) + \delta$, then it also satisfies $\langle\psi| A |\psi\rangle \geq b$, and so

$$\langle\psi| (b-\mu)H + \delta A |\psi\rangle \geq \lambda(H)(b-\mu) + \delta b \geq \lambda^*(b-\mu) + \delta b = b'$$

where we have used $\langle\psi| H |\psi\rangle \geq \lambda(H) \geq \lambda^*$ and $b - \mu \geq 0$. Otherwise, if $|\psi\rangle$ is high energy with $\langle\psi| H |\psi\rangle \geq \lambda(H) + \delta$, then

$$\begin{aligned} \langle\psi| (b-\mu)H + \delta A |\psi\rangle &\geq (\lambda(H) + \delta)(b-\mu) + \delta\lambda(A) \\ &= \lambda(H)(b-\mu) + \delta b + \delta(\lambda(A) - \mu) \geq \lambda^*(b-\mu) + \delta b = b' \end{aligned}$$

where we have used $\langle\psi| A |\psi\rangle \geq \lambda(A)$ and $\lambda(A) - \mu \geq 0$. Thus, we reject. ◀

Proof of Lemma 13. We proceed by contradiction. Let $x \in \{0,1\}^m$ ($y \in \{0,1\}^m$) denote a correct (incorrect) query string which has lowest energy among all *correct* (*incorrect*) query strings against H. (Note that x and y are well-defined, though they may not be unique; in this latter case, any such x and y will suffice for our proof.) For any $z \in \{0,1\}^m$, define λ_z as the smallest eigenvalue in $\mathcal{H}_z$.

Since y is an incorrect query string, there exists at least one $i \in \{1, \ldots, m\}$ such that y_i is the wrong answer to a valid query $H_{\mathcal{Y}_i}$. If query i is a YES-instance, the smallest eigenvalue of M_i corresponds to setting $\mathcal{X}_i$ to (the correct query answer) $|1\rangle$, and is at most a_i. On the other hand, the space with $\mathcal{X}_i$ set to $|0\rangle$ has all eigenvalues equaling $(a_i + b_i)/2$. A similar argument shows that in the NO-case, the $|0\rangle$-space has eigenvalues equaling $(a_i + b_i)/2$, and the $|1\rangle$-space has eigenvalues at least b_i. We conclude that flipping query bit i to the correct query answer $\overline{y}_i$ allows us to "save" an energy penalty of $(b_i - a_i)/2$ against M_i, and since all other terms act invariantly on $\mathcal{X}_i \otimes \mathcal{Y}_i$, we save $(b_i - a_i)/2$ against H as well.

Let y' denote y with bit i flipped. If y' is also an incorrect query string, we have $\lambda_{y'} < \lambda_y$, a contradiction due to the minimality of y. Conversely, if y' is a correct query string, then we must have $\lambda_{y'} \geq \lambda_x + (b_i - a_i)/2 \geq \lambda + \epsilon$, as otherwise we contradict the minimality of x. ◀

The following technical lemma will be used in the proofs of Lemmas 12 and 15.

▶ **Lemma 19.** *Let H be a Hamiltonian and ρ a density matrix satisfying* $\mathrm{Tr}(H\rho) \leq \lambda(H)+\delta$. *Let P be the projector onto the space of eigenvectors of H with energy less than $\lambda(H)+\delta'$. Then,*

$$\frac{1}{2}\|\rho-\rho'\|_1 \leq \sqrt{\frac{\delta}{\delta'}}, \qquad \text{where } \rho' = P\rho P/\,\mathrm{Tr}(P\rho).$$

Proof of Lemma 19. First, bound the trace distance by the fidelity in the usual way (using one of the Fuchs-van de Graf inequalities [16]):

$$\frac{1}{2}\|\rho-\rho'\|_1 \leq \sqrt{1-F(\rho,\rho')^2} \tag{3}$$

where

$$F(\rho,\rho') = \mathrm{Tr}\left(\sqrt{\sqrt{\rho}\rho'\sqrt{\rho}}\right) = \mathrm{Tr}\left(\sqrt{\frac{\sqrt{\rho}P\rho P\sqrt{\rho}}{\mathrm{Tr}(P\rho)}}\right) = \frac{1}{\sqrt{\mathrm{Tr}(P\rho)}}\,\mathrm{Tr}(\sqrt{\rho}P\sqrt{\rho}) = \sqrt{\mathrm{Tr}(P\rho)},$$

where the third equality follows since $(\sqrt{\rho}P\sqrt{\rho})^2 = \sqrt{\rho}P\rho P\sqrt{\rho}$ and since the latter is positive semi-definite. Now, it remains to bound $\mathrm{Tr}(P\rho)$. We note that H has eigenvalues at least $\lambda(H)+\delta'$ on the space annihilated by P and eigenvalues at least $\lambda(H)$ everywhere else, and so $H \succeq (\lambda(H)+\delta')(I-P)+\lambda(H)P = (\lambda(H)+\delta')I-\delta' P$. Therefore, using the bound on $\mathrm{Tr}(H\rho)$, we have

$$\lambda(H)+\delta \geq \mathrm{Tr}(H\rho) \geq (\lambda(H)+\delta')\,\mathrm{Tr}(\rho) - \delta'\,\mathrm{Tr}(P\rho) \quad \Leftrightarrow \quad 1-\mathrm{Tr}(P\rho) \leq \frac{\delta}{\delta'}.$$

Substituting this back into Equation (3) proves the result. ◀

Proof of Lemma 12. We split the Hilbert space into three parts $\mathcal{W}$, $\mathcal{X} = \bigotimes_i \mathcal{X}_i$, $\mathcal{Y} = \bigotimes_i \mathcal{Y}_i$ and have a Hamiltonian of the form $H = H_1 + H_2$, where H_1 acts on $\mathcal{W}$ and $\mathcal{X}$, and H_2 acts on $\mathcal{X}$ and $\mathcal{Y}$. H_2 is the query Hamiltonian of Equation (1), and therefore by Lemma 13 the space of eigenvectors of H_2 with eigenvalues less than $\lambda(H_2)+\epsilon$ is spanned by states of the form: $|x\rangle_{\mathcal{X}} \otimes |\phi\rangle_{\mathcal{Y}}$, where x is a correct string of answers for the queries to the C oracle.

$H_1 = H_{\mathrm{prop}} + H_{\mathrm{in}}$ is the classical Hamiltonian encoding the evolution of a classical P circuit, using the Cook-Levin construction of Section 3, where H_{prop} is as defined in Equation (2). For clarity, H_{prop} and H_{in} act on $\mathcal{W}$ and $\mathcal{W}\otimes\mathcal{X}$, respectively. We think of $\mathcal{W}$ as "laid out in a 2D grid" as in Figure 1, and of $\mathcal{X}$ as playing the role of a "message" register passing information between H_1 and H_2. We modify the Hamiltonian H_{in} which initializes the qubits at the start of the classical circuit. For each qubit $\mathcal{X}_i$ in $\mathcal{X}$, we initialize a corresponding qubit of the first ($t=0$) row of $\mathcal{W}$ into the same state with a penalty term $|1\rangle\langle 1|_{\mathcal{X}_i}\otimes|0\rangle\langle 0|_{\mathcal{W}_i} + |0\rangle\langle 0|_{\mathcal{X}_i}\otimes|1\rangle\langle 1|_{\mathcal{W}_i}$. All other qubits in the first ($t=0$) row of $\mathcal{W}$ are initialized to $|0\rangle$ with a penalty $|1\rangle\langle 1|$. The full construction is depicted diagrammatically in Figure 1. Note that as stated in the claim, H is of the form $H = H_{\mathrm{cl}} + \sum_i^m |1\rangle\langle 1|_i \otimes H_i$, where H_{cl} contains H_1 and the local terms of H_2 which are tagged with $|0\rangle\langle 0|$ in registers $\mathcal{X}_i$.

We can argue about the low-energy eigenspace of H as follows. Since the ground spaces of H_1 and H_2 have non-trivial intersection, $\lambda(H) = \lambda(H_1)+\lambda(H_2) = \lambda(H_2)$. Moreover, since $[H_1,H_2]=0$ (they overlap only on the $\mathcal{X}$ register, on which they are both diagonal in the standard basis), and since we may assume without loss of generality that $\lambda(H_2)+\epsilon$ is inverse polynomially bounded below 1 (otherwise, we can scale H_1 by an appropriate fixed polynomial), we conclude the space of eigenstates of H with eigenvalue less than $\lambda(H)+\epsilon$,

henceforth denoted $\mathcal{H}_{\text{low}}$, is spanned by states of the form $|\Phi\rangle = |w\rangle_{\mathcal{W}} \otimes |x\rangle_{\mathcal{X}} \otimes |\phi\rangle_{\mathcal{Y}}$, where x is a string of correct answers to the oracle queries and w is the classical string encoding the correct computation of the P circuit acting on x. The qubit corresponding to the output bit of the P circuit will be in the state $|1\rangle$ (resp. $|0\rangle$) in a YES (resp. NO) instance of $\forall$-APX-SIM.

To complete the proof let the observable $A = Z_{\text{out}}$, a Pauli Z measurement on the qubit corresponding to the output bit of the P circuit, and let $\delta = \epsilon/16$ and $\delta' = \epsilon$. Consider any state $|\psi\rangle$ with $\langle\psi| H |\psi\rangle \leq \lambda(H) + \delta$. Then by Lemma 19, there exists a state $|\psi'\rangle \in \mathcal{H}_{\text{low}}$ such that $\langle\psi'| H |\psi'\rangle \leq \lambda(H) + \delta' = \lambda(H) + \epsilon$ which satisfies $\| |\psi\rangle\langle\psi| - |\psi'\rangle\langle\psi'| \|_1 \leq 1/2$. So,

$$\langle\psi'| Z_{\text{out}} |\psi'\rangle = \begin{cases} -1 & \text{in a YES instance} \\ 1 & \text{in a NO instance} \end{cases}$$

which implies by Hölder's inequality that $\langle\psi| A |\psi\rangle$ is $\leq -1/2$ in a YES instance and $\geq 1/2$ in a NO instance, as required. ◀

Proof of Theorem 2. The containment $\mathrm{P}^{\mathrm{C[log]}} \subseteq \mathrm{P}^{||\mathrm{C}}$ follows directly from the same argument that $\mathrm{P}^{\mathrm{NP[log]}} \subseteq \mathrm{P}^{||\mathrm{NP}}$ of [5], which we summarized in Section 1. By Lemma 11, APX-SIM is contained in $\mathrm{P}^{\mathrm{C[log]}}$ for Hamiltonians and observables from $\mathcal{F}$. And by Lemma 12 $\forall$-APX-SIM is $\mathrm{P}^{||\mathrm{C}}$-hard for Hamiltonians from $\mathcal{F}$, even when the observable is a single Pauli Z measurement, which is contained in $\mathcal{F}$ by the assumption that $\mathcal{F}$ contains any classical Hamiltonian H_{cl}. Since $\forall$-APX-SIM trivially reduces to APX-SIM, we thus have that APX-SIM is similarly $\mathrm{P}^{||\mathrm{C}}$-hard, and the result follows. ◀

C Proofs for Section 4

Proof of Lemma 15. Let $\Pi = (H, A, k, \ell, a, b, \delta)$ be an instance of $\mathcal{F}$-$\forall$-APXSIM. We will demonstrate that one can efficiently compute $H' \in \mathcal{F}'$ and A', k', ℓ', a', b', and δ' such that $\Pi' = (H', A', k', \ell', a', b', \delta')$ is a YES (respectively NO) instance of $\forall$-APX-SIM if Π is a YES (resp. NO) instance of $\forall$-APX-SIM; further, we will have that $\ell' \in O(\ell), a' = a + (b-a)/3, b' = b - (b-a)/3$ and $\delta - \delta' \geq 1/\text{poly}(n)$. To do so, we shall pick parameters Δ, η, ϵ so that $\Delta, 1/\eta, 1/\epsilon$ are $O(\text{poly } n)$, upon which the definition of efficient simulation (Definition 14) guarantees we can efficiently compute a Hamiltonian H' being a (Δ, η, ϵ)-simulation of H, which we claim will preserve YES and NO instances H.

