Quantum PCPs: on Adaptivity, Multiple Provers and Reductions to Local Hamiltonians

Harry Buhrman[1], Jonas Helsen[2], and Jordi Weggemans[2]

[1]QuSoft, University of Amsterdam & Quantinuum, Partnership House, Carlisle Place, London
[2]QuSoft & CWI, Amsterdam, the Netherlands

We define a general formulation of quantum PCPs, which captures adaptivity and multiple unentangled provers, and give a detailed construction of the quantum reduction to a local Hamiltonian with a constant promise gap. This reduction turns out to be a versatile subroutine to prove properties of quantum PCPs, allowing us to show:

(i) Non-adaptive quantum PCPs can simulate adaptive quantum PCPs when the number of proof queries is constant. In fact, this can even be shown to hold when the non-adaptive quantum PCP picks the proof indices *uniformly* at random from a subset of all possible index combinations, answering an open question by Aharonov, Arad, Landau and Vazirani (STOC '09).

(ii) If the q-local Hamiltonian problem with constant promise gap can be solved in QCMA, then $\mathsf{QPCP}[q] \subseteq \mathsf{QCMA}$ for any $q \in \mathcal{O}(1)$.

(iii) If $\mathsf{QMA}[k]$ has a quantum PCP for any $2 \leq k \leq \mathrm{poly}(n)$, then $\mathsf{QMA}[2] = \mathsf{QMA}$, connecting two of the longest-standing open problems in quantum complexity theory.

Moreover, we also show that there exist (quantum) oracles relative to which certain quantum PCP statements are false. Hence, any attempt to prove the quantum PCP conjecture requires, just as was the case for the classical PCP theorem, (quantumly) non-relativizing techniques.

Jordi Weggemans: jrw@cwi.nl, https://jordiweggemans.github.io

Contents

1 Introduction

Arguably the most fundamental result in classical complexity theory is the Cook-Levin Theorem [1, 2], which states that constraint satisfaction problems (CSPs) are NP-complete. Kitaev showed that there exists a natural analogue in the quantum world: local Hamiltonian problems, which can be viewed as a quantum generalisation of classical constraint satisfaction problems, are complete for the class QMA, which is a quantum generalisation of NP [3].[1]

There are many variants of NP, which a priori seem either more powerful or more restrictive, but turn out to be equal to NP. Examples of these are interactive proof systems of various flavours with a deterministic verifier, and perhaps most surprisingly and importantly, probabilistically checkable proof systems (PCPs) [5]. A PCP has a polynomial-time probabilistic verifier with query access to a proof (provided by a prover). Usually, a PCP is specified by two parameters: the number of random coins the verifier is allowed to use (which also upper bounds the length of the proof as some exponential in the number of coins) and the number of queries it can make to the proof. The PCP theorem [5, 6], which originated from a long line of research on the complexity of interactive proof systems, states that all problems in NP can be decided, with a constant probability of error, by only using a logarithmic number of coin flips and by querying a constant number of bits of the proof. Equivalently, it implies that it is NP-hard to decide whether an instance of a CSP is either completely satisfiable or no more than a constant fraction of its constraints can be satisfied. This is usually referred to as the *hardness of approximation* formulation of the PCP theorem. It is also possible to prove the PCP theorem by reducing a CSP *directly* to another CSP which has the above property. This transformation, due to Dinur [7], is usually referred to as *gap amplification*, referring to the increase in the difference (the gap) in the fraction of constraints which can be satisfied in both the YES- and NO-instances.

Naturally, quantum complexity theorists have proposed proof-checking and hardness of approximation versions of a quantum PCP conjecture. The hardness of approximation formulation states that the local Hamiltonian problem with constant promise gap, relative to the operator norm of the Hamiltonian, is QMA-hard. This formulation of the quantum PCP conjecture has been the predominant focus of the quantum PCP literature, and progress has been made in giving evidence both in favour and against the conjecture. Amongst the positive are the NLTS theorem and its cousins, excluding a large set of potential NP-witnesses [8, 9, 10, 11, 12]. Evidence against (assuming $\mathsf{NP} \neq \mathsf{QMA}$) are results showing that under many extra constraints one poses on the local Hamiltonian problem (constraining e.g. the interaction graph or ground space structure) the problem is classically solvable when the relative promise gap is constant, whilst the problem remains (quantumly) hard when the gap is inverse polynomial [13, 14, 15, 16, 17, 18, 19].

The proof-checking formulation of the quantum PCP conjecture, which states that one can solve any promise problem in QMA by using a quantum verifier which only accesses a constant number of qubits from a quantum proof, has received considerably less attention. The reason for this is that it is not hard to show that both conjectures are equivalent under *quantum* reductions. This was already pointed out in the first work which proposed a quantum PCP formulation [20].[2] Perhaps that is why, even after more than two decades

[1] There are many possible quantum generalisations of NP, see [4] for a recent review.

[2] Whilst this is widely known in the community, this reduction has (as far as the authors are aware) never been written down in full detail except in the works of [21] and [12], which both consider quantum PCPs that are not (or at least shown to be) fully general. In both of these works, a quantum PCP is defined using a single POVM, where each POVM element acts on an ancilla register and a small part of the

since the question of the existence of a quantum PCP was first posited [22], many basic questions regarding the proof-checking formulation have not been addressed. For example, as already raised in [20], how robust are definitions of quantum PCPs under subtle changes, e.g. the choice of picking the distribution over which the proof qubits are selected? Are adaptive queries to the proof more powerful than non-adaptive in the constant query setting or are they the same as is the case for the classical setting? Are there non-equivalent variations of quantum PCPs, similar to the many natural variations of QMA?

Motivated by these questions, this work is concerned with studying the properties of quantum PCPs through proof verification. Studying the connection between quantum PCPs and local Hamiltonians turns out to be a powerful tool for showing basic properties of quantum PCPs.

1.1 Quantum PCPs are constant promise gap local Hamiltonians

We give a detailed definition of a quantum PCP verifier, which allows for adaptive queries as well as quantum proofs consisting of multiple unentangled quantum states given by different unentangled provers. Informally (see Definition 4 for the full definition) the verifier makes q queries to a quantum proof in the following way:

- The verifier takes as an input a string x of size n (which will be hardcoded into the circuits), has a polynomial-sized ancilla register (which we will call the workspace) and a proof register containing k unentangled quantum proofs of some specified size depending on n. Generally, $k = 1$ unless specified otherwise.
- The verifier is going to do the following q times. At stage $t \in [q]$, suppose the quantum PCP has decided to act on proof qubits from the set $I = \{i_1, \ldots, i_{t-1}\}$. The quantum PCP then applies a polynomial-time quantum circuit to the workspace and the proof qubits with indices $i_1, \ldots, i_{t-1}$ followed by a measurement of some designated qubits to determine which next index i_t is going to be added to the set I.
- Finally, another polynomial-time circuit is applied to the workspace and the proof qubits with indices $i_1, \ldots, i_q$ and a single designated output qubit is measured in the computational basis. The quantum PCP accepts if and only if the outcome is $|1\rangle$.

Note that this quantum PCP is adaptive; similarly, a non-adaptive quantum PCP can be defined using only a single measurement to decide the indices $i_1, \ldots, i_q$. One can then construct complexity classes based on this verifier in a general setting, where we also include the possibility of multiple provers.

Our first goal is to show that there exists a local Hamiltonian with PSD terms whose expectation value has a one-to-one correspondence with the acceptance probabilities of the quantum PCP given a quantum proof. This involves some subtleties compared to the classical case, as each intermediate measurement to determine the next index can alter the *entire* quantum state of the quantum verifier due to entanglement across different registers. Nevertheless, we prove the following result.

Lemma 1.1 (Informal) (from Lemma 5)**.** *Let x be an input and V_x a q-query quantum PCP verifier with x hardcoded into it. Then there exists a q-local Hamiltonian H_x such that*

$$\Pr[V_x \text{ accepts } \xi] = 1 - \mathrm{tr}[H_x \xi],$$

provided quantum proof. The verifier selects the POVM element by sampling from a classical probability distribution (in [21] this is simply the uniform distribution).

where $H_x = \sum_{i\in[m]} H_{x,i}$ *with each* $H_{x,i}$ *PSD and* $\|H_x\| \leq 1$.

The core idea is that instead of focusing on conditional probabilities arising from specific query paths, we consider the overall probability of following a particular adaptive query path *and* accepting. For a fixed input, we show that all input, ancilla, and circuit information can be directly encoded into a single POVM element whose expectation value captures this probability. Then, by linearity of expectation, we simply add all POVM elements together to form our final Hamiltonian term.

However, the quantum PCP conjecture is typically framed differently from the form of the Hamiltonian in Lemma 1.1. First, it is often assumed that the interaction graph corresponding to the qubits on which the local terms act has constant degree.[3] However, this so-called *interaction degree* depends on the support of the distribution over which the indices are selected, which could, in principle, include all possible choices of indices. As a result, the interaction graph can have degree $\Omega(n)$. Second, one assumes that H is of the form $H = \sum_{i\in[m]} H_i$ with $0 \preceq H_i \preceq 1$ and that $\lambda_{\min}(H) \leq \alpha m$ or $\lambda_{\min}(H) \geq \beta m$. These Hamiltonians can be easily rewritten in the same form as H_x by letting $H' = \sum_{i\in[m]} H'_i$ with $H'_i := H_i/m$, such that $\lambda_{\min}(H') \leq \alpha$ or $\lambda_{\min}(H') \geq \beta$. However, this formulation implicitly assumes that in the latter case, at least a constant fraction of the terms contribute on the order of $\sim \frac{1}{m}$ to the ground state energy, meaning that the energy is relatively spread out over all terms. If the Hamiltonian's interaction degree were in fact constant, then such energy spreading would indeed hold in the large energy case. However, for our definition of quantum PCPs, this does not necessarily have to be the case. This difference is subtle but important: one can no longer pick a constant number of terms uniformly at random and measure their energy to solve the corresponding local Hamiltonian problem, as only a small $o(1)$ fraction of the terms might have large energy. Nevertheless, we show that there exists a transformation from any such Hamiltonian to another one which does satisfy this property, at the cost of increasing the locality by a factor of two and decreasing the promise gap exponentially in the locality, which is still constant if the locality is constant (Lemma 8).

Now we have shown that the probability that a QPCP-verifier accepts a certain proof is equivalent to the expectation value of some Hamiltonian, we still need to show that the descriptions of the local terms can be obtained efficiently. For this, it is well-known that one can adopt a *quantum reduction* [20]. We show that a quantum reduction also exists for our general formulation of quantum PCPs. Moreover, by a simple trick, we find that we can round the Hamiltonian in such a way that every time the reduction succeeds (which can be done with a success probability exponentially close to one) it produces *exactly* the same Hamiltonian.

Theorem 1.1 (Informal) (from Theorem 1 and Corollary 1)**.** *Let* x *be an input. There exists a quantum reduction from a* q*-query quantum PCP with circuit* V_x *to a* q*-local Hamiltonian* $\hat{H}_x$ *such that*

$$\left|\Pr[V_x \text{ accepts } \xi] - \left(1 - \operatorname{tr}\left[\hat{H}_x \xi\right]\right)\right| \leq \epsilon,$$

which succeeds with probability $1-\delta$ *and runs in time* $\operatorname{poly}(n, 1/\epsilon, \log(1/\delta))$*. Moreover, this reduction can be made such that it produces the same Hamiltonian every time it succeeds.*

[3]The interaction degree here is defined as the maximum number of terms that act simultaneously on a single qubit.

The specificity of the reduction (i.e. it always yields the same Hamiltonian and the acceptance probability is directly encoded in the energy of the Hamiltonian) makes it a useful tool in deriving a variety of statements about quantum PCPs, since quantum PCPs themselves can perform the reduction.

1.2 Results from applying the reduction

All applications are derived by adopting variations of the following general strategy:

1. The reduction of a quantum PCP to a local Hamiltonian is used as a pre-processing step of some larger protocol, rounding to a fixed Hamiltonian with high success probability.

2. One uses tricks of Hamiltonian complexity to manipulate and solve the corresponding local Hamiltonian problems.

It is important to emphasise that by this strategy, the reductions we propose are simply *classical* polynomial-time reductions, even though they involve quantum reductions. The reason for this is that the quantum reductions are absorbed into the quantum protocols as pre-processing steps. A disadvantage of this strategy is that it in general fails to preserve perfect completeness.

1.2.1 Reduction to average particle energy formulation

As a first application, we consider a specific formulation of the quantum PCP conjecture in terms of an error constant relative to the number of sites (i.e., qubits or qudits) rather than the total number of terms. This formulation has appeared in several works (see, for example, the recent preprints [23] and [24]), but to our knowledge, it has never been shown to be complete for the class $\mathsf{QPCP}[\mathcal{O}(1)]$. Using our reductions above, combined with standard techniques from tail bounds for sums of random matrices, we can show that even under this restriction, the local Hamiltonian with a constant promise gap remains complete for $\mathsf{QPCP}[\mathcal{O}(1)]$.

Theorem 1.2 (Informal) (from Corollary 3)**.** *There are constants $c < d$ such that deciding if the ground energy of an n-qubit local Hamiltonian H with $m = \Theta(n)$ terms is (yes case) $\leq c$ or (no case) $\geq d$ is $\mathsf{QPCP}[\mathcal{O}(1)]$-complete under quantum polynomial-time reductions.*

1.2.2 Upper bounds on quantum PCPs

By a result of [19], it seems unlikely that there exists a *classical* reduction that satisfies the same properties as Theorem 1.1, as that would show that $\mathsf{BQP} \subseteq \mathsf{QCPCP} \subseteq \mathsf{NP}$.[4] The other direction (going from a local Hamiltonian with constant to a QPCP verification circuit) is easy to show through Kitaev's original QMA verification circuit for the local Hamiltonian problem (see Protocol 1), once the Hamiltonian has been transformed in the required form (equally bounded operator norm on all local terms). Hence, if one shows that $\mathsf{QPCP}[q] \subseteq \mathcal{C}$ for some complexity class $\mathcal{C}$ that can perform polynomial-time classical reductions (so any class $\mathcal{C} \supseteq \mathsf{P}$ would suffice), then the local Hamiltonian problem with constant promise gap is in $\mathcal{C}$ as well. However, since the reduction from QPCP to a local

[4]Here QCPCP is just as QPCP but with the promises for classical proofs instead of quantum proofs, similar to what QCMA is to QMA.

Hamiltonian is a quantum reduction, such a statement does not hold in the other direction. Weaker conclusions can be drawn however. In particular we can show that an upper bound on the local Hamiltonian problem with a constant promise gap implies an upper bound on quantum PCPs with the same locality.

Theorem 1.3 (Informal) (from Theorem 2)**.** *If the q-local Hamiltonian problem with constant promise gap is in* QCMA*, then* $\mathsf{QPCP}[q] \subseteq \mathsf{QCMA}$*.*

This entails that, assuming $\mathsf{QMA} \neq \mathsf{QCMA}$, disproving QMA-hardness for either the proof verification or Hamiltonian formulation of the quantum PCP conjecture (by placing it in QCMA) does indeed disprove the other (similar to how proving QMA hardness of one implies QMA-hardness of the other.)

1.2.3 Adaptive and non-adaptive quantum PCPs

It is a well-known fact that non-adaptive PCPs can simulate adaptive PCPs in the classical setting when the number of proof queries is only allowed to be constant. This can easily be shown by simulating the adaptive PCP verifier over all possible local proof settings for a fixed setting of the randomness, after which the actual proof can be read at the used locations to check which setting was correct. However, such tricks seem to have no quantum analogue, as it is generally impossible to fix the inherent randomness (or "quantumness") in a quantum algorithm [25] (see also the discussion in Section 1.1). However, we can show that a quantum analogue of this result still holds.

Theorem 1.4 (Informal) (from Theorem 3)**.** *Non-adaptive quantum PCPs can simulate adaptive quantum PCPs when the number of proof queries is constant.*

In fact, we are able to prove something stronger: we show that any constant query PCP can be simulated by a quantum PCP that picks the indices of the proof bits uniformly at random from a subset of all possible index choices, resolves an open question of [20] (see Remark 3 in the main text).[5] We prove these results by showing that there exists a non-adaptive quantum PCP verifier for the local Hamiltonian induced by the adaptive quantum PCP, which only has its promise gap exponentially smaller in the locality. Weak error reduction can then be used to boost the promise gap back to its original value.