Let us leave Δ, η, ϵ arbitrary for now, and assume we have a simulation of the form given in Definition 14. Then, there exists an isometry $\widetilde{V} : \mathcal{H} \to \mathcal{H}'$ ($\mathcal{H}$ and $\mathcal{H}'$ are the spaces H and H' act on, respectively) which maps onto the space of eigenvectors of H' with eigenvalues less than Δ, i.e. onto $S_{\leq\Delta} := \text{Span}\{|\psi\rangle : H' |\psi\rangle = \lambda |\psi\rangle, \lambda \leq \Delta\}$. In addition, $\widetilde{V}$ satisfies $\|\widetilde{V} - \bigotimes_i V_i\| \leqslant \eta$ and $\|H_{\leqslant\Delta} - \widetilde{V} H \widetilde{V}^\dagger\| \leqslant \epsilon$.

Let $|\psi'\rangle$ be a low-energy state of H' satisfying $\langle\psi'| H' |\psi'\rangle \leq \lambda(H') + \delta'$ for δ' to be set later. First, we show that $|\psi'\rangle$ is close to a state $\widetilde{V} |\psi\rangle$ where $|\psi\rangle$ is a low-energy state of H; then, we will show that there exists an observable A', depending only on A and the isometries V_i, such that $\langle\psi'| A' |\psi'\rangle$ approximates $\langle\psi| A |\psi\rangle$ for any choice of $|\psi\rangle$. Since by Definition 14 A is efficiently computable, our choice of A' will be as well.

Let $|\phi\rangle = P_{\leq\Delta(H')} |\psi'\rangle / \|P_{\leq\Delta(H')} |\psi'\rangle\|$ be the (normalized) component of $|\psi'\rangle$ in $S_{\leq\Delta}$. By Lemma 19, we have

$$\frac{1}{2} \| |\psi'\rangle\langle\psi'| - |\phi\rangle\langle\phi| \|_1 \leq \sqrt{\frac{\delta'}{\Delta - \lambda(H')}}.$$

Since $S_{\leq\Delta} = \mathrm{Im}(\widetilde{V})$, there must exist a state $|\psi\rangle$ in $\mathcal{H}$ such that $\widetilde{V}|\psi\rangle = |\phi\rangle$; next, we will show that $|\psi\rangle$ has low-energy with respect to H. Note that $|\psi'\rangle = \sqrt{p}|\phi\rangle + \sqrt{1-p}|\phi^\perp\rangle$ for some $p \in [0,1]$ and a state $|\phi^\perp\rangle$ in $S^\perp_{\leq\Delta}$ which has higher energy: $\langle\phi^\perp|H'|\phi^\perp\rangle \geq \Delta \geq \langle\phi|H'|\phi\rangle$. Therefore,

$$\langle\psi'|H'|\psi'\rangle = p\langle\phi|H'|\phi\rangle + (1-p)\langle\phi^\perp|H'|\phi^\perp\rangle \geqslant \langle\phi|H'|\phi\rangle,$$

which implies that

$$\begin{aligned}
\langle\psi|H|\psi\rangle - \langle\psi'|H'|\psi'\rangle &\leqslant \langle\psi|H|\psi\rangle - \langle\phi|H'|\phi\rangle && (4)\\
&= \langle\phi|\widetilde{V}H\widetilde{V}^\dagger|\phi\rangle - \langle\phi|H'|\phi\rangle && (5)\\
&\leqslant \|H'_{\leq\Delta} - \widetilde{V}H\widetilde{V}^\dagger\| \leqslant \epsilon. && (6)
\end{aligned}$$

So, $\langle\psi|H|\psi\rangle \leqslant \lambda(H') + \delta' + \epsilon \leqslant \lambda(H) + \delta' + 2\epsilon$, where the final inequality follows from Lemma 27 of [13], which roughly states that eigenvalues are preserved up to error ϵ in a simulation (in particular, the minimum eigenvalues satisfy $|\lambda(H') - \lambda(H)| \leq \epsilon$).

For any local measurement A_S acting on subset of S qubits $\mathcal{H}_S$ (here $\mathcal{H}_S$ is the Hilbert space for qudits in set $S \subseteq [n]$), we can define the local measurement $A'_S = V_S A_S V_S^\dagger$ on $\mathcal{H}'_S$ where $V = \bigotimes V_i$ is the local isometry in the definition of simulation and $V_S := \bigotimes_{i\in S} V_i$. Note that A'_S acts only on the $O(|S|)$ qudits which V_S maps to. Furthermore, $V^\dagger(A'_S \otimes I)V = A_S \otimes I$ and so

$$\begin{aligned}
|\langle\psi'|A'_S\otimes I|\psi'\rangle - \langle\psi|A_S\otimes I|\psi\rangle| &= |\langle\psi'|A'_S\otimes I|\psi'\rangle - \langle\psi|V^\dagger(A'_S\otimes I)V|\psi\rangle| && (7)\\
&\leqslant \|A'_S\|\||\psi'\rangle\langle\psi'| - V|\psi\rangle\langle\psi|V^\dagger\|_1 && (8)\\
&\leqslant \|A_S\|\left(\||\psi'\rangle\langle\psi'| - |\phi\rangle\langle\phi|\|_1 + \|\widetilde{V}|\psi\rangle\langle\psi|\widetilde{V}^\dagger - V|\psi\rangle\langle\psi|V^\dagger\|_1\right) && (9)\\
&\leqslant \|A_S\|\left(\||\psi'\rangle\langle\psi'| - |\phi\rangle\langle\phi|\|_1 + 2\|\widetilde{V} - V\|\right) && (10)\\
&\leqslant \|A_S\|\left(2\sqrt{\frac{\delta'}{\Delta - \lambda(H')}} + 2\eta\right) && (11)
\end{aligned}$$

where to get to (10), we have used the triangle inequality to bound:

$$\begin{aligned}
&\|\widetilde{V}|\psi\rangle\langle\psi|\widetilde{V}^\dagger - V|\psi\rangle\langle\psi|V^\dagger\|_1 && (12)\\
&\leq \|\widetilde{V}|\psi\rangle\langle\psi|\widetilde{V}^\dagger - V|\psi\rangle\langle\psi|\widetilde{V}^\dagger\|_1 + \|V|\psi\rangle\langle\psi|\widetilde{V}^\dagger - V|\psi\rangle\langle\psi|V^\dagger\|_1 && (13)\\
&= \|\widetilde{V} - V\|\left(\||\psi\rangle\langle\psi|\widetilde{V}^\dagger\|_1 + \|V|\psi\rangle\langle\psi|\|_1\right) && (14)\\
&= 2\|\widetilde{V} - V\| && (15)
\end{aligned}$$

Therefore, to ensure that Π' is a YES (resp. NO) instance if Π is a YES (resp. NO) instance, we will choose $a' = a + (b-a)/3$ and $b' = b - (b-a)/3$. Choosing $\delta', \Delta, \epsilon, \eta$ such that

$$0 < \delta' + 2\epsilon < \delta \qquad \text{and} \qquad 0 < \|A\|\left(2\sqrt{\frac{\delta'}{\Delta - \lambda(H')}} + 2\eta\right) < \frac{b-a}{3}$$

completes the proof. ◀

Proof of Theorem 5. We first discuss containment in the claimed complexity classes, and then hardness.

Containment. In the first case it is trivial to simulate the outcome of 1-local measurements on the ground state of a 1-local Hamiltonian, as the ground state is an easily calculated product state. For the other three cases, it was shown in [12] and [7], that k-LH for these three families of Hamiltonians is complete for the classes NP, StoqMA, QMA, respectively. Therefore, by Lemma 11, APX-SIM is contained in $\mathrm{P}^{\mathrm{NP[log]}}, \mathrm{P}^{\mathrm{StoqMA[log]}}$ and $\mathrm{P}^{\mathrm{QMA[log]}}$, respectively. (Note that the precondition of Lemma 11 is met, i.e. for H and A given as a linear combination of terms from $\mathcal{S}$ and I, we have that k-LH for $\alpha H + \beta A$ is contained in the respective complexity class of NP, StoqMA, or QMA, for any $0 \leq \alpha, \beta \leq \mathrm{poly}(n)$, and for all $k \geq 1$.)

Hardness. Starting with the referenced completeness results of [12, 7] above, we now wish to show APX-SIM is hard for $\mathrm{P}^{\mathrm{NP[log]}}, \mathrm{P}^{\mathrm{StoqMA[log]}}$ and $\mathrm{P}^{\mathrm{QMA[log]}}$ for cases 2–4 of our claim. At first glance, it may seem that Theorem 2 already yields this result, since that theorem says that APX-SIM is $\mathrm{P}^{\mathrm{C[log]}}$-complete when restricted to k-local Hamiltonians and observables from a family $\mathcal{F}$. Unfortunately, however, a precondition of Theorem 2 is that $\mathcal{F}$ must contain all classical (i.e. diagonal in standard basis) Hamiltonians, which is *not* necessarily true for cases 2–4 of our claim here. Thus, some work is required get the hardness claims of cases 2–4 here.

To achieve this, we first apply Lemma 12 to conclude that $\forall$-APX-SIM is hard for classes $\mathrm{P}^{||\mathrm{NP}}, \mathrm{P}^{||\mathrm{StoqMA}}$ and $\mathrm{P}^{||\mathrm{QMA}}$ for the families of classical, stoquastic and arbitrary local Hamiltonians, respectively. (In contrast to the Hamiltonians of cases 2–4 of our claim here, the sets of classical, stoquastic and arbitrary local Hamiltonians *do* contain all diagonal Hamiltonians, and thus satisfy the preconditions of Lemma 12.) We then use simulations, in combination with Lemma 15, to reduce the sets of classical, stoquastic, and arbitrary local Hamiltonians to the Hamiltonians in cases 2,3,4 of our claim here, respectively.

Specifically, it was shown in [13] that the three families of Hamiltonians in cases 2–4 of our claim can efficiently simulate all classical, stoquastic and arbitrary local Hamiltonians, respectively, via some local isometry V (see Definition 14). It follows by Lemma 15 (which states that simulations act like hardness reductions) that $\forall$-APX-SIM is hard for $\mathrm{P}^{||\mathrm{NP}}, \mathrm{P}^{||\mathrm{StoqMA}}$ and $\mathrm{P}^{||\mathrm{QMA}}$ respectively, with respect to (using the notation of Lemma 15) a local observable A' (in the larger, simulating, space) such that $A' = VAV^\dagger$ (where in our case A will equal Pauli Z due to the proof of Lemma 12). The only obstacle to achieving our current claim is that we also require A' to be chosen as a linear combination of terms from $\mathcal{S}$ and I. This is what the remainder of the proof shall show.

Observation (*). To begin, note the proof of Lemma 12 used single qubit observable Z, since we encoded the P machine's output in a single bit, which we assumed was set to $|0\rangle$ for "reject" and $|1\rangle$ for "accept". However, without loss of generality, we may alter the starting P machine to encode its output in some more general function on two bits, such as the parity function. (For example, the P machine can be assumed to output a 2-bit string q, such that q has odd parity if and only if the P machine wishes to accept.) We use this observation as follows. Consider any classical observable A with two distinct eigenvalues $\lambda_x < \lambda_y$ corresponding to eigenstates $|x\rangle$ and $|y\rangle$, respectively, for distinct strings $x, y \in \{0,1\}^2$. Then, assuming the specification of A is independent of the number of qubits in the system (thus, A is specified to within constant bits of precision, and so $\lambda_y - \lambda_x \in \Theta(1)$), if we set the P machine to output x when it wishes to accept and y when it wishes to reject, a measurement with observable A suffices to distinguish these two cases. With this observation in hand, we consider cases 2–4 of our claim, in particular with respect to the action of isometry V.

Case 2: $\mathrm{P}^{||\mathrm{NP}}$*-completeness.* First note that in this case we can assume without loss of generality that all interactions in $\mathcal{S}$ are diagonal (by performing a global basis change of $U^{\otimes n}$ if necessary) . Since we are not in the first case we know also that there is a 2-local interaction in $\mathcal{S}$ with at least two distinct eigenvalues. By Observation (*), it will suffice to simulate such an observable on a particular pair of qubits in the original system; call this operator A. For the $\mathrm{P}^{\mathrm{NP[log]}}$ case, the isometry V appends some ancilla qubits in a computational basis state (in the $U^{\otimes n}$ basis) [15]. We can therefore choose A' to be the same 2-local observable A, but acting on the corresponding qubits in the larger, simulating system; that is, if we let $A' = A \otimes I$ (where the identity term acts on the ancilla qubits), then $V^\dagger A' V = A$ as desired.

Case 3: $\mathrm{P}^{||\mathrm{StoqMA}}$*-completeness.* For the third case, one can check that the reductions in [7] correspond to a simulation with an isometry V which maps each qubit $|0\rangle \mapsto |0011\rangle$ and $|1\rangle \mapsto |1100\rangle$ and appends some additional ancilla qubits in a computational basis state (see discussion in Section 9.4 of [13]). Thus, a classical 2-local observable $Z \otimes Z + \mathrm{diag}(A) \otimes I + I \otimes \mathrm{diag}(B)$ (which we may use by Observation (*)) can be simulated in the larger, simulating space on physical qubits $1, 2, 3, 4$ (logical qubit 1) and $5, 6, 7, 8$ (logical qubit 2) via:

$$V^\dagger (Z_1 Z_5 + A_1 + B_5) V = Z \otimes Z + \mathrm{diag}(A) \otimes I + I \otimes \mathrm{diag}(B),$$

where $\mathrm{diag}(A)$ denotes the diagonal part of A, i.e. $\mathrm{diag}(A) = \sum_{i=0}^{1} |i\rangle\langle i|A|i\rangle\langle i|$. Thus, measuring observable $(Z_1 Z_5 + A_1 + B_5)$ on the larger, simulating Hamiltonian H' (which has the desired form of Case 3 here) is equivalent to measuring $Z \otimes Z + \mathrm{diag}(A) \otimes I + I \otimes \mathrm{diag}(B)$ on the starting Hamiltonian H in the simulation (again, using notation of Lemma 15).

Case 4: $\mathrm{P}^{||\mathrm{QMA}}$*-completeness.* The final case is slightly more complicated. When showing that these Hamiltonians are universal, the one step with a non-trivial isometry is simulating $\{X, Z, XX, ZZ\}$-Hamiltonians with $\{XX + YY\}$-Hamiltonians or $\{XX + YY + ZZ\}$-Hamiltonians in Theorem 41 of [13]. In both of these cases, the isometry V maps each qubit via action

$$|0\rangle \mapsto |\Psi^-\rangle_{13} |\Psi^-\rangle_{24} \qquad |1\rangle \mapsto \tfrac{2}{\sqrt{3}} |\Psi^-\rangle_{12} |\Psi^-\rangle_{34} - \tfrac{1}{\sqrt{3}} |\Psi^-\rangle_{13} |\Psi^-\rangle_{24}.$$

In the proof of Theorem 41 of [13], it is shown that a single Z observable can be reproduced by choosing $A = h_{13}$ (where either $h = XX + YY$ or $h = XX + YY + ZZ$), that is $V^\dagger h_{13} \otimes I_{24} V$ is proportional to Z.

The proof is completed by Corollary 3 (i.e. logarithmic adaptive queries are equivalent to polynomially many parallel queries). ◀

D Proofs for Section 5

Proof of Lemma 17. We will construct a Hamiltonian on the registers $\mathcal{W}$, $\mathcal{X}_i$ and $\mathcal{Y}_i$ for $i \in \{1, \ldots m\}$, for which the problem $\forall$-APX-SIM encodes the output of a $\mathrm{P}^{||\mathrm{QMA}}$ circuit, where m is the number of parallel queries to the QMA oracle.

Let the qubits of $\mathcal{W}$ and $\mathcal{Y}_i$ be arranged on distinct parts of a square lattice. For each qubit of $\mathcal{Y}_i$, there is a corresponding qubit in $\mathcal{X}_i$, and $\mathcal{X}_i$ contains a path of qubits leading from $\mathcal{Y}_i$ to $\mathcal{W}$. See Figure 2 for an example layout in the case $m = 3$.

Let E_i be the set of edges of the square lattice of qubits of $\mathcal{Y}_i$ (i.e. not including the edges connecting $\mathcal{Y}_i$ to $\mathcal{X}_i$ in Figure 2) and let $H_{\mathcal{Y}_i} = \sum_{(j,k) \in E_i} h^i_{\mathcal{Y}_i(j,k)}$ be a 2D nearest neighbor Hamiltonian on $\mathcal{Y}_i$ corresponding to the i-th query. We have used the subscript notation $\mathcal{Y}_i(j,k)$ to denote the action of an operator on the j-th and k-th qubits of the $\mathcal{Y}_i$ register.

$H_{\mathcal{Y}_i}$ has ground state energy less than a_i if query i is a YES instance and energy greater than b_i in a NO instance. Then, let $H_2 = \sum_i H_2^{(i)}$ where

$$H_2^{(i)} = \frac{a_i + b_i}{2}|0\rangle\langle 0|_{\mathcal{X}_i(1)} \otimes I_{\mathcal{Y}_i} + \sum_{(j,k)\in E_i} \left(|1\rangle\langle 1|_{\mathcal{X}_i(g(j,k))} \otimes h^i_{\mathcal{Y}_i(j,k)}\right),$$

where $g(j,k)$ is the location of the "nearest" qubit in $\mathcal{X}_i$ to edge (j,k) in $\mathcal{Y}_i$. Here, the choice "nearest" is somewhat arbitrary; for concreteness, one can set $g(j,k) = j$, i.e. pick the vertex in $\mathcal{X}_i$ which aligns with the first coordinate of the edge (j,k). (In this sense, Figure 2 is not entirely accurate, since it depicts the 3-local constraint $|1\rangle\langle 1|_{\mathcal{X}_i(g(j,k))} \otimes h^i_{\mathcal{Y}_i(j,k)}$ as a pair of 2-local constraints. This is done solely for the purpose of simplifying the illustration, as otherwise one would need to draw hyperedges of size 3.)

Let $H_1 = H_{\text{prop}} + H_{\text{in}}$ be the Cook-Levin Hamiltonian where H_{prop} is exactly as in Lemma 12. Let H_{in} initialize the qubits of the first ($t = 1$) row of the qubits in $\mathcal{W}$. For each query i, we have a penalty term $|1\rangle\langle 1|_{\mathcal{X}_i(1)}|0\rangle\langle 0| + |0\rangle\langle 0|_{\mathcal{X}_i(1)}|1\rangle\langle 1|$ which effectively copies the state of $\mathcal{X}_i(1)$, the qubit in $\mathcal{X}_i$ nearest to $\mathcal{W}$, onto the i-th qubit of the first row of $\mathcal{W}$. For all the remaining qubits in the first ($t = 1$) row of $\mathcal{W}$, we have a penalty term $|1\rangle\langle 1|$, effectively initializing the qubit into the $|0\rangle$ state.

Restricted to the subspace $\mathcal{H}$ where each $\mathcal{X}_i$ register is either all $|0\rangle$ or all $|1\rangle$, $H_1 + H_2$ is exactly the same Hamiltonian as in Lemma 12. It remains to give a high energy penalty to all other states not in this subspace. We do this with $H_3 = \sum_{i=1}^m H_3^{(i)}$ where each term $H_3^{(i)}$ acts on $\mathcal{X}_i$:

$$H_3^{(i)} = \Delta_i \sum_{(j,k)\in G_i} \left(|0\rangle\langle 0|_{\mathcal{X}_i(j)}|1\rangle\langle 1|_{\mathcal{X}_i(k)} + |1\rangle\langle 1|_{\mathcal{X}_i(j)}|0\rangle\langle 0|_{\mathcal{X}_i(k)}\right)$$

where G_i is the set of edges between the qubits of the $\mathcal{X}_i$ register. G_i consists of edges between nearest neighbors on the square lattice E_i and on the path of qubits from $\mathcal{Y}_i$ to $\mathcal{W}$. The overall Hamiltonian $H = H_1 + H_2 + H_3$ is therefore spatially sparse.

$H_3^{(i)}$ is a classical Hamiltonian, so all of its eigenstates can be taken to be of form $|x\rangle$ for some $x \in \{0,1\}^{n_i}$. Its ground space $\mathcal{G}_i$ contains $|0\rangle^{\otimes n_i}$ and $|1\rangle^{\otimes n_i}$; and all states in $\mathcal{G}_i^\perp$ have energy at least Δ_i. Choosing $\Delta_i > \delta + \sum_{(j,k)\in E_i} \|h^i_{\mathcal{Y}_i(j,k)}\|$ ensures that all states in $\mathcal{G}_i^\perp$ have energy greater than $\lambda(H) + \delta$.

Then $H = H_1 + H_2 + H_3$ is block diagonal with respect to the split of each subspace $\mathcal{G}_i \oplus \mathcal{G}_i^\perp$; restricted to the spaces $\mathcal{G}_i$, H is exactly the Hamiltonian from Lemma 12, and all states in spaces $\mathcal{G}_i^\perp$ have energy greater than $\lambda(H) + \delta$. The result then follows just as in the proof of Lemma 12. ◀

E Proofs for Section 6

We now give all details of our 1D hardness construction from Section 6, and prove correctness thereof in Section E.1.

Our 1D hardness construction. We give a reduction from $\text{P}^{||\text{QMA}}$ to $\forall$-APX-SIM, which by Theorem 2 yields the claim. Let Π be a $\text{P}^{||\text{QMA}}$ computation which takes in an input of size n and which consists of a uniformly generated polynomial-size classical circuit C making $m = O(\log n)$ 2-LH queries $\pi_i := (H_i, a_i, b_i)$ to a QMA oracle. As in Lemma 12, we treat the "answer register" in which C receives answers to its m queries as a proof register.

Our high-level approach consists of three steps: (1) construct a "master" circuit V composed of the verification circuits V_i corresponding to each query π_i and of the circuit C; (2) run V through the 1D circuit-to-Hamiltonian construction of [26] to obtain a 1D Hamiltonian G with local dimension 8 constructed such that the low-energy space $\mathcal{S}$ of G must consist of history states (of the form described in [26]); and (3) carefully add additional 1-local penalty terms acting on the output qubits corresponding to each verification circuit V_i to obtain final Hamiltonian H such that the low-energy space must encode satisfying proofs to each V_i whenever possible. This final step of "fine-grained splitting" of S forces the output qubits of the circuits V_i to encode correct answers to query π_i, and thus the final circuit C receives a correct proof, hence leading the history states of step (2) to encode a correct simulation of Π. The answer to the computation Π can then be read off the ground state of H via an appropriate single qudit measurement.