1.2.4 Quantum PCPs for $\mathsf{QMA}[k]$

Finally, we consider a quantum PCP in the $\mathsf{QMA}[k]$ setting: the complexity class $\mathsf{QMA}[k]$ is a generalisation of QMA where there are multiple uninteracting provers, which are guaranteed to be unentangled with each other. It is known that $\mathsf{QMA}[k] = \mathsf{QMA}[2]$ by the result of Harrow and Montanaro [26], but is generally believed that $\mathsf{QMA} \neq \mathsf{QMA}[2]$ since there are problems known to be in $\mathsf{QMA}[2]$ but not known to be in QMA [27, 28, 29, 30]. Moreover, the best upper bound we have on $\mathsf{QMA}[2]$ is that it is contained in NEXP, which follows by just guessing exponential-size classical descriptions of the two quantum proofs. The reason for this is that the maximum acceptance probability of a $\mathsf{QMA}[2]$ verifier is with respect to separable states, which means that the corresponding maximisation problem is no longer convex. Another surprising fact is that the separable local Hamiltonian problem is QMA-complete, and we only have that a separable Hamiltonian problem becomes $\mathsf{QMA}[2]$-complete when one considers *sparse* Hamiltonians, where the individual terms can

[5] This also shows that the quantum PCP formulation as in [21] is in fact fully general when one considers quantum PCPs which make a constant number of proof queries.

only have a polynomial number of non-zero entries but do not have to be local [31]. Since our quantum reduction from a quantum PCP always produces a local Hamiltonian irrespective of the number of provers, a generalisation of the protocol of [31] to the k-separable local Hamiltonian setting allows us to show the following.

Theorem 1.5 (Informal) (from Theorem 5)**.** *If* $\mathsf{QMA}[k]$ *has a* k*-prover quantum PCP for any* $2 \leq k \leq \mathrm{poly}(n)$*, then we have that* $\mathsf{QMA}[2] = \mathsf{QMA}$*.*

This shows a connection between two of the biggest open problems in quantum complexity. In particular, if one believes that a quantum PCP conjecture for QMA should hold and that $\mathsf{QMA}[2] \neq \mathsf{QMA}$, then any proof of the quantum PCP conjecture should not translate to the multiple unentangled prover setting.

As a bonus, we improve the parameter range of QMA-containment of the consistency of local density matrices (CLDM) problem ([32, 33]), by giving a new protocol.

1.3 Proofs of the quantum PCP theorem do not relativize

In the final section, we will show that given Aaronson and Kuperberg's quantum oracle [34] (under which $\mathsf{QCMA} \neq \mathsf{QMA}$) we have that the quantum PCP conjecture is false. Their oracle separation is fundamentally based on a lower bound combined with a counting argument, which exploits the fact that there are doubly exponential quantum states (with small pairwise fidelities) but only an exponential number of classical proofs. The reason this lower bound cannot be applied directly to QPCP comes from the fact that the total number of quantum proofs that can be given is the same as for QMA. However, from the verifiers perspective, the total number of proofs he observes is limited by a set of local density matrices and indices indicating from what part of the global state this density matrix comes from. Using this observation, we can define an artificial variation of QPCP, denoted QPCP_ϵ, which can be viewed as a "proof-discretized"-version of QPCP. Informally, this class can be viewed as a variant to QPCP where after the verifier decides which parts of the proof are going to be accessed, some "magical operation" takes this part of the proof and projects it onto the density matrix closest to it from some fixed finite set. We can show that with respect to all oracles, this new class contains QPCP, so any oracle separation between QMA and QPCP_ϵ also separates QPCP and QMA.

Theorem 1.6 (Informal) (from Theorem 6 and Corollary 4)**.** *There exists a quantum (classical) oracle relative to which the polylog (constant) quantum PCP conjecture is false.*

Since for any oracle X encoding a PSPACE-complete language we have $\mathsf{QPCP}[q]^X = \mathsf{QMA}^X$, and since the inclusions $\mathsf{QPCP}[q] \subseteq \mathsf{QMA} \subseteq \mathsf{PSPACE}$ all relativize with respect to classical oracles and $\mathsf{PSPACE}^{\mathsf{PSPACE}} = \mathsf{PSPACE}$, it follows that this result implies proving the quantum PCP conjecture by showing $\mathsf{QPCP}[q] = \mathsf{QMA}$ for some constant $q \in \mathcal{O}(1)$ requires non-relativizing techniques—just as was the case for the classical PCP theorem [35].

Perhaps the most important takeaway from this result is that it should apply to any quantum oracle separation involving QMA which exploits the fact that there are doubly exponentially many proofs in QMA. Therefore, all oracle separations between QMA and QCMA based on such an argument immediately yield an oracle relative to which the quantum PCP conjecture is false.

1.4 Discussion and open problems

We have studied the quantum PCP conjecture in the proof-checking formulation (as opposed to the more popular local Hamiltonian formulation). Ironically, many of our results

follow from leveraging the reduction to the local Hamiltonian problem. Our results take the form of both novel statements, and formal proofs of what we think of as "folklore knowledge", i.e., results we suspect are known to be true but for which we could not find a proof in the literature. Given the renewed interest in the quantum PCP conjecture we believe this was an effort worth undertaking. We conclude by listing several questions on quantum PCPs which we have not yet resolved but believe are interesting avenues for future work.

Locality reductions We do not know whether Theorem 2 already holds when only the 2-local Hamiltonian problem is in QCMA. In [14], it is claimed that the 2-local Hamiltonian problem is complete for QPCP (which would imply the above statement), but it is not clear to us that this is actually the case. The reason for this is that the gadget constructions of [36] mentioned in [14], which transform a q-local Hamiltonian into a 2-local Hamiltonian whilst preserving the relative promise gap, only work when every qubit is involved in only a *constant* number of terms. In fact, the existence of a transformation which maps a q-local Hamiltonian with an interaction degree $\Omega(n)$ to a 2-local Hamiltonian with an interaction degree $\Omega(n)$ would directly imply that the local Hamiltonian problem with constant promise gap is in NP, disproving the quantum PCP conjecture for any constant q directly. The reason for this is as follows[6]: take any q-local Hamiltonian H, $q \geq 2$ constant, with some degree d interaction graph where $d \geq 2$, $\|H\| = 1$ (which can be assumed w.l.o.g.), and completeness and soundness parameters b and a, respectively, satisfying $(b-a)/\|H\| = \Omega(1)$. Then we have the Hamiltonian $H' = H^2$ is $2q$-local, has operator norm still equal to 1, has an interaction graph of degree n, and completeness and soundness parameters b', a' with $b' - a'/\|H'\| \geq (b-a)^2 = \Omega(1)$. If there would exist a transformation from H' to H'' with H'' 2-local, interaction degree $\mathcal{O}(n)$ and with completeness a'' and soundness b'' satisfying $(b'' - a'')/\|H''\| = \Omega(1)$, then by Brandao-Harrow [14, Cor. 5] one could decide the correct energy of H.

Quantum PCPs with perfect completeness A major downside of our technique—leveraging the QPCP to local Hamiltonian quantum reduction—is that it fails to preserve perfect completeness, which is due to the Hamiltonian being learned only up to some small error. It is an open question whether other techniques can be used to obtain similar results in the perfect completeness setting, for example whether non-adaptive quantum PCPs can simulate adaptive ones when the number of quantum proof queries is only allowed to be constant.

Strong error reduction It is easy to show that non-adaptive quantum PCPs with near-perfect completeness allow, just like QMA [37], for strong error reduction (Appendix A). However, we do not know whether this also holds in the general case. The idea of taking polynomials of the Hamiltonian to manipulate its spectrum seems not to be compatible with Kitaev's energy estimation protocol, as this in general introduces coefficients that can be negative and have large absolute values, which is incompatible with Kitaev's energy estimation protocol (Section 4.3) which requires each term to have to be PSD and have operator norm at most one.

[6]This argument is based on a similar—unpublished—argument due to Anshu and Nirkhe, which can be found at (at time of writing) https://anuraganshu.seas.harvard.edu/files/anshu/files/bh_4local.pdf

Acknowledgements JW was supported by the Dutch Ministry of Economic Affairs and Climate Policy (EZK), as part of the Quantum Delta NL programme. JH acknowledges support from the Quantum Software Consortium (NWO Zwaartekracht 024.003.037) and a Veni grant (NWO Veni 222.331).