1. Construction of V. Suppose each query π_i has corresponding QMA verification circuit V_i. Without loss of generality, we may henceforth assume that the completeness/soundness error of V_i is at most $p \leq 2^{-n}$, for p to be set later, by standard error reduction [2, 34]; thus, if a particular query (H_i, a_i, b_i) is valid (i.e. $\lambda(H) \notin (a_i, b_i)$), then either there exists a proof such that V_i outputs YES with probability at least $1-p$ or no proof causes V_i to output YES with probability greater than p. Next, since Π is a $\mathrm{P}^{||\mathrm{QMA}}$ computation, all queries and corresponding V_i can be precomputed in polynomial-time. We view the "master circuit" V as consisting of two phases:

1. (Verification phase) Given supposed proofs for each query, V runs all verification circuits V_i in parallel, where V_i acts on space $\mathcal{Y}_i \otimes \mathcal{W}_i \otimes \mathcal{X}_i$, for proof register $\mathcal{Y}_i$, ancilla register $\mathcal{W}_i$, and single-qubit output register $\mathcal{X}_i$.
2. (Simulated classical phase) The simulated P circuit C now receives the query answers $\mathcal{X} := \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_m$ as its proof register as well as an ancilla register $\mathcal{W}_0$. It outputs a single qubit to an output register $\mathcal{X}_0$.

This completes the construction of V, which acts on $\mathcal{Y} \otimes \mathcal{W} \otimes \mathcal{X}$, where $\mathcal{Y} = \bigotimes_{i=1} \mathcal{Y}_i$,$\mathcal{W} = \bigotimes_{i=1} \mathcal{W}_i$, and $\mathcal{X} = \bigotimes_{i=1} \mathcal{X}_i$. Crucially, note that given a set of proofs in register $\mathcal{Y}$, V does *not* necessarily yield the same answer as Π, since a malicious prover could intentionally send a "bad" proof to a YES query, flipping the final answer of V.

2. Construction of G. We now plug V into the circuit-to-Hamiltonian construction of Hallgren, Nagaj, and Narayanaswami [26] to obtain a nearest-neighbor 1D Hamiltonian $G' = \Delta_{\mathrm{in}} H_{\mathrm{in}} + \Delta_{\mathrm{prop}} H_{\mathrm{prop}} + \Delta_{\mathrm{pen}} H_{\mathrm{pen}} + H_{\mathrm{out}}$, where $\Delta_{\mathrm{in}}, \Delta_{\mathrm{prop}}$, and Δ_{pen} are at most polynomials in n which we will set as needed; we review this construction more closely below. Set $G = G' - H_{\mathrm{out}}$, since in our setting the task of "checking the output" will be delegated to the observable A. Note that as an intermediate step, [26] maps V to a circuit V' which it then maps to G'; we describe the role of V' in the following review. Our construction will make two trivial assumptions about the behavior of V', including how it arranges its query answers between the verification phase and the simulated classical phase and how it stores its output in the final timestep; we defer details about these assumptions until we define our "fine-grained splitting" in step 3 and when we define our observable.

Review of 1D QMA construction [26]. Suppose an arbitrary circuit U acts on n qubits. Begin by arbitrarily arranging these qubits along a line. The circuit U is then "linearized", meaning it is mapped to a new circuit U' which consists of R rounds in which each round applies a sequence of $n-1$ two-qubit gates acting on nearest neighbors. The i-th gate in

a round acts on qubits $(i, i+1)$. This "linearization" is achieved in polynomial time by inserting swap and identity gates as needed, and U' is at most polynomially larger than U.

To reduce U' to an instance of k-LH, we wish to design a mapping similar to Kitaev's circuit-to-Hamiltonian construction for showing QMA-hardness of 5-LH on general geometry [31]. In both settings, the goal is to design an H which enforces a structure on any state in its low-energy space. In the construction of [31], $H = H_{\text{in}} + H_{\text{prop}} + H_{\text{stab}} + H_{\text{out}}$, and the minimizing state of H has the form of a *history state*:

$$|\eta\rangle = \frac{1}{\sqrt{L+1}} \sum_{t=0}^{L} U_t \cdots U_1 |\psi\rangle_{\mathcal{Y}} |0 \cdots 0\rangle_W |t\rangle_C .$$

Intuitively, H_{stab} forces a structure on the clock register C of basis states $|0\rangle, |1\rangle, \ldots,$ such that each will correspond to a timestep of U. Then, H_{in} ensures the ancilla register W is set to the all $|0\rangle$ state when $|t\rangle = |0\rangle$. The term H_{prop} ensures that the workspaces entangled with timesteps $|t\rangle$ and $|t+1\rangle$ are related by the 2-qubit gate U_{t+1}. Together, these terms ensure that a minimizing state $|\psi_{\text{hist}}\rangle$ encodes a correct simulation of the circuit U, and that all low-energy states are close to $|\psi_{\text{hist}}\rangle$. In fact, a valid $|\psi_{\text{hist}}\rangle$ lies in the nullspace of $H_{\text{in}} + H_{\text{prop}} + H_{\text{stab}}$. Finally, H_{out} penalizes the low-energy space if the output qubit has overlap with $|0\rangle$.

Now in the 1D setting, the goal remains the same: design H such that the structure of its low-energy state is a superposition over a sequence of states corresponding to timesteps in the computation of U'. But, we now appear unable to entangle the workspace with a separate clock register using nearest neighbor interactions. Instead, the constructions of [1, 26] employ qudits of higher dimension as a means to label the qubits, with each labeling encoding a particular timestep. [26] then doubles the number of qudits in order to lower the necessary number of labels. The construction of [26] thus maps U' to a Hamiltonian $H = H_{\text{in}} + H_{\text{prop}} + H_{\text{out}} + H_{\text{pen}}$ acting on $2nR$ qudits of dimension 8, where the qudits are arranged on a 1D line in R blocks of $2n$ qudits (i.e. one block per round in U').

Let us further describe the idea of labeling, or "marking", of qudits. For example, a qubit $\alpha|0\rangle + \beta|1\rangle$ may be encoded as $\alpha|A\rangle + \beta|B\rangle$ if that qubit is ready for a gate to be applied or as $\alpha|C\rangle + \beta|D\rangle$ if that round's gate has already been applied, where $|A\rangle, |B\rangle, |C\rangle, |D\rangle$ are some basis states. The possible configurations, or arrangements, of labels along the line form a set of orthogonal spaces. [26] thus introduces a Hamiltonian term H_{pen} which enforces a set of "legal configurations" of the workspace, penalizing all other configurations. We then map each of the configurations which remain in the low-energy space of H to timesteps in the computation of U', effectively assigning the job of encoding the workspace in a particular timestep to a particular configuration of qudits. We note that the crucial feature of the set of legal configurations developed by [26] is that they are sufficiently identifiable solely by 2-local nearest neighbor checks[5] such that penalties can be correctly assigned when constructing 1D analogs of the terms $H_{\text{in}}, H_{\text{prop}}, H_{\text{out}}$. Similar to the general geometry case of [31], the construction of [26] enforces that the nullspace of $H_{\text{in}} + H_{\text{prop}} + H_{\text{pen}}$ consists of history states

$$|\psi_{\text{hist}}\rangle = \frac{1}{\sqrt{L+1}} \sum_{t=0}^{L} |\psi_t\rangle , \tag{16}$$

[5] For clarity, in [26] not all illegal configurations are immediately detectable by H_{pen}. Any such undetectable illegal configurations are instead shown to eventually evolve under H_{prop} into detectable illegal configurations.

such that $|\psi_{\text{hist}}\rangle$ is a superposition over states in each legal configuration, $|\psi_0\rangle$ encodes a properly initialized workspace, and each pair $|\psi_t\rangle$ and $|\psi_{t+1}\rangle$ are related according to the corresponding timestep of U'. Finally, again similar to the general geometry case, all low-energy states must be close to $|\psi_{\text{hist}}\rangle$ (we make these two claims explicit and give proofs in Lemma 22).

The full description of the labeling, the legal configurations, and their mapping to timesteps by [26] is rather involved. Here, we introduce sufficient details for our later analysis. We begin with a single block of $2n$ qudits, where recall each block is used to encode a single round (taken from [26]):

$$\|\blacktriangleright\circledcirc\,|\square\circledcirc\,|\cdots|\square\circledcirc\,|\square\bigcirc\| \tag{17}$$

Recall the design of U' began by arranging the qubits of U arbitrarily on the line; the i-th qubit on that line corresponds to qudits $2i-1$ and $2i$ in (17). Thus, each qubit of U', henceforth denoted a *logical qubit*, is encoded into two consecutive qudits. Each pair of qudits representing a logical qubit is depicted as separated by a $|$ for clarity. The standard basis for each 8-dimensional qudit is labeled by

$$\{\,|\bigcirc\rangle,\ |\ominus\rangle,\ |\circledcirc\rangle,\ |\otimes\rangle,\ |\blacktriangleright_0\rangle,\ |\blacktriangleright_1\rangle,\ |\square_0\rangle,\ |\square_1\rangle\,\},$$

where, as described earlier, the current state of a qudit can be used to encode a logical qubit and to label the qudit. The first four states should be thought of as 1-dimensional labels; they are used to ensure the correct propagation of the circuit and do not encode a logical qubit. The final four states are used to either label a qudit with $\blacktriangleright$, in which case a logical qubit is encoded as a superposition of $|\blacktriangleright_0\rangle$ and $|\blacktriangleright_1\rangle$, or with $\square$, in which case a logical qubit is encoded as a superposition of $|\square_0\rangle$ and $|\square_1\rangle$. To make this example more concrete, a product state of $(\alpha|0\rangle+\beta|1\rangle)^{\otimes n}$ on n logical qubits could be encoded as

$$(\alpha\,|\blacktriangleright_0\rangle+\beta\,|\blacktriangleright_1\rangle)\otimes|\circledcirc\,\rangle\otimes(\alpha\,|\square_0\rangle+\beta\,|\square_1\rangle)\otimes|\circledcirc\,\rangle\otimes\cdots \tag{18}$$

$$\otimes\,(\alpha\,|\square_0\rangle+\beta\,|\square_1\rangle)\otimes|\bigcirc\,\rangle\,. \tag{19}$$

Next, here is an example depicting multiple blocks (from Table 2 of [26]):

$$\cdots\otimes\otimes\|\blacktriangleright\circledcirc\,|\square\circledcirc\,|\square\bigcirc\|\bigcirc\,\bigcirc\,|\bigcirc\,\bigcirc\,|\bigcirc\,\bigcirc\|\bigcirc\,\bigcirc\cdots, \tag{20}$$

where the blocks are delineated by $\|$. The labels $\otimes$ to the left depict "dead" qudits, while the labels $\bigcirc$ to the right depict "unborn" qudits. By construction, all logical qubits are encoded in a block between the dead and unborn labels. In this example, the logical qubits line up with the beginning of a new block, beginning with $\|\blacktriangleright$ and ending with the first $\bigcirc\|$.

At a high level, the set of legal configurations is mapped to a sequence of timesteps as follows. The first timestep corresponds to a configuration similar to (17), with n logical qubits encoded in the leftmost block of $2n$ qudits, with no $\otimes$ labels anywhere, and with the "gate" label $\blacktriangleright$ on the first qudit. The second configuration has the $\blacktriangleright$ label shifted to the right, on the second qudit. Next, the third configuration has the second qudit labeled $\square$ and the third qudit labeled $\blacktriangleright$. This propagation of the $\blacktriangleright$ label rightwards continues, with each step corresponding to another legal configuration, until it reaches the end of the block. As the $\blacktriangleright$ passes between logical qubits $(i, i+1)$, the corresponding configurations map to timesteps i and $i+1$ of round 1, and H_{prop} enforces that configurations are related by the application of gate U'_i. Thus, when we reach a configuration with $\blacktriangleright$ at the end of the block, i.e. $\blacktriangleright\|$, all gates in the current round will have been applied. Next, before encoding the

next round of gates, our goal becomes to shift all of the logical qubits encoded in the current block rightwards $2n$ spots into the second block. To do this, the $\blacktriangleright$ label becomes a special $\ominus$ label and moves to the left one spot at a time until it reaches the end of the logical qubits (here, the leftwards $\|$). As the label $\ominus$ moves left, it shifts each logical qubit to the right one spot, i.e. $|\square\ominus\rangle \to |\ominus\square\rangle$. This process repeats, with a label propagating rightwards to the end of the logical qubits (now past the rightwards $\|$), then the label $\ominus$ propagating to the left, shifting logical qubits to the right, and so on, until the logical qubits have shifted entirely into the second block. Then, the gate label $\blacktriangleright$ once again transitions down the line, with successive configurations encoding the second round of gates of U'. Throughout this sequence, $\bigcirc$ labels to the right are consumed, while all qudits to the left are labeled $\otimes$. This procedure continues until the entire circuit has been simulated.

Lastly, we observe that the final timestep of U' is encoded by [26] in the following configuration:

$$\cdots \otimes \otimes \| \otimes \otimes \| \otimes \square | \circledcirc \square | \cdots | \circledcirc \square | \circledcirc \blacktriangleright \| \tag{21}$$

3. Adding 1-local "sifters". We now add 1-local Hamiltonian terms which serve to "sift" through bad proofs, or more accurately to split the ground space of G, so as to force low-energy states to encode correct query answers. As previously described, even a correct simulation of the circuit V may not output the correct answer for instance Π if a malicious prover supplies incorrect proofs to the query registers $\mathcal{Y}_i$; in particular, a prover might send a proof which accepts with low probability even though π_i is a YES-instance. Intuitively, we wish to penalize states encoding a proof $|\psi_i\rangle$ which leads verifier V_i to reject with high probability when there exists a proof $|\phi_i\rangle$ such that V_i would have accepted with high probability (here, query π_i is a YES instance). For answer register $\mathcal{X}_i$, we add a "sifter" penalty term $\epsilon\,|0\rangle\langle 0|_{\mathcal{X}_i}$, for ϵ some inverse polynomial to be set later. These terms are similar to the H_{out} term from other Hamiltonian constructions; but, here we are not only concerned about the ground space but also about the low-energy space. As in other constructions, we must penalize NO answers enough to ensure the ground space encodes YES answers when possible. But, given a correct NO answer, the penalty must be small enough that the energy is gapped lower than any state which encodes an incorrect YES, such as those which by encode an invalid computation leading to YES.

However, because the encoding enforced by G shifts the block of logical qubits rightwards along the line as the computation progresses, the location of a particular logical qubit's encoding depends on the current timestep. Thus, in order to properly act on logical qubit $\mathcal{X}_i$, we must be careful to specify the configuration which the penalty term acts on.

We may assume that once V' finishes simulating all of the circuits V_i, it arranges each of the outputs in the first m logical qubits on the line, finishing by the end of some round $r^* - 1$, such that the i-th logical qubit on the line is the qubit which V stored in $\mathcal{X}_i$. (The value of r^* can be determined during the construction of V'.) We may also assume that V' then "pauses" by applying only identity gates in round r^*. This round is encoded in block r^*, and since each block is comprised of $2n$ qudits, the answers to queries 1 to m are thus simultaneously stored in qudits

$$q_i := (2n)(r^* - 1) + (2i - 1). \tag{22}$$

The m sifter terms are given by

$$H_{\text{out},i} = \epsilon\,|\blacktriangleright_0\rangle\langle\blacktriangleright_0|_{q_i},$$

where the subscript denotes the qudit which the term acts on and ϵ is to be set later. Note that there is a unique legal configuration in which any given qudit is labeled $\blacktriangleright$, so $H_{\text{out},i}$ will apply to at most one state $|\psi_t\rangle$ in the history state of Equation (16). Finally, we define $H_{\text{out}} = \sum_{i=1}^{m} H_{\text{out},i}$.

The final Hamiltonian. Our final Hamiltonian is $H := G + H_{\text{out}} = \Delta_{\text{in}} H_{\text{in}} + \Delta_{\text{prop}} H_{\text{prop}} + \Delta_{\text{pen}} H_{\text{pen}} + H_{\text{out}}$, with $\Delta_{\text{in}}, \Delta_{\text{prop}}, \Delta_{\text{pen}}$ polynomials to be set later.

The observable. Recall the configuration from (21), which corresponds to the final timestep in the computation of a circuit passed to the construction of [26]. Note that this is the unique timestep in which the final qudit is labeled $\blacktriangleright$. We assume, without loss of generality, that V' places its final output in the rightmost logical qubit on the line. Thus, we choose single-qudit observable $A = |\blacktriangleright_0\rangle\langle\blacktriangleright_0|_{2nR}$, where the subscript denotes that A acts on the rightmost qudit on the line, where R is the number of rounds in V'.

Setting parameters. Let L denote the number of legal configurations which the history state in (16) is summed over, which is at most polynomial in n. We have that H is k-local and A is ℓ-local for $k := 2$ and $\ell := 1$. Set $\epsilon = 1/(8m)$, where recall m is the (polynomial) number of queries. Then, set p, the completeness/soundness error of each V_i, to some inverse-exponential in n such that $p < \epsilon$ for all n. Set $a = 1/(4L)$ and $b = 3/(4L)$. We will set δ to a sufficiently small fixed inverse polynomial in n in the proof of Lemma 23, which will then set $\Delta_{\text{in}}, \Delta_{\text{prop}}, \Delta_{\text{pen}}$ to sufficiently large fixed polynomials in n via the proof of Lemma 22.

This concludes our deterministic polynomial-time mapping of the input $\text{P}^{||\text{QMA}}$ computation Π to the 1D instance $\tilde{\Pi} := (H, A, k, \ell, a, b, \delta)$ of $\forall$-APX-SIM.

E.1 Correctness

We now prove Theorem 10 by showing correctness of our construction from Section 6. A number of lemmas required in the proof are deferred to Section E.1.1 to ease the exposition; in particular, we require Lemma 22, which explicitly proves two facts about the low-energy space of the construction of [26], Lemma 23, which shows that a history state in our construction must simultaneously encode nearly correct answers for all valid queries π_i, and Lemma 24, which states a Commutative Quantum Union Bound.

Proof of Theorem 10. Containment in $\text{P}^{\text{QMA}[\log]}$ was already shown for up to $O(\log n)$-local H by [3], with no restriction on the geometry. Our goal is now to show $\text{P}^{||\text{QMA}}$-hardness, which by Theorem 2 yields $\text{P}^{\text{QMA}[\log]}$-hardness. We show hardness for the problem $\forall$-APX-SIM, which recall from Section 1 trivially reduces to APX-SIM, thus yielding hardness for APX-SIM. Let Π be a $\text{P}^{||\text{QMA}}$ computation and map it to the $\forall$-APX-SIM instance $\tilde{\Pi} = (H, A, k, l, a, b, \delta)$ as described in Section 6. The proof proceeds in two parts: We first show that low energy states must necessarily encode correct query answers, and subsequently apply this to show correctness in YES and NO cases for Π.

Low energy states approximately encode correct query answers. Recall that $H = G + H_{\text{out}}$. Let δ, γ denote arbitrary inverse polynomials in n which will be set later in Lemma 23. Consider any state $|\psi\rangle$ such that $\langle\psi| H |\psi\rangle \leq \lambda(H) + \delta$. Since $H_{\text{out}} \succeq 0$, $\langle\psi| G |\psi\rangle \leq \lambda(H) + \delta$ as well. By Lemma 22, for sufficiently large fixed polynomials $\Delta_{\text{in}}, \Delta_{\text{prop}}, \Delta_{\text{pen}}$, two statements thus hold: First, the nullspace $\mathcal{S}$ of Hamiltonian $G = \Delta_{\text{in}} H_{\text{in}} + \Delta_{\text{prop}} H_{\text{prop}} + \Delta_{\text{pen}} H_{\text{pen}}$ is

the span of all correctly encoded history states, as defined in Equation (16); Second, there exists a correctly encoded history state $|\psi_{\text{hist}}\rangle$ such that

$$\| |\psi\rangle\langle\psi| - |\psi_{\text{hist}}\rangle\langle\psi_{\text{hist}}| \|_{\text{tr}} \leq \gamma. \tag{23}$$

Combining Equation (23) with the Hölder Inequality and the fact that $\| H_{\text{out}} \|_\infty = m\epsilon$ yields that

$$|\operatorname{Tr}[H_{\text{out}} |\psi\rangle\langle\psi|] - \operatorname{Tr}[H_{\text{out}} |\psi_{\text{hist}}\rangle\langle\psi_{\text{hist}}|]| \leq \gamma \| H_{\text{out}} \|_\infty = m\epsilon\gamma.$$

Since $|\psi_{\text{hist}}\rangle$ is a nullstate of G and $\langle\psi| H_{\text{out}} |\psi\rangle \leq \langle\psi| H |\psi\rangle \leq \lambda(H) + \delta$, we conclude

$$\langle\psi_{\text{hist}}| H |\psi_{\text{hist}}\rangle \leq \lambda(H) + \delta + m\epsilon\gamma. \tag{24}$$

Next, let $I \subseteq \{1, \ldots, m\}$ be the set of indices corresponding to valid queries π_i, and for all $i \in I$ define $x_i = 1$ if π_i is a YES-instance and $x_i = 0$ if π_i is a NO-instance.[6] Recall now from Section 6 that at the beginning of round r^*, V' has encoded the answer to the i-th QMA query in qudit q_i (defined in Equation (22)). Let $|\psi_{t^*}\rangle$ denote the unique (normalized) state in the superposition comprising $|\psi_{\text{hist}}\rangle$ in which q_1 is labeled ▶ (i.e. the first timestep corresponding to round r^*). Since during round r^*, V' only applies identity gates, the qubits encoded in qudits q_i during timestep t^*, in which q_1 is labeled ▶ and all other q_i are labeled □, are exactly the same as in successive timesteps in which other q_i are labeled by ▶. More formally, $|\langle\psi_{t^*}|\square_{x_i}\rangle_{q_i}|^2 = L|\langle\psi_{\text{hist}}|\blacktriangleright_{x_i}\rangle_{q_i}|^2$ for any $i \in I$, and so by Lemma 23,

$$\left|\langle\psi_{t^*}|\square_{x_i}\rangle_{q_i}\right|^2 \geq 1 - \epsilon, \tag{25}$$

where[7] we substitute the label ▶ for □ when $i = 1$, and where the factor of L^{-1} is removed due to the normalization of $|\psi_{t^*}\rangle$.

This is for any single query $\pi_i, i \in I$; from this, we can obtain that $|\psi_{t^*}\rangle$ simultaneously encodes nearly correct query answers to *all* valid queries. To do so, define $\Gamma := \Pi_{i\in I} |\square_{x_i}\rangle\langle\square_{x_i}|_{q_i}$ (where again, we replace label ▶ for □ when $i = 1$). Then, by the Commutative Quantum Union Bound (Lemma 24),

$$\langle\psi_{t^*}| \Gamma |\psi_{t^*}\rangle \geq 1 - |I|\,\epsilon \geq 1 - m\epsilon. \tag{26}$$

It follows that we may write $|\psi_{t^*}\rangle = \alpha |\phi_1\rangle + \beta |\phi_2\rangle$ for unit vectors $|\phi_1\rangle, |\phi_2\rangle$ such that $\Gamma |\phi_1\rangle = |\phi_1\rangle$ and $\Gamma |\phi_2\rangle = 0$, and where $\alpha, \beta \in \mathbb{C}, |\alpha|^2 + |\beta|^2 = 1$, and $|\alpha|^2 \geq 1 - m\epsilon$. Intuitively, $|\phi_1\rangle$ is the part of $|\psi_{t^*}\rangle$ that encodes correct strings of query answers on I, while $|\phi_2\rangle$ encodes strings with at least one incorrect query answer in I – for clarity, $|\phi_1\rangle$ may encode a superposition of multiple *distinct* correct strings of query answers, since queries with indices not in I may be answered arbitrarily.