2 Preliminaries

2.1 Notation

For a Hilbert space $\mathcal{H}$ of dimension d, we denote $\mathcal{D}(\mathcal{H}) = \{\rho \in \mathrm{PSD}(\mathcal{H}), \mathrm{tr}[\rho] = 1\}$, where $\mathrm{PSD}(\mathcal{H})$ is the set of all positive semidefinite matrices in $\mathcal{H}$, for the set of all density matrices in $\mathcal{H}$. For the set of all d-dimensional pure states, we write $\mathcal{P}(\mathcal{H}) = \{\rho \in \mathrm{PSD}(\mathcal{H}), \mathrm{tr}[\rho^2] = 1\}$ (normally defined as $\mathbb{CP}^{d-1}$, i.e. complex projective space with dimension $d-1$). For a tuple consisting of q distinct elements, i.e., $(i_1, i_2, \ldots, i_q)$, we write $\{(i_1, i_2, \ldots, i_q)\}!$ for the set of all permutations of the tuple $(i_1, i_2, \ldots, i_q)$. We denote $\binom{[n]}{k}$ for the set of all k-element subsets of $[n]$ (so unordered and without repetitions). When the base is not explicitly stated, log always refers to the base-2 logarithm.

2.2 Complexity theory

All verifiers used to define quantum complexity classes in this work will be defined in terms of the quantum circuit model. We say that a quantum verifier V_n, taking an input $x \in 0, 1^n$ for some integer n, is polynomial-time if it uses a workspace of $\mathrm{poly}(n)$ ancilla qubits initialized in $|0\rangle$ (tensored with the input in $|x\rangle$) and at most $\mathrm{poly}(n)$ elementary quantum gates and elementary (intermediate) POVMs (or PVMs). To efficiently generate descriptions of its components depending only on the input size n, we need the notion of P-*uniformity*.

Definition 1. *A set of verifiers $\{V_n : n \in \mathbb{N}\}$ is polynomial-time uniform (abbreviated as P-uniform) if there exists a polynomial-time deterministic Turing machine that, on input 1^n, outputs a classical description of V_n.*

We start by recalling the basic definitions from complexity theory classes used in this work.

Definition 2 ($\mathsf{QMA}[k]$)**.** *Let $n \in \mathbb{N}$ be the input size. A promise problem $A = (A_{yes}, A_{no})$ is in $\mathsf{QMA}[k, c, s]$ if and only if there exists a P-uniform family of polynomial-time quantum verifiers $\{V_n\}$ and a polynomial p, where V_n takes as input a string $x \in \{0,1\}^n$ and a $kp(n)$-qubit witness quantum state $|\psi\rangle$ and decides on acceptance or rejection of x such that*

- *if $x \in A_{\mathrm{YES}}$ then there exist $p(n)$-qubit witness states $|\psi_j\rangle \in (\mathbb{C}^2)^{\otimes p(n)}$, $j \in [k]$, such that V_n accepts $(x, \otimes_{j\in[k]} |\psi_j\rangle)$ with probability $\geq c$,*
- *if $x \in A_{\mathrm{NO}}$ then for all $|\psi_j\rangle \in (\mathbb{C}^2)^{\otimes p(n)}$, $j \in [k]$, we have that V_n accepts $(x, \otimes_{j\in[k]} |\psi_j\rangle)$ with probability $\leq s$,*

where $c - s = \Omega(1/\mathrm{poly}(n))$. If $c = 2/3$ and $s = 1/3$, we abbreviate to $\mathsf{QMA}[k]$, and if $k = 1$, we have $\mathsf{QMA} = \mathsf{QMA}[1]$.[7]

[7] In the literature, $\mathsf{QMA}(k)$ is commonly used, but we chose to use "[.]"-brackets, as this is customary in complexity theory for classes that depend on some parameters.

Definition 2a (QCMA)**.** *This has the same definition as* QMA *but where the promises hold with respect to computational basis states* $|y\rangle$*, where* $|y| = p(n)$*. In this case we trivially have* $\mathsf{QCMA}[k] = \mathsf{QCMA}$ *for all* $k \leq \text{poly}(n)$*.*

By a result of Harrow and Montanaro, it is known that QMA with 2 unentangled provers can simulate any polynomial number of unentangled provers.

Lemma 1 ([26])**.** *For any* $2 \leq k \leq \text{poly}(n)$*, we have*

$$\mathsf{QMA}[k] = \mathsf{QMA}[2].$$

Fact 1. *For any* $1 \leq k \leq \text{poly}(n)$*, the completeness and soundness parameters in* $\mathsf{QMA}[k]$ *and* QCMA *can be made exponentially close to* 1 *and* 0*, respectively, i.e.* $c = 1 - 2^{-\mathcal{O}(n)}$ *and* $s = 2^{-\mathcal{O}(n)}$*.*

The q-local k-separable local Hamiltonian problem generalises the q-local Hamiltonian problem (which is the case for $k = 1$) by giving a promise on the energies of all low-energy of k-fold tensor products, where each quantum state in the tensor product lives on a pre-determined number of qubits. It is shown to be QMA-complete in the case where $k = 2$ in [31].

We will define local Hamiltonian problems such that all local terms are PSD, as this will turn out to be the relevant setting in the context of quantum PCPs. Generally, one can apply the transformation $H_i \mapsto (H_i + \mathbb{I})/2$ for each local term, which makes every local term PSD.

Definition 3 (k-separable q-local Hamiltonian problem)**.** (k, q)-$\mathsf{LH}[\delta]$
Input: *A classical description of a collection of* q*-local PSD operators* $\{H_i\}_{i\in[m]}$*, which define an* n*-qubit* k*-local Hamiltonian* $H = \sum_{i=1}^{m} H_i$ *with* $\|H\| \leq 1$*, and two parameters* $a, b \in \mathbb{R}$ *such that* $b - a = \delta$*. We call* δ *the promise gap.*
Promise: *We have that either one of the following two holds:*

(i) There exists a quantum state $\xi = \otimes_{j\in[k]}\xi_j$*, such that* $\text{tr}[H\xi] \leq a$

(ii) For all quantum states $\xi = \otimes_{j\in[k]}\xi_j$ *we have that* $\text{tr}[H\xi] \geq b$*.*

Output:

- *If case (i) holds, output* YES*.*
- *If case (ii) holds,, output* NO*.*

If $k = 1$*, the above problem reduces to the standard local Hamiltonian problem which we abbreviate as* q-$\mathsf{LH}[\delta]$*.*

3 Quantum probabilistically checkable proofs

The bulk of this section is taken up a definition of a quantum PCP. Our definition is more detailed than that given in [38], and more general than the definitions in [12, 21], in that we also include the ability to query proof qubits adaptively. Moreover, we allow for the possibility of having multiple provers (i.e. unentangled proofs). Once we have this definition in place we can define an associated complexity class QPCP. For convenience, we also explicitly define a non-adaptive version. We finish up by arguing that weak error reduction (and limited strong error reduction) is possible in QPCP.

Definition 4 (quantum PCP verifier)**.** *Let $p_1, p_2, p_3 : \mathbb{N} \to \mathbb{N}$, and $n \in \mathbb{N}$ be the input size. A (k, q, p_1, p_2, p_3)-QPCP-verifier V consists of the following:*

- *a n-qubit input register A, initialised in input $|x\rangle$, $x \in \{0,1\}^n$;*
- *a p_1-qubit ancilla register B, initialised in $|0\rangle^{\otimes p_1(n)}$;*
- *a kp_2-qubit proof register C, initialised in $\xi = \otimes_{j=1}^{k} \xi_j$ for some quantum witnesses $\xi_j \in \mathcal{D}\left((\mathbb{C}^2)^{\otimes p_2(n)}\right)$ for all $j \in [k]$;*
- *a collection of PVMs $\Pi^t = \{\Pi_i^t\}$, $i \in [kp_2(n)]$, with $\Pi_i^t = |i\rangle\langle i| \otimes \mathbb{I}$ for all $t \in [q]$, $\Pi^{\text{output}} = \{\Pi_0^{\text{output}}, \Pi_1^{\text{output}}\}$, with $\Pi_0^{\text{output}} = |0\rangle\langle 0| \otimes \mathbb{I}$ and $\Pi_1^{\text{output}} = |1\rangle\langle 1| \otimes \mathbb{I}$;*
- *collection of circuits $V^{t'}$, $t' \in [q+1]$, where circuit $V^{t'}$ only acts on at most t' qubits of the proof register and consist of at most p_3 gates from some universal gate set.*

Let $I = \emptyset$ be the set of all proof indices to be accessed. The quantum PCP verifier V acts as follows:

1. *For $t \in [q]$: it applies the circuit V^t to registers A, B and qubits I from C, performs the measurement Π^t and adds outcome i_t to the set I;*
2. *It applies V^{q+1} followed by a measurement of Π^{output} of the first qubit and returns "accept" if the outcome was $|1\rangle$, and "reject" if the outcome was $|0\rangle$.*

If p_1, p_2, and p_3 are all polynomially bounded functions, then we abbreviate to a (k, q)-QPCP verifier, and to a (q)-QPCP verifier if additionally $k = 1$. For the remainder of this work, we will assume that in Step 1 for any t the probability of measuring any outcome i_t for which there exists a $i_{t'} \in I$, $t' \in [t-1]$ such that $i_t = i_{t'}$, is zero.