---

[6] Without loss of generality, we may assume at least one query is valid ($I \neq \emptyset$). This is because if all queries are invalid, then all simulations of the P circuit C must output the same answer no matter the sequence of query answers C receives. Thus, all history states will encode the same final answer, and α (defined after (26)) equals 1, satisfying the lower bound found of $\alpha \geq 1 - m\epsilon$.

[7] We implicitly apply identity on all qudits other than q_i, i.e. $\left|\langle\psi_{\text{hist}}|\square_{x_i}\rangle_{q_i}\right|^2 := \operatorname{Tr}\left[|\psi_{\text{hist}}\rangle\langle\psi_{\text{hist}}| \left(I \otimes |\square_{x_i}\rangle\langle\square_{x_i}|_{q_i} \otimes I\right)\right]$.

Application to YES versus NO cases for Π. We have shown that for any low energy state $|\psi\rangle$, there exists a history state $|\psi_{\text{hist}}\rangle$ close to $|\psi\rangle$ which has large amplitude on all the correct query answers for set I in round r^*. We can now analyze the YES and NO cases for our $\text{P}^{\text{QMA[log]}}$ problem Π.

Recall that $|\phi_1\rangle$ may be a superposition over *multiple* correct query strings (due to invalid queries π_i for $i \notin I$). Nevertheless, since the classical circuit C for the $\text{P}^{\text{QMA[log]}}$ machine is required to output the *same* answer regardless of how invalid queries are answered (i.e. for any given correct string of query answers), all query strings which $|\phi_1\rangle$ is a superposition over lead C to output the same, correct final answer. Thus, setting $y = 0$ if Π is a YES-instance and $y = 1$ if Π is a NO-instance, we have

$$\left|\langle\psi_{\text{hist}}| A |\psi_{\text{hist}}\rangle - \frac{y}{L}\right| \leq \frac{m\epsilon}{L},$$

where the factor of L^{-1} is due to the fact A applies only to the final configuration/time step. Combining Equation (23) with the Hölder inequality yields that

$$|\operatorname{Tr}[A|\psi\rangle\langle\psi|] - \operatorname{Tr}[A|\psi_{\text{hist}}\rangle\langle\psi_{\text{hist}}|]| \leq \gamma,$$

since $\|A\|_\infty = 1$, and so

$$\left|\langle\psi| A |\psi\rangle - \frac{y}{L}\right| \leq \frac{m\epsilon}{L} + \gamma,$$

Given that we set $\delta = \gamma = 1/(256m^2L) < 1/(8L)$ in Lemma 23 and $\epsilon = 1/(8m)$, we have that $\gamma + m\epsilon/L < 1/(4L)$. We conclude that for all low-energy states $|\psi\rangle$ (i.e. states satisfying $\langle\psi| H |\psi\rangle \leq \lambda(H) + \delta$), if Π is a YES-instance then $\langle\psi| A |\psi\rangle \leq 1/(4L)$ (i.e. we have a YES instance of ∀-APX-SIM), and if Π is a NO-instance then $\langle\psi| A |\psi\rangle \geq 3/(4L)$ (i.e. we have a NO instance of ∀-APX-SIM), as desired. ◀

E.1.1 Required lemmas for proof of Theorem 10

We begin by restating a known lemma and corollary.

▶ **Lemma 20** (Kempe, Kitaev, Regev [28]). *Let $H = H_1 + H_2$ be the sum of two Hamiltonians operating on some Hilbert space $\mathcal{H} = \mathcal{S} + \mathcal{S}^\perp$. The Hamiltonian H_1 is such that $\mathcal{S}$ is a zero eigenspace and the eigenvectors in $\mathcal{S}^\perp$ have eigenvalue at least $J > 2\|H_2\|_\infty$. Then,*

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

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

▶ **Corollary 21** ([22]). *Let $H = H_1 + H_2$ be the sum of two Hamiltonians operating on some Hilbert space $\mathcal{H} = \mathcal{S} + \mathcal{S}^\perp$. The Hamiltonian H_1 is such that $\mathcal{S}$ is a zero eigenspace and the eigenvectors in $\mathcal{S}^\perp$ have eigenvalue at least $J > 2\|H_2\|_\infty$. Let $K := \|H_2\|_\infty$. Then, for any $\delta \geq 0$ and vector $|\psi\rangle$ satisfying $\langle\psi| H |\psi\rangle \leq \lambda(H) + \delta$, there exists a $|\psi'\rangle \in \mathcal{S}$ such that*

$$\||\psi\rangle\langle\psi| - |\psi'\rangle\langle\psi'|\|_{\text{tr}} \leq 2\left(\frac{K + \sqrt{K^2 + \delta(J - 2K)}}{J - 2K}\right).$$

We now prove the lemmas required for Theorem 10.

▶ **Lemma 22.** *Assume the notation of Section 6. For $G = \Delta_{\text{in}}H_{\text{in}} + \Delta_{\text{prop}}H_{\text{prop}} + \Delta_{\text{pen}}H_{\text{pen}}$, the following hold:*

1. *For sufficiently large (efficiently computable) polynomials* $\Delta_{\text{in}}, \Delta_{\text{prop}}, \Delta_{\text{pen}}$*, the null space of* G *is the span of all correctly encoded history states, i.e. of the form in Equation (16).*
2. *For any fixed inverse polynomials* δ *and* γ*, there exist efficiently computable polynomials* $\Delta_{\text{in}}, \Delta_{\text{prop}}, \Delta_{\text{pen}}$ *such that for any* $|\psi\rangle$ *attaining* $\langle\psi| G |\psi\rangle \leq \lambda(G) + \delta$*, there exists a correctly encoded history state* $|\psi_{\text{hist}}\rangle$ *such that*

$$\| |\psi\rangle\langle\psi| - |\psi_{\text{hist}}\rangle\langle\psi_{\text{hist}}| \|_{\text{tr}} \leq \gamma.$$

Proof. The analysis of G is more subtle than that of, say, the 5-local Kitaev circuit-to-Hamiltonian construction [31]. The latter required the analysis of two orthogonal subspaces acted on invariantly by the Hamiltonian in question; the span of all correctly encoded history states, and the span of all states with an incorrectly encoded clock register (i.e. illegal configurations). In [26], however, due to the restrictions of encoding in 1D, there are *two* types of illegal configurations which can arise – those which are detectable by local checks, and those which are not – and G does not act invariantly on the spaces of legal and illegal configurations. The soundness analysis of the QMA-hardness construction of [26] (see Section 6 therein, which we follow below) hence independently analyzes *three* types of subspaces which are acted on invariantly by H_{prop}: (1) The span of legal configurations and certain locally detectable illegal configurations, (2) the span of certain other locally detectable illegal configurations, and (3) the span of illegal configurations which are not locally detectable. We shall henceforth refer to these subspaces as S_1, S_2, and S_3, respectively.

Proof of claim 1. This claim is implicit in [26]; we sketch a proof to make it explicit here. Claim 2 of [26] and the subsequent discussion explicitly show that any valid history state is a null state of G. For the reverse containment, Section 6.2 of [26] shows that for sufficiently large polynomials $\Delta_{\text{in}}, \Delta_{\text{prop}}, \Delta_{\text{pen}}$, $\lambda((\Delta_{\text{prop}} H_{\text{prop}} + \Delta_{\text{pen}} H_{\text{pen}})|_{S_3}) \in \Omega(1)$. That $\lambda(G|_{S_2}) \geq \Delta_{\text{pen}}$ follows since H_{pen} is a sum of pairwise commuting projectors. Thus, $\text{Null}(G)$ resides in S_1. Section 6.1 of [26] shows that $\text{Null}(H_{\text{prop}}|_{S_1 \cap \text{Null}(H_{\text{pen}})})$ is spanned by valid history states. We conclude that the span of all valid history states contains $\text{Null}(G)$.

Proof of claim 2. We know from claim 1 that $\text{Null}(G)$ is precisely the span of all correctly encoded history states. Let $\mathcal{C}$ denote the orthogonal complement of $\text{Null}(G)$. Then, we know from the proof of claim 1 that $\lambda(G|_{\mathcal{C} \cap S_2}) \geq \Delta_{\text{pen}} \in \Omega(1)$, and that $\lambda((\Delta_{\text{prop}} H_{\text{prop}} + \Delta_{\text{pen}} H_{\text{pen}})|_{\mathcal{C} \cap S_3}) \in \Omega(1)$. (Here we have used the fact that $S_2 \cup S_3 \subseteq \mathcal{C}$.) Since δ is assumed to be inverse polynomial in n, and since we know from claim 1 that $\lambda(H) \leq 0$, it follows that no vector $|\psi\rangle$ from S_2 or S_3 can attain $\langle\psi| G |\psi\rangle \leq \lambda(G) + \delta$.

We are thus reduced to the case $|\psi\rangle \in S_1$, which we prove using three applications of Corollary 21. (To reduce notation, in the remainder of this proof all operators are implicitly restricted to S_1.) In the first application, let $H_1 = \Delta_{\text{pen}} H_{\text{pen}}$ and $H_2 = \Delta_{\text{in}} H_{\text{in}} + \Delta_{\text{prop}} H_{\text{prop}}$. Suppose $\langle\psi| H_1 + H_2 |\psi\rangle \leq \lambda(H) + \delta$. Then by Lemma 21, there exists a vector $|\psi'\rangle \in \text{Null}(H_{\text{pen}})$ such that

$$\| |\psi\rangle\langle\psi| - |\psi'\rangle\langle\psi'| \|_{\text{tr}} \leq 2 \left(\frac{K_1 + \sqrt{K_1^2 + \delta(J_1 - 2K_1)}}{J_1 - 2K_1} \right) =: 2\gamma_1,$$

for $K_1 := \|H_2\|_\infty$ and $J_1 > 2K_1$. (Note that since $\Delta_{\text{pen}} H_{\text{pen}}$ is a sum of commuting projectors, its smallest non-zero eigenvalue is at least Δ_{pen}, i.e. $J \geq \Delta_{\text{pen}}$.) By the Hölder inequality,

$$|\text{Tr}((H_1 + H_2) |\psi\rangle\langle\psi|) - \text{Tr}((H_1 + H_2) |\psi'\rangle\langle\psi'|)| \leq 2\gamma_1 \|H_1 + H_2\|_\infty =: \epsilon_1. \tag{27}$$

Combining these facts, we have

$$\begin{aligned}\langle\psi'|\,(H_1+H_2)|_{\mathrm{Null}(H_{\mathrm{pen}})}\,|\psi'\rangle &= \langle\psi'|\,(H_1+H_2)\,|\psi'\rangle \\ &\leq \lambda((H_1+H_2)) + \delta + \epsilon_1 \\ &\leq \lambda((H_1+H_2)|_{\mathrm{Null}(H_{\mathrm{pen}})}) + \delta + \epsilon_1 \\ &=: \lambda((H_1+H_2)|_{\mathrm{Null}(H_{\mathrm{pen}})}) + \delta_2, \end{aligned} \tag{28}$$

where the first statement holds since $|\psi'\rangle \in \mathrm{Null}(H_{\mathrm{pen}})$, the second by Equation (27), and the third by the Projection Lemma (this follows directly since projections can only increase the smallest eigenvalue).