We make the final assumption in Definition 4 because it simplifies our notation, and there would be no benefit in querying the same proof index multiple times. This assumption can easily be enforced in a QPCP-verifier by adding a random sampling step whenever a duplicate index is observed.

Remark 1. *We will often write V_x to denote the verifier with the input x hardcoded. Moreover, in the case of multiple provers, an index i_t will denote a tuple (j, k), where j indicates the corresponding proof and k indicates the index of the qubit in this proof.*

Definition 4a (Non-adaptive quantum PCP verifier)**.** *A non-adaptive (k, q)-QPCP_{NA}-verifier is just like the (k, q)-QPCP verifier but instead has a single PVM $\Pi = \{\Pi_{i_1,\ldots,i_q}\}$ with $\Pi_{i_1,\ldots,i_q} = |i_1, \ldots, i_q\rangle\langle i_1, \ldots, i_q| \otimes \mathbb{I}$, which determines all q qubits to be accessed. If $k = 1$, we simply refer to a (q)-QPCP_{NA} verifier.*

Some additional explanations are in order. First, one might wonder why Definition 4 is not defined entirely in terms of PVMs that absorb the circuits $V^{t'}$. As we will see later, this split is necessary to ensure an efficient unitary decomposition whenever the PVMs are at most $\mathcal{O}(\log)$-local. Throughout this work, we assume an ordering on the tuples (j, l), where (j, l) represents the lth qubit of the jth proof, and we use basis state notation, treating strings and their integer representations interchangeably. Since there are only $kp_2(n)$ possible settings, we need to allocate at most $\lceil \log(kp_2(n)) \rceil = \mathcal{O}(\log n)$ qubits to determine the next proof index to be queried when $k, p_2(n) \in \text{poly}(n)$. This ensures that the PVMs remain $\mathcal{O}(\log n)$-local. In the constant-query setting (which is our focus), this property still holds in the non-adaptive case, even when a single PVM represents

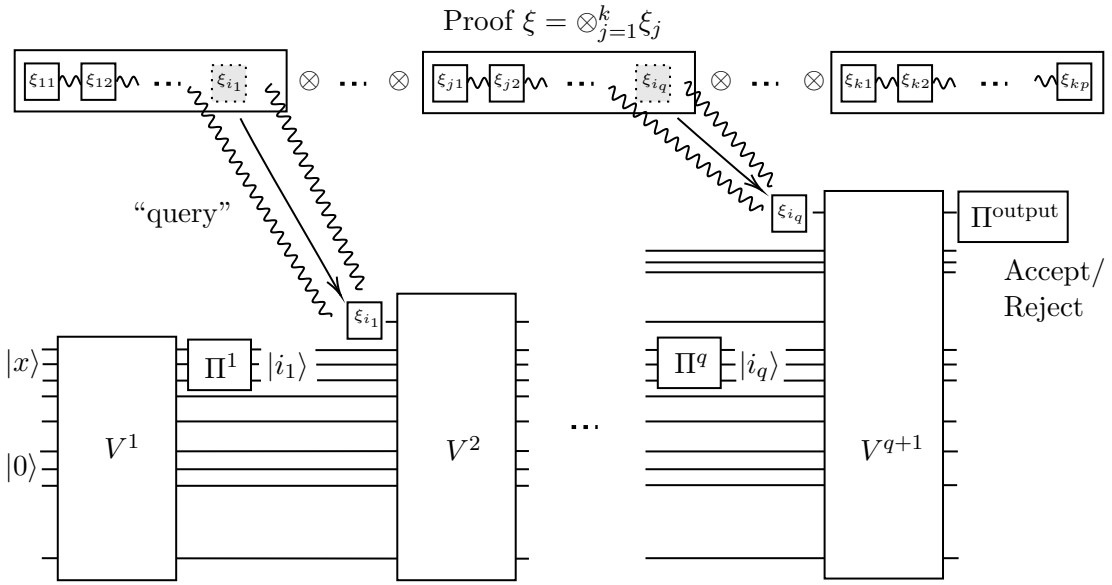

Figure 1: Visualisation of a k-prover quantum PCP which allows for q adaptive queries to the provided quantum proof through intermediate measurements as per Definition 4. Note that an index measurement result i_t denotes a tuple (j, l) where j indicates the corresponding proof and l the index of the qubit in this proof.

the parallel application of all single-index PVMs. As a result, each PVM has at most polynomial circuit complexity, making them efficiently implementable.

We can now define the associated complexity classes.

Definition 5 (QPCP)**.** *Let $n \in \mathbb{N}$ be the input size, and $x \in \{0,1\}^n$ be some input. A promise problem $A = (A_{\text{yes}}, A_{\text{NO}})$ belongs to* QPCP$[k, q, c, s]$ *if and only if there exist polynomially bounded functions $p_1, p_2, p_3 : \mathbb{N} \to \mathbb{N}$ and a* P*-uniform family of (k, q, p_1, p_2, p_3)-*QPCP *verifiers $\{V_n\}$ such that*

- *If $x \in A_{\text{YES}}$, then there exist quantum states $\xi_j \in \mathcal{D}\left((\mathbb{C}^2)^{\otimes p_2(n)}\right)$, $j \in [k]$, such that V_n accepts $(x, \otimes_{j=1}^{k} \xi_j)$ with probability at least c,*
- *If $y \in A_{\text{NO}}$, then for all quantum states $\xi_j \in \mathcal{D}\left((\mathbb{C}^2)^{\otimes p_2(n)}\right)$, $j \in [k]$, we have that V_n accepts $(x, \otimes_{j=1}^{k} \xi_j)$ with probability at most s.*

If $c = 2/3$ and $s = 1/3$, we simply write QPCP$[k, q]$*, and use* QPCP$[q]$ *if also $k = 1$.*

Definition 5a (QPCP$_{\text{NA}}$)**.** *This follows the same definition as for* QPCP *but now with a non-adaptive* QPCP*-verifier as per Definition 3.1a.*

We can easily show that in the single-prover setting we have weak error reduction by parallel repetition by using a similar argument as in [22].

Lemma 2 (Weak error reduction for the single prover case)**.** *Let $c - s \in \Omega(1)$. Then*

$$\text{QPCP}[1, q, c, s] \subseteq \text{QPCP}[1, \mathcal{O}(qt), 1 - 2^t, 2^{-t}].$$

Proof. This follows from a standard parallel repetition argument, with special care given to the fact that the proof can be entangled. Let V be the (k, q, p_1, p_2, p_3)-QPCP verifier with completeness c and soundness s, where $p_1, p_2, p_3 \in \text{poly}(n)$. Arthur expects to receive the proof $\xi^{\otimes R}$, runs V R times in parallel (acting only on q qubits of each ξ), measures the output qubit, and accepts if at least a $(c+s)/2$-fraction of the outcomes are accepting. Completeness follows directly from a Chernoff bound. If Merlin provides the correct proof $\xi^{\otimes R}$, then each run of the verifier accepts with probability at least c. Let $X_i \in \{0, 1\}$ be the random variable that indicates whether the ith run of the parallel repetition accepted ($X_i = 1$) or not ($X_i = 0$). Let $\{X_i\}_{i\in[R]}$ be the outcomes of the R runs, with $\mu = \mathbb{E}[X_1] = c$. The total number of accepting runs is given by $S_R = \sum_{i=1}^{R} X_i$. By the Chernoff bound, the probability that fewer than a $(c + s)/2$-fraction of the runs accept is given by

$$\Pr\left[S_R < \frac{c+s}{2} \cdot R\right] \leq \exp\left(-2R\left(\frac{c-s}{2}\right)^2\right).$$

To ensure that $\Pr[S_R < (c + s)/2 \cdot R] \leq 2^{-t}$, it suffices to choose

$$R := \left\lceil \frac{2t \ln 2}{(c-s)^2} \right\rceil. \tag{1}$$

For soundness, let ρ be the $p_2(n)t$-qubit proof that Merlin provides instead of $\xi^{\otimes t}$. From the soundness property of the verifier, we know that $\mathbb{E}[X_1] \leq s$. Now consider the expectation of X_2, which depends on the outcome of X_1. However, the soundness condition ensures that $\mathbb{E}[X_2 \mid X_1 = j] \leq s$ for all possible outcomes $j \in \{0, 1\}$. By repeating this argument, we see that for any i, $\mathbb{E}[X_i \mid X_1, \ldots, X_{i-1}] \leq s$. Since this holds for any i, we can upper bound the acceptance probability by polynomially many independent Bernoulli trials with mean $\mu = s$, again with bias $(c-s)/2$. Applying a Chernoff bound for dependent variables (with bounded conditional expectations), we find that the acceptance probability is at most 2^{-t}. Finally, the total number of queries to the proof is $qR = \mathcal{O}(qt)$. □

We leave open the question of weak error reduction for the multiple-prover case, but we expect that it can be done using the ideas in [26].