We now repeat the process for $H_1 = \Delta_{\mathrm{prop}} H_{\mathrm{prop}}|_{\mathrm{Null}(H_{\mathrm{pen}})}$ and $H_2 = \Delta_{\mathrm{in}} H_{\mathrm{in}}|_{\mathrm{Null}(H_{\mathrm{pen}})}$. The key observation (used also in [26]) is that restricted to $S_1 \cap \mathrm{Null}(H_{\mathrm{pen}})$, H_{prop} is now positive semidefinite, has a 1-dimensional null space spanned by the correct history state (the action of H_{prop} ignores the initial setting of ancilla qubits, including the proof register, which in general leads to multiple correct history states), and its smallest non-zero eigenvalue is at least $1/(2(L+1)^2)$ (recall L is the number of time steps a valid history state sums over). Thus, by Lemma 21, there exists a vector $|\psi''\rangle \in \mathrm{Null}(H_{\mathrm{pen}}) \cap \mathrm{Null}(H_{\mathrm{prop}})$ such that

$$\| |\psi'\rangle\langle\psi'| - |\psi''\rangle\langle\psi''| \|_{\mathrm{tr}} \leq 2\left(\frac{K_2 + \sqrt{K_2^2 + \delta_2(J_2 - 2K_2)}}{J_2 - 2K_2}\right) =: 2\gamma_2,$$

for $K_2 := \|H_2\|_\infty$ and $J_2 > 2K_2$. Note that $J_2 \geq \Delta_{\mathrm{prop}}/(2(L+1)^2)$. By the Hölder inequality,

$$|\mathrm{Tr}((H_1+H_2)\,|\psi'\rangle\langle\psi'|) - \mathrm{Tr}((H_1+H_2)\,|\psi''\rangle\langle\psi''|)| \leq 2\gamma_2 \|H_1+H_2\|_\infty =: \epsilon_2,$$

which yields

$$\begin{aligned}\langle\psi''|\,(H_1+H_2)|_{\mathrm{Null}(H_{\mathrm{prop}})}\,|\psi''\rangle &= \langle\psi''|\,(H_1+H_2)\,|\psi''\rangle \\ &\leq \lambda((H_1+H_2)) + \delta_2 + \epsilon_2 \\ &\leq \lambda((H_1+H_2)|_{\mathrm{Null}(H_{\mathrm{prop}})}) + \delta_2 + \epsilon_2 \\ &=: \lambda((H_1+H_2)|_{\mathrm{Null}(H_{\mathrm{prop}})}) + \delta_3. \end{aligned}$$

Finally, we repeat the process for $H_1 = \Delta_{\mathrm{in}} H_{\mathrm{in}}|_{\mathrm{Null}(H_{\mathrm{pen}})\cap\mathrm{Null}(H_{\mathrm{prop}})}$ and $H_2 = 0$. Since by claim 1 we know the joint null space of $H_{\mathrm{in}}, H_{\mathrm{prop}}, H_{\mathrm{pen}}$ is non-empty, by Lemma 21, there exists a vector $|\psi'''\rangle \in \mathrm{Null}(H_{\mathrm{pen}}) \cap \mathrm{Null}(H_{\mathrm{prop}}) \cap \mathrm{Null}(H_{\mathrm{in}})$ such that

$$\| |\psi''\rangle\langle\psi''| - |\psi'''\rangle\langle\psi'''| \|_{\mathrm{tr}} \leq 2\sqrt{\frac{\delta_3}{J_3}} =: 2\gamma_3,$$

for $J_3 > 0$. Note that $J_3 \geq \Delta_{\mathrm{in}}$ since H_{in} is a sum of commuting projectors. By claim 1, since $|\psi'''\rangle$ is in the joint null space of $H_{\mathrm{in}}, H_{\mathrm{prop}}, H_{\mathrm{pen}}$, it is a correctly encoded history state; denote it $|\psi_{\mathrm{hist}}\rangle$. By the triangle inequality we have

$$\| |\psi\rangle\langle\psi| - |\psi_{\mathrm{hist}}\rangle\langle\psi_{\mathrm{hist}}| \|_{\mathrm{tr}} \leq 2(\gamma_1+\gamma_2+\gamma_3).$$

The claim now follows by observing that all variables involved, i.e. δ_2, δ_3, ϵ_1, ϵ_2, γ_1, γ_2, γ_3, J_1, J_2, J_3, decrease inverse polynomially in (a non-empty subset of) polynomials $\Delta_{\mathrm{in}}, \Delta_{\mathrm{prop}}, \Delta_{\mathrm{pen}}$. Thus, for any desired target accuracy q, we may attain the claim by setting $\Delta_{\mathrm{in}}, \Delta_{\mathrm{prop}}, \Delta_{\mathrm{pen}}$ as sufficiently large polynomials. (Note that this requires upper bounding terms of the form $K_2 := \|H_2\|_\infty$, which is easily done via triangle inequality of the spectral norm and the fact that projections can only decrease maximum eigenvalues.) ◀

▶ **Lemma 23.** *Assume the notation of Section E.1. For all $i \in I$, it holds that*

$$\left| \langle \psi_{\text{hist}} | \blacktriangleright_{x_i} \rangle_{q_i} \right|^2 \geq \frac{1-\epsilon}{L}, \tag{29}$$

where recall q_i is the index of the qudit which encodes the output corresponding to query π_i following the verification phase.

Proof. For clarity, the factor of L^{-1} comes from the L configurations which $|\psi_{\text{hist}}\rangle$ is a sum over. Recall there is a unique configuration in which any given qudit is labeled $\blacktriangleright$, implying all history states $|\psi_{\text{hist}}\rangle$ satisfy

$$\left| \langle \psi_{\text{hist}} | \blacktriangleright_0 \rangle_{q_i} \right|^2 + \left| \langle \psi_{\text{hist}} | \blacktriangleright_1 \rangle_{q_i} \right|^2 = \frac{1}{L}. \tag{30}$$

We prove our claim by contradiction via an exchange argument. Suppose there exists a valid query[8] π_j with correct answer x_j such that

$$\left| \langle \psi_{\text{hist}} | \blacktriangleright_{x_j} \rangle_{q_j} \right|^2 < \frac{1-\epsilon}{L}.$$

Since $|\psi_{\text{hist}}\rangle$ is a correctly encoded history state, we claim π_j must be a YES-instance. For if π_j were a NO-instance, then all simulations of V_j (on any possible proof) output NO with probability at least $1-p$. Thus, $|\psi_{\text{hist}}\rangle$ always encodes an output qubit such that

$$\left| \langle \psi_{\text{hist}} | \blacktriangleright_0 \rangle_{q_j} \right|^2 \geq \frac{1-p}{L} \geq \frac{1-\epsilon}{L},$$

which would contradict our supposition.

Given that π_j is a YES-instance, we have that $\left| \langle \psi_{\text{hist}} | \blacktriangleright_1 \rangle_{q_j} \right|^2 \leq (1-\epsilon)/L$, and so by Equation (30), $\langle \psi_{\text{hist}} | H_{\text{out},j} | \psi_{\text{hist}} \rangle \geq \epsilon^2/L$. Further, since π_j is a YES-instance, there exists a QMA proof $|\omega\rangle$ which causes V_j to output YES with probability at least $1-p$. By exchanging the QMA proof which $|\psi_{\text{hist}}\rangle$ encodes for circuit V_j with the proof $|\omega\rangle$, we obtain a new history state $|\psi'_{\text{hist}}\rangle$ which satisfies

$$\left| \langle \psi'_{\text{hist}} | \blacktriangleright_1 \rangle_{q_j} \right|^2 \geq \frac{1-p}{L},$$

and so $\langle \psi'_{\text{hist}} | H_{\text{out},j} | \psi'_{\text{hist}} \rangle \leq p\epsilon/L$. Hence,

$$\langle \psi_{\text{hist}} | H_{\text{out},j} | \psi_{\text{hist}} \rangle - \langle \psi'_{\text{hist}} | H_{\text{out},j} | \psi'_{\text{hist}} \rangle \geq \frac{(\epsilon-p)\epsilon}{L}, \tag{31}$$

i.e. flipping the incorrect query answer saves a non-trivial energy penalty on $H_{\text{out},j}$.

We now use this to obtain the desired contradiction. Recall that $H = G + H_{\text{out}}$. We make two observations: First, because all the QMA queries are made in parallel, flipping the answer to query π_j does not affect the other queries the P machine makes or the answers it receives. Thus, $|\psi_{\text{hist}}\rangle$ and $|\psi'_{\text{hist}}\rangle$ obtain the same energy on all terms of H_{out} other than $H_{\text{out},j}$, and Equation (31) holds for H_{out} in place of $H_{\text{out},j}$. (Analyzing adaptive queries, rather than parallel, would require that penalties for later queries be carefully weighted less than penalties for earlier queries [3], leading to a significantly more involved analysis.)

[8] If all queries are invalid, then Lemma 23 holds vacuously.

Second, both $|\psi_{\text{hist}}\rangle$ and $|\psi'_{\text{hist}}\rangle$ are null states of G, and so we may substitute H for H_{out}, yielding

$$\langle\psi_{\text{hist}}| H |\psi_{\text{hist}}\rangle - \langle\psi'_{\text{hist}}| H |\psi'_{\text{hist}}\rangle \geq \frac{(\epsilon - p)\epsilon}{L}. \tag{32}$$

Now, recall from Equation (24) that $\langle\psi_{\text{hist}}| H |\psi_{\text{hist}}\rangle \leq \lambda(H) + \delta + m\epsilon\gamma$. Since δ and γ are inverse polynomials which (by Lemma 22) we are free to choose as needed (the choice of δ and γ, in turn, will mandate the choices of $\Delta_{\text{in}}, \Delta_{\text{prop}}, \Delta_{\text{pen}}$ via Lemma 22), we set $\delta = \gamma = 1/(256m^2L)$ (where recall L and m are fixed polynomials in n). These choices of δ, γ satisfy $\delta + m\epsilon\gamma < (\epsilon - p)\epsilon/L$, which combined with Equation (32) gives that $\langle\psi_{\text{hist}}| H |\psi_{\text{hist}}\rangle > \lambda(H) + \delta + m\epsilon\gamma$, i.e. $|\psi_{\text{hist}}\rangle$ could not have been close to the ground state energy of H. Hence, we have a contradiction, completing the proof. ◀

Finally, we require a known quantum analogue of the union bound for commuting operators (see, e.g. [38]). Generalizations to *non-commuting* projectors are given in [43, 17, 38].

▶ **Lemma 24** (Commutative Quantum Union Bound). *Let $\{P_i\}_{i=1}^m$ be a set of pairwise commuting projectors, each satisfying $0 \preceq P_i \preceq I$. Then for any quantum state ρ,*

$$1 - \text{Tr}(\Pi_m \cdots P_1 \rho P_1 \cdots \Pi_m) \leq \sum_{i=1}^{m} \text{Tr}((I - P_i)\rho).$$

The simple proof of Lemma 24 is given below for completeness.

Proof of Lemma 24. We proceed by induction on m. The case of $m = 1$ is trivial. Consider $m > 1$. Since the P_i pairwise commute, $\text{Tr}(P_m \cdots P_1 \rho P_1 \cdots P_m) = \text{Tr}(P_m \cdots P_1 \rho) := \text{Tr}(P_m M \rho)$ for brevity, and M is a projector. Then,

$$\begin{aligned} 1 - \text{Tr}(P_m M\rho) = \quad & \text{Tr}((I - P_m)M\rho) + \text{Tr}(P_m(I - M)\rho) + \text{Tr}((I - P_m)(I - M)\rho) \\ = \quad & \text{Tr}((I - P_m)\rho) + \text{Tr}((I - M)\rho) - \text{Tr}((I - P_m)(I - M)\rho) \\ \leq \quad & \text{Tr}((I - P_m)\rho) + \text{Tr}((I - M)\rho), \end{aligned}$$

where the second equality holds since $\text{Tr}((I - P_m)(I - M)\rho)$ equals

$$\begin{aligned} &\text{Tr}((I - P_m)\rho) + \text{Tr}((I - M)\rho) - (\text{Tr}((I - P_m)M\rho) + \\ &\text{Tr}(P_m(I - M)\rho) + \text{Tr}((I - P_m)(I - M)\rho)). \end{aligned}$$

Applying the induction hypothesis completes the proof. ◀

F General simulations

In this section we will give a full proof of Lemma 15 and show that *any* efficient simulation will preserve hardness of ∀-APX-SIM, not just the special case considered in Definition 14. To state the full definition of simulation, we must first introduce the notion of an encoding.