When we have near-perfect completeness, strong error reduction is also possible for non-adaptive quantum PCPs.

Claim 1 (Strong error reduction for non-adaptive QPCPs with near-perfect completeness)**.** *For $l \in \mathcal{O}(1)$ it holds that*

$$\mathsf{QPCP}_{\text{NA}}[1, q, c, s] = \mathsf{QPCP}_{\text{NA}}[1, lq, c', s']$$

with $c = 1 - 2^{-\Omega(n)}$, $s = 1/2$ and $c' = 1 - 2^{-\Omega(n)}$, $s' = 2^{-\mathcal{O}(l)}$.

The proof can be found in Appendix A.

4 Local Hamiltonians from quantum PCPs

The core of this section is Theorem 1, which argues that we can, from a QPCP verification circuit (as in Definition 4) efficiently produce a local Hamiltonian, such that the expectation value of a proof state is given by its acceptance probability by the verifier. Our construction broadly follows the ideas in [21], but also includes the more general adaptive verifier case. This leads to a slightly more general class of local Hamiltonians than is usually assumed in the quantum PCP literature, and we have to perform some extra tricks to

obtain the traditional Hamiltonian.

We begin by proving a basic lemma, which expresses the probability that a proof ξ gets rejected by the quantum PCP verifier V conditioned on taking the query path $(i_1, \ldots, i_q)$, in terms of the PVMs and circuits of V. Throughout this work, we make a distinction between indices indicated surrounded by brackets (e.g. "$(i_1, \ldots, i_q)$") and those that are not (e.g. "$i_1, \ldots, i_q$") to make a distinction where the order does matter (the former) and where it does not (the latter).

Lemma 3. *Let V_x be a (k, q, p_1, p_2, p_3)-QPCP verifier as in Definition 4, with hardcoded input $x \in \{0,1\}^n$. Define $M^{t'}_{i_q,x} = \Pi^{t'}_{i_q} V^{t'}_x$ for all $i_q \in [kp_2(n)]$, $t' \in [q]$. The probability that the quantum PCP rejects a proof ξ, conditioned on taking the query path $(i_1, \ldots, i_q)$, is given by*

$$\Pr[V_x \text{ rejects } \xi | (i_1, \ldots, i_q)] = \frac{\operatorname{tr}\left[P_{x,(i_1,\ldots,i_q)} \rho^0\right]}{\Pr[(i_1, \ldots, i_q)]},$$

where $\Pr[(i_1, \ldots, i_q)]$ is the probability that $i_1, \ldots, i_q$ are queried (and in this order), $\rho^0 = |0\rangle\langle 0|^{\otimes n+p_1(n)} \otimes \xi$, and $P_{x,(i_1,\ldots,i_q)}$ is a $(k + n + p_1(n))$-local operator given by

$$P_{x,(i_1,\ldots,i_q)} = M^{1\dagger}_{i_1,x} \ldots M^{q\dagger}_{i_q,x} V^{q+1\dagger}_x \Pi^{\text{output}}_0 V^{q+1}_x M^q_{i_q,x} \ldots M^1_{i_1,x}. \tag{2}$$

Proof. This follows from a straightforward application of the rules for post-measurement states in projective measurements. Let $\rho^0 = |0\rangle\langle 0|^{\otimes n+p_1(n)} \otimes \xi$ be the initial state (note the extra n term for the original input register).

Suppose the first PVM of the quantum PCP returns outcome i_1. The post-measurement state after this step is:

$$\rho^1 = \frac{\Pi^1_{i_1} V^1_x \rho^0 V^{1\dagger}_x \Pi^1_{i_1}}{\operatorname{tr}\left[\Pi^1_{i_1} V^1_x \rho^0 V^{1\dagger}_x\right]} = \frac{\Pi^1_{i_1} V^1_x \rho^0 V^{1\dagger}_x \Pi^1_{i_1}}{\Pr[i_1]} = \frac{M^1_{i_1,x} \rho^0 M^{1\dagger}_{i_1,x}}{\Pr[i_1]}.$$

Similarly, assuming outcome i_2 for the second query, the state becomes:

$$\rho^2 = \frac{\Pi^2_{i_2} V^2_x \rho^1 V^{2\dagger}_x \Pi^2_{i_2}}{\operatorname{tr}\left[\Pi^2_{i_2} V^2_x \rho^1 V^{2\dagger}_x\right]} = \frac{\Pi^2_{i_2} V^2_x \rho^1 V^{2\dagger}_x \Pi^2_{i_2}}{\Pr[i_2|i_1]} = \frac{M^2_{i_2,x} \rho^1 M^{2\dagger}_{i_2,x}}{\Pr[i_2|i_1]} = \frac{M^2_{i_2,x} M^1_{i_1,x} \rho^0 M^{1\dagger}_{i_1,x} M^{2\dagger}_{i_2,x}}{\Pr[i_2|i_1]\Pr[i_1]}.$$

Repeating this procedure $q-2$ more times, assuming outcomes $i_3, \ldots, i_q$, we find that the state after the q'th query becomes

$$\rho^q = \frac{M^q_{i_q,x} \ldots M^1_{i_1,x} \rho^0 M^{1\dagger}_{i_1,x} \ldots M^{q\dagger}_{i_q,x}}{\prod_{l=1}^q \Pr[i_l|(i_1, \ldots, i_{l-1})]} = \frac{M^q_{i_q,x} \ldots M^1_{i_1,x} \rho^0 M^{1\dagger}_{i_1,x} \ldots M^{q\dagger}_{i_q,x}}{\Pr[(i_1, \ldots, i_q)]}.$$

Now in the final step of the quantum PCP, a final circuit V_l is applied, followed by the PVM Π^{output}. The expected value of rejection is then given by

$$\Pr[V_x \text{ rejects } \xi | (i_1, \ldots, i_q)] = \operatorname{tr}\left[\Pi^{\text{output}}_0 V^q_x \rho^q V^{q\dagger}_x\right] = \frac{\operatorname{tr}\left[\Pi_0 V^q_x M^q_{i_q,x} \ldots M^1_{i_1,x} \rho^0 M^{1\dagger}_{i_1,x} \ldots M^{q\dagger}_{i_q,x} V^{q\dagger}_x\right]}{\Pr[(i_1, \ldots, i_q)]}.$$

Using the cyclic property of the trace, we can write:

$$\begin{aligned}\Pr[V_x \text{ rejects } \xi|(i_1,\ldots,i_q)] &= \frac{\mathrm{tr}\Big[M^{1\dagger}_{i_1,x}\ldots M^{q\dagger}_{i_q,x}V_x^{q\dagger}\Pi_0^{\text{output}}V_x^q M^q_{i_q,x}\ldots M^1_{i_1,x}\rho^0\Big]}{\Pr[(i_1,\ldots,i_q)]}\\ &= \frac{\mathrm{tr}\Big[P_{x,(i_1,\ldots,i_q)}\rho^0\Big]}{\Pr[(i_1,\ldots,i_q)]},\end{aligned}$$

with

$$P_{x,(i_1,\ldots,i_q)} = M^{1\dagger}_{i_1,x}\ldots M^{q\dagger}_{i_q,x}V_x^{q\dagger}\Pi_0^{\text{output}}V_x^q M^q_{i_q,x}\ldots M^1_{i_1,x}.$$

□

The next idea is that the expectation value of an operator A acting on an n-qubit state consisting of a product state of a q-qubit state and a fixed $(n-q)$-qubit state can be represented as an expectation value of a q-qubit operator B acting only on the q-qubit state. The following lemma proves this and gives an explicit expression of the local operator, assuming that the fixed state is pure.

Lemma 4. *Let A be an operator acting on an n-qubit Hilbert space consisting of a variable q-qubit input state ρ in a tensor product with some fixed $(n-q)$-qubit state ρ_{fixed}. Then we have that*

$$\mathrm{tr}[A(\rho\otimes\rho_{\text{fixed}})] = \mathrm{tr}[B\rho],$$

where $B = B(\rho_{\text{fixed}})$ is some q-local operator which depends on ρ_{fixed}. Moreover, if $\rho_{\text{fixed}} = |\psi\rangle\langle\psi|$ for some pure state $|\psi\rangle$, we have that the (α,α')-entry of B in some basis $\{\alpha\}$ is given by

$$b_{\alpha,\alpha'} = \langle\alpha|\langle\psi|A|\alpha'\rangle|\psi\rangle.$$

Proof. We can decompose A in two arbitrary bases $\{\alpha\}$ and $\{\beta\}$ for each part of the cut in the product state as

$$A = \sum_{\alpha,\alpha',\beta,\beta'} a_{\alpha,\alpha',\beta,\beta'}\,|\alpha\rangle\langle\alpha'|\otimes|\beta\rangle\langle\beta'|.$$

Using the linearity of the trace and the tensor product property,

$$\begin{aligned}\mathrm{tr}[A(\rho\otimes\rho_{\text{fixed}})] &= \mathrm{tr}\left[\sum_{\alpha,\alpha',\beta,\beta'} a_{\alpha,\alpha',\beta,\beta'}\,|\alpha\rangle\langle\alpha'|\otimes|\beta\rangle\langle\beta'|\,(\rho\otimes\rho_{\text{fixed}})\right]\\ &= \sum_{\alpha,\alpha',\beta,\beta'} a_{\alpha,\alpha',\beta,\beta'}\,\mathrm{tr}\big[|\alpha\rangle\langle\alpha'|\,\rho\big]\,\mathrm{tr}\big[|\beta\rangle\langle\beta'|\,\rho_{\text{fixed}}\big]\\ &= \sum_{\alpha,\alpha'}\left(\sum_{\beta,\beta'} a_{\alpha,\alpha',\beta,\beta'}\,\mathrm{tr}\big[|\beta\rangle\langle\beta'|\,\rho_{\text{fixed}}\big]\right)\mathrm{tr}\big[|\alpha\rangle\langle\alpha'|\,\rho\big]\\ &= \sum_{\alpha,\alpha'} b_{\alpha,\alpha'}\,\mathrm{tr}\big[|\alpha\rangle\langle\alpha'|\,\rho\big]\\ &= \mathrm{tr}[B\rho],\end{aligned}$$

where the operator B, given by

$$B = \sum_{\alpha,\alpha'} b_{\alpha,\alpha'} \left|\alpha\right\rangle\!\left\langle\alpha'\right|, \quad b_{\alpha,\alpha'} = \sum_{\beta,\beta'} a_{\alpha,\alpha',\beta,\beta'} \operatorname{tr}\left[\left|\beta\right\rangle\!\left\langle\beta'\right| \rho_{\text{fixed}}\right],$$

is indeed q-local.

For the second part of the lemma, we note that under the assumption that $\rho_{\text{fixed}} = |\psi\rangle\langle\psi|$ for some pure state $|\psi\rangle$, we have

$$\begin{aligned}
\langle\alpha| \langle\psi| A |\alpha'\rangle |\psi\rangle &= \langle\alpha| \langle\psi| \left(\sum_{\alpha,\alpha',\beta,\beta'} a_{\alpha,\alpha',\beta,\beta'} \left|\alpha\right\rangle\!\left\langle\alpha'\right| \otimes \left|\beta\right\rangle\!\left\langle\beta'\right| \right) |\alpha'\rangle |\psi\rangle \\
&= \sum_{\beta,\beta'} a_{\alpha,\alpha',\beta,\beta'} \langle\psi| \left|\beta\right\rangle\!\left\langle\beta'\right| |\psi\rangle \\
&= \sum_{\beta,\beta'} a_{\alpha,\alpha',\beta,\beta'} \operatorname{tr}\left[\left|\beta\right\rangle\!\left\langle\beta'\right| |\psi\rangle\langle\psi|\right] \\
&= \sum_{\beta,\beta'} a_{\alpha,\alpha',\beta,\beta'} \operatorname{tr}\left[\left|\beta\right\rangle\!\left\langle\beta'\right| \rho_{\text{fixed}}\right] \\
&= b_{\alpha,\alpha'},
\end{aligned}$$

completing the proof. □

With these lemmas in hand we can argue that, given a verifier, there always exists a local Hamiltonian that captures the probability of acceptance of a proof.

Lemma 5 (Hamiltonians from general quantum PCPs)**.** *Let $q \in \mathbb{N}$ be some constant and let p_1, p_2, and p_3 be polynomials. Let V_x be a (k, q, p_1, p_2, p_3)-*QPCP*-verifier as in Definition 4, with hardcoded input x, $|x| = n$ for some $n \in \mathbb{N}$, and a $kp_2(n)$-qubit quantum proof $\xi = \otimes_{j=1}^{k} \xi_j$, where $\xi_j \in \mathcal{D}\left((\mathbb{C}^2)^{\otimes p_2(n)}\right)$. Then there exists a Hamiltonian H_x consisting of q-local PSD terms acting on $kp_2(n)$-qubits such that for all ξ, we have*

$$\Pr[V_x \text{ accepts } \xi] = 1 - \operatorname{tr}[H_x \xi]. \tag{3}$$

Proof. Let Ω be the set of all unordered subsets of size q drawn from $[kp_2(n)]$, i.e.,

$$\Omega = \binom{[kp_2(n)]}{q} = \{\{i_1, \ldots, i_q\} \mid i_j \in [kp_2(n)],\ i_j \neq i_k \text{ for } j \neq k\}.$$

By the definition of conditional probability,

$$\begin{aligned}
\Pr[V_x \text{ accesses qubits } (i_1, \ldots, i_q) \text{ from } |\xi\rangle \text{ and rejects }] &= \Pr[(i_1, \ldots, i_q)] \cdot \Pr[V \text{ rejects } \xi | (i_1, \ldots, i_q)] \\
&= \Pr[(i_1, \ldots, i_q)] \cdot \frac{\operatorname{tr}\left[P_{x,(i_1,\ldots,i_q)} \rho\right]}{\Pr[(i_1, \ldots, i_q)]} \\
&= \operatorname{tr}\left[P_{x,(i_1,\ldots,i_q)} \sigma_{i_1,\ldots,i_q}\right],
\end{aligned}$$

where

$$\sigma_{i_1,\ldots,i_q} = \operatorname{tr}_{\bar{C}_{i_1,\ldots,i_q}} \left[\xi \otimes |0\rangle\langle 0|^{\otimes n + p_2(n)}\right]$$

with $\bar{C}_{i_1,\ldots,i_q}$ denoting all qubits of ξ except for those with indices $i_1,\ldots,i_q$. Hence, we can write the probability that V_x rejects ξ as

$$\Pr[V_x \text{ rejects } \xi] = \sum_{\{i_1,\ldots,i_q\}\in\Omega} \sum_{(i_1,\ldots,i_q)\in\{(i_1,\ldots,i_q)\}!} \operatorname{tr}\Big[P_{x,(i_1,\ldots,i_q)}\sigma_{i_1,\ldots,i_q}\Big].$$