▶ **Definition 25** ([13]). *We say a map $\mathcal{E} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}')$ is an encoding if it is of the form*

$$\mathcal{E}(M) = V(M \otimes P + \overline{M} \otimes Q)V^\dagger$$

where $\overline{M}$ denotes the complex conjugate of M, P and Q are orthogonal projectors (i.e. $PQ = 0$) on an ancilla space E; and V is an isometry $V : \mathcal{H} \otimes E \to \mathcal{H}'$.

When $\mathcal{H}$ is a many body system with a decomposition $\mathcal{H} = \bigotimes_{i=1}^n \mathcal{H}_i$, we say $\mathcal{E}$ is a local encoding if $E = \bigotimes_{i=1}^n E_i$ such that:

- $V = \bigotimes_{i=1}^{n} V_i$ *where each* V_i *acts on* $\mathcal{H}_i \otimes E_i$.
- *for each* i, *there exist orthogonal projectors* P_{E_i} *and* Q_{E_i} *on* E *which act non-trivially only on* E_i, *and satisfy* $PP_{E_i} = P$ *and* $QQ_{E_i} = Q$.

We are now ready to give the full definition of simulation.

▶ **Definition 26** ([13]). *We say that* H' *is a* (Δ, η, ϵ)*-simulation of* H *if there exists a local encoding* $\mathcal{E}(M) = V(M \otimes P + \overline{M} \otimes Q)V^\dagger$ *such that:*

1. *There exists an isometry* $\widetilde{V} : \mathcal{H} \otimes E \to \mathcal{H}'$ *such that* $\|\widetilde{V} - V\| \leq \eta$; *and that the encoding* $\widetilde{\mathcal{E}}(M) = \widetilde{V}(M \otimes P + \overline{M} \otimes Q)\widetilde{V}^\dagger$ *satisfies* $\widetilde{\mathcal{E}}(I) = P_{\leq \Delta(H')}$.
2. $\|H'_{\leq \Delta} - \widetilde{\mathcal{E}}(H)\| \leq \epsilon$.

We say that a family $\mathcal{F}'$ *of Hamiltonians can simulate a family* $\mathcal{F}$ *of Hamiltonians if, for any* $H \in \mathcal{F}$ *and any* $\eta, \epsilon > 0$ *and* $\Delta \geq \Delta_0$ *(for some* $\Delta_0 > 0$*), there exists* $H' \in \mathcal{F}'$ *such that* H' *is a* (Δ, η, ϵ)*-simulation of* H. *We say that the simulation is efficient if, in addition, for* H *acting on* n *qudits,* $\|H'\| = \mathrm{poly}(n, 1/\eta, 1/\epsilon, \Delta)$; H' *and* $\{ V_i \}$ *are efficiently computable given* H, Δ, η *and* ϵ; *and each local isometry* V_i *in the decomposition* $V = \bigotimes_i V_i$ *maps to* $O(1)$ *qudits.*

We note that Definition 14 is just the special case of Definition 26 where $\mathcal{E}(M) = VMV^\dagger$. We are now ready to restate and prove Lemma 15.

▶ **Lemma 15** (Simulations preserve hardness of $\forall$-APX-SIM). *Let* $\mathcal{F}$ *be a family of Hamiltonians which can be efficiently simulated by another family* $\mathcal{F}'$. *Then* $\mathcal{F}$-$\forall$-$APXSIM$ *reduces to* $\mathcal{F}'$-$\forall$-$APXSIM$.

Proof. For brevity, let $P_{\leq \Delta} := P_{\leq \Delta(H')}$. Let $\rho' = |\psi'\rangle\langle\psi'|$ be a state on $\mathcal{H}'$ such that $\langle\psi'| H' |\psi'\rangle \leq \delta'$ and let $\widetilde{\rho} = P_{\leq \Delta}\rho' P_{\leq \Delta} / \operatorname{Tr}(P_{\leq \Delta}\rho')$, so that by Lemma 19, we have $\|\rho' - \widetilde{\rho}\|_1 \leq 2\sqrt{\frac{\delta'}{\Delta - \lambda(H')}}$.

Since $P_{\leq \Delta}$ commutes with H', we have

$$\operatorname{Tr}(H'\rho') = \operatorname{Tr}(H' P_{\leq \Delta}\rho' P_{\leq \Delta}) + \operatorname{Tr}(H'(I - P_{\leq \Delta})\rho'(I - P_{\leq \Delta})) \tag{33}$$

$$= p \operatorname{Tr}(H'\widetilde{\rho}) + (1-p)\operatorname{Tr}(H'\widetilde{\rho}^\perp) \geq \operatorname{Tr}(H'\widetilde{\rho}), \tag{34}$$

where $p = \operatorname{Tr}(P_{\leq \Delta}\rho')$, $\widetilde{\rho}^\perp = (I - P_{\leq \Delta})\rho'(I - P_{\leq \Delta}) / \operatorname{Tr}((I - P_{\leq \Delta})\rho')$, and the final inequality follows because $\operatorname{Tr}(H'\widetilde{\rho}^\perp) \geq \Delta \geq \operatorname{Tr}(H'\widetilde{\rho})$.

Now let

$$\rho = \operatorname{Tr}_E\left(\widetilde{V}^\dagger \widetilde{\rho} \widetilde{V}(I \otimes P)\right) + \operatorname{Tr}_E\left(\overline{\widetilde{V}^\dagger \widetilde{\rho} \widetilde{V}(I \otimes Q)}\right)$$

and note that for any operator A on $\mathcal{H}$, we have $\operatorname{Tr}(\widetilde{\mathcal{E}}(A)\widetilde{\rho})$ equals

$$\operatorname{Tr}\left(\widetilde{V}(A \otimes P + \overline{A} \otimes Q)\widetilde{V}^\dagger \widetilde{\rho}\right) = \operatorname{Tr}\left(A \otimes P \widetilde{V}^\dagger \widetilde{\rho} \widetilde{V}\right) + \operatorname{Tr}\left(\overline{A} \otimes Q \widetilde{V}^\dagger \widetilde{\rho} \widetilde{V}\right) = \operatorname{Tr}(A\rho).$$

Therefore,

$$\begin{aligned}
\operatorname{Tr}(H\rho) &= \operatorname{Tr}(\widetilde{\mathcal{E}}(H)\widetilde{\rho}) \\
&\leq \operatorname{Tr}(H'\widetilde{\rho}) + \|H'_{\leq \Delta} - \widetilde{\mathcal{E}}(H)\| \\
&\leq \operatorname{Tr}(H'\rho') + \epsilon \\
&\leq \lambda(H') + \delta' + \epsilon \\
&\leq \lambda(H) + \delta' + 2\epsilon,
\end{aligned}$$

where the second inequality follows from Equation (34) and the last inequality from Lemma 27 of [13], which roughly states that eigenvalues are preserved up to additive error ϵ in a simulation.

At this point the proof diverges from the simpler case because ρ may be a mixed state, even when $\rho' = |\psi'\rangle\langle\psi'|$ is pure. Despite having a bound on $\mathrm{Tr}(H\rho)$, this bound may not hold for all pure states in the spectral decomposition of ρ. Let $\rho_\delta = P_\delta \rho P_\delta / \mathrm{Tr}(P_\delta)$, where P_δ is the projector onto eigenvectors of H with energy less than δ. By Lemma 19, $\|\rho - \rho_\delta\|_1 \leq 2\sqrt{\frac{\delta'+2\epsilon}{\delta}}$. We will use the spectral decomposition of $\rho_\delta = \sum_i \mu_i |\phi_i\rangle\langle\phi_i|$ where the $|\phi_i\rangle$ are orthogonal states with energy $\langle\phi_i| H |\phi_i\rangle \leq \lambda(H) + \delta$ and thus, for observable A given as part of of $\mathcal{F}$-$\forall$-APXSIM input,

$$\mathrm{Tr}(A\rho_\delta) = \sum_i \mu_i \langle\phi_i| A |\phi_i\rangle \quad \begin{cases} \leq a \text{ in a YES instance} \\ \geq b \text{ in a NO instance.} \end{cases}$$

Let $U = V\widetilde{V}^\dagger$, which satisfies $U\widetilde{\mathcal{E}}(A) = \mathcal{E}(A)U$ for any A, and so $\mathcal{E}(I)U\widetilde{\rho}U^\dagger = U\widetilde{\mathcal{E}}(I)\widetilde{\rho}U^\dagger = U\widetilde{\rho}U^\dagger$. Now we need to choose A' such that $A'\mathcal{E}(I) = \mathcal{E}(A)$. (Two notes: First, $\mathcal{E}(I) \neq I$ necessarily, as P and Q need not sum to identity. Second, setting $A' = \mathcal{E}(A)$ is not necessarily desirable, as P and Q may be non-local projectors.) For example if $A = B_i \otimes I$, let $A' = V_i(B_i \otimes P_{E_i} + \overline{B_i} \otimes Q_{E_i})V_i^\dagger \otimes I$. We note that the locality of A' depends on the number of qudits which V_i maps to, which is $O(1)$ by the definition of efficient simulation. Then

$$\mathrm{Tr}(A\rho) = \mathrm{Tr}\left(\widetilde{\mathcal{E}}(A)\widetilde{\rho}\right) = \mathrm{Tr}\left(\mathcal{E}(A)U\widetilde{\rho}U^\dagger\right) = \mathrm{Tr}(A'\mathcal{E}(I)U\widetilde{\rho}U^\dagger) = \mathrm{Tr}(A'U\widetilde{\rho}U^\dagger)$$

and therefore

$$\begin{aligned} |\mathrm{Tr}(A'\rho') - \mathrm{Tr}(A\rho_\delta)| &\leq |\mathrm{Tr}(A'\rho') - \mathrm{Tr}(A'U\widetilde{\rho}U^\dagger)| + |\mathrm{Tr}(A\rho) - \mathrm{Tr}(A\rho_\delta)| \\ &\leq \|A'\|\left(\|\rho' - \widetilde{\rho}\|_1 + \|\widetilde{\rho} - U\widetilde{\rho}U^\dagger\|_1\right) + \|A\|\|\rho - \rho_\delta\|_1 \\ &\leq \|A\|\left(2\sqrt{\frac{\delta'}{\Delta - \lambda(H')}} + 2\eta + 2\sqrt{\frac{\delta'+2\epsilon}{\delta}}\right), \end{aligned}$$

We note that $\|\widetilde{\rho} - U\widetilde{\rho}U^\dagger\|_1 \leq 2\eta$ follows from $\|U - \widetilde{V}\widetilde{V}^\dagger\| \leq \eta$, and that $\widetilde{V}\widetilde{V}^\dagger\widetilde{\rho} = P_{\leq\Delta}\widetilde{\rho} = \widetilde{\rho}$. Therefore we just need to choose $\Delta, \epsilon, \eta, \delta'$ such that this is less than $(b-a)/3$ and then set $a' = a + (b-a)/3$ and $b' = b - (b-a)/3$. ◀