For all $\{i_1,\ldots,i_q\}\in\Omega$, we define the $2^q\times 2^q$ matrix $H_{x,i_1,\ldots,i_q}$ as

$$\langle\alpha| H_{x,i_1,\ldots,i_q} |\beta\rangle := \sum_{(i_1,\ldots,i_q)\in\{(i_1,\ldots,i_q)\}!} \left(\langle 0|^{\otimes p_1(n)+n}\otimes\langle\alpha|\right) P_{x,(i_1,\ldots,i_q)} \left(|0\rangle^{\otimes p_1(n)+n}\otimes|\beta\rangle\right), \tag{4}$$

where $\alpha,\beta\in\{0,1\}^q$. By Lemma 4, we have that for any q-local density matrix ρ we have that the expression

$$\operatorname{tr}\big[H_{x,i_1,\ldots,i_q}\rho\big] = \operatorname{tr}\left[\sum_{(i_1,\ldots,i_q)\in\{(i_1,\ldots,i_q)\}!} P_{x,(i_1,\ldots,i_q)} |0\rangle\langle 0|^{\otimes n+p_2(n)}\otimes\rho\right] \tag{5}$$

holds. Moreover, since Eq. (5) is the sum of all probabilities that a query path $(i_1,\ldots,i_1)$ is taken and the proof is rejected, taken over all possible permutations of the indices, we must have that $H_{x,i_1,\ldots,i_q}$ is PSD. Now consider the q-local Hamiltonian H defined as

$$H_x = \sum_{\{i_1,\ldots,i_q\}\in\Omega} H_{x,i_1,\ldots,i_q}.$$

We have that

$$\Pr[V_x \text{ accepts } \xi] = 1 - \operatorname{tr}[H_x\xi],$$

since by the linearity of the trace

$$\begin{aligned}
\operatorname{tr}[H_x\xi] &= \operatorname{tr}\left[\sum_{\{i_1,\ldots,i_q\}\in\Omega} \left(H_{x,i_1,\ldots,i_q}\otimes\mathbb{I}\right)\xi\right] \\
&= \sum_{\{i_1,\ldots,i_q\}\in\Omega} \operatorname{tr}\left[\left(H_{x,i_1,\ldots,i_q}\otimes\mathbb{I}\right)\xi\right] \\
&= \sum_{\{i_1,\ldots,i_q\}\in\Omega} \operatorname{tr}\Big[H_{x,i_1,\ldots,i_q}\operatorname{tr}_{\bar{C}_{i_1,\ldots,i_q}}[\xi]\Big] \\
&= \sum_{\{i_1,\ldots,i_q\}\in\Omega} \sum_{(i_1,\ldots,i_q)\in\{(i_1,\ldots,i_q)\}!} \operatorname{tr}\Big[P_{x,(i_1,\ldots,i_q)}\sigma_{(i_1,\ldots,i_q)}\Big] \\
&= \sum_{\{i_1,\ldots,i_q\}\in\Omega} \sum_{(i_1,\ldots,i_q)\in\{(i_1,\ldots,i_q)\}!} \Pr[(i_1,\ldots,i_q)]\Pr[V \text{ rejects } \xi|(i_1,\ldots,i_q)] \\
&= \Pr[V \text{ rejects } \xi] \\
&= 1-\Pr[V \text{ accepts } \xi],
\end{aligned}$$

which also implies that $H_{x,i_1,\ldots,i_q}\preceq 1$ for all $i_1,\ldots,i_1\in\Omega$. □

4.1 Learning the Hamiltonian

We have shown that the probability that a QPCP accepts a certain proof is equivalent to the expectation value of some Hamiltonian. We still have to show how to obtain the entries

of each term efficiently. Before we state the final protocol, let us argue that we can indeed learn each local term up to arbitrary (inverse polynomial) precision. To do this we will use the Hadamard test, introduced in [39].

We need a simple generalisation of the one presented in [39], as we require two different input states on both sides of the inner product. A proof of this can be found in Chapter 2 of [21].

Lemma 6 (Hadamard test [39])**.** *Let $|\psi\rangle, |\phi\rangle \in \mathbb{C}^{2^n}$ be quantum states with state preparation unitaries U_ϕ, U_ψ, i.e. $|\psi\rangle = U_\psi |0^n\rangle$ and $|\phi\rangle = U_\phi |0^n\rangle$. Let $W \in \mathbb{U}(2^n)$ be some unitary. Then there exists a polynomial-time quantum algorithm that outputs an estimate $\hat{z}$ such that*

$$|\hat{z} - \mathrm{Re}(\langle\psi| W |\phi\rangle)| \leq \epsilon$$

with probability $\geq 1 - \delta$, in

$$\mathcal{O}\left(\frac{\log\left(\frac{1}{\delta}\right)}{\epsilon^2}\right).$$

(controlled) queries to U_ψ, U_ϕ and W. Similarly, there exists a quantum algorithm to estimate $\mathrm{Im}(\langle\psi| W |\phi\rangle)$ at the cost of applying one additional single-qubit gate.

Note that the Hadamard test only works for unitaries. Therefore, for our purposes we also need the following lemma, which shows that every (local) projector on a basis state can be written as a linear combination of unitaries with short circuit depth.

Lemma 7. *Let $\Pi_q = |q\rangle\langle q|$, where $q \in \{0,1\}^k$ is a basis state. Then Π_q can be written as*

$$\Pi_q = \bigotimes_{i\in[k]} \frac{Z_i + (1-2q_i)\mathbb{I}}{2} = \frac{1}{2^k} \sum_{j\in\{0,1\}^k} a_j U_j,$$

where $U_j \in \{I, Z\}^{\otimes k}$ and $a_j \in \{-1, +1\}$.

Proof. We have that $|0\rangle\langle 0| = (Z + \mathbb{I})/2$ and $|1\rangle\langle 1| = (Z - \mathbb{I})/2$. Hence, for $q_i \in \{0,1\}$, we obtain

$$|q_i\rangle\langle q_i| = \frac{Z_i + (1-2q_i)\mathbb{I}}{2}.$$

Since $|q\rangle\langle q| = \bigotimes_{i\in[k]} |q_i\rangle\langle q_i|$, we have that

$$\Pi_q = \bigotimes_{i\in[k]} |q_i\rangle\langle q_i| = \bigotimes_{i\in[k]} \frac{Z_i + (1-2q_i)\mathbb{I}}{2} = \frac{1}{2^k} \sum_{j\in\{0,1\}^k} a_j U_j, \tag{6}$$

where each U_j is of the form $V_1^j \otimes V_2^j \otimes \cdots \otimes V_k^j$ for $V_i^j \in \{\mathbb{I}, Z\}$ and $a_j \in \{-1, 1\}$. □

Before we move to prove the existence of the reduction, we need to define some parameters. The operators $H_{x,i_1,\ldots,i_q}$ (whose entries are defined in Eq. (4)) are composed of $q+1$ unitaries $\{V^t\}$, a total of q of $\mathcal{O}(\log n)$-local PVM elements (see Section 3), and a single 1-local PVM element (which is Π_{output}). If we use $\log(kp_2(n)) + 1$ qubits for each Π^t and decompose each PVM element in $H_{x,i_1,\ldots,i_q}$ into unitaries, as per Lemma 7, we can write

$$\langle\alpha| H_{x,i_1,\ldots,i_q} |\beta\rangle = \frac{1}{\Gamma} \sum_{(i_1,\ldots,i_q)\in S(\{(i_1,\ldots,i_q)\})} \sum_{j\in[\Gamma]} a_j \left(\langle 0|^{\otimes p_1(n)+n} \otimes \langle\alpha|\right) U_{j,x} \left(|0\rangle^{\otimes p_1(n)+n} \otimes |\beta\rangle\right)$$

,

with

$$U_{j,x} = \prod_{l\in[4q+3]} U^l_{j,x},$$

where the unitaries $U^l_{j,x}$ are composed of a polynomial number of elementary gates and $a_j \in \{-1,+1\}$. The range of index l can be seen from inspecting Eq. (2) of Lemma 3: each $P_{x,(i_1,\ldots,i_q)}$, which makes up a $H_{x,(i_1,\ldots,i_q)}$, consists of $q+1$ V_t's and their conjugate transposes, two unitaries for each of the first q PVM elements coming from its decomposition and a final single one for the final outcome measurement (which is "sandwiched" in the middle of Eq. (2)). Hence, we have a total of $2(q+1)+2q+1 = 4q+3$ unitaries in the product. The total number of unitaries in the linear combination for each $(i_1,\ldots,i_q)$ is given by

$$\Gamma := \left(2^{\log(kp_2(n))+1}\right)^q \cdot 2 = 2^{q+1}(kp_2(n))^q.$$

We define

$$z^{\alpha,\beta,j}_{(i_1,\ldots,i_q)} := \left(\langle 0|^{\otimes p_1(n)+n} \otimes \langle\alpha|\right) U_{j,x} \left(|0\rangle^{\otimes p_1(n)+n} \otimes |\beta\rangle\right), \tag{7}$$

and

$$h^{\alpha,\beta}_{i_1,\ldots,i_q} = \sum_{(i_1,\ldots,i_q)\in S(\{(i_1,\ldots,i_q)\})} \sum_{j\in[\Gamma]} a_j z^{\alpha,\beta,j}_{(i_1,\ldots,i_q)}. \tag{8}$$

such that

$$\langle\alpha| H_{x,i_1,\ldots,i_q} |\beta\rangle = \frac{h^{\alpha,\beta}_{i_1,\ldots,i_q}}{\Gamma},$$

for all $\alpha,\beta \in \{0,1\}^q$ and $\{i_1,\ldots,i_q\} \in \Omega$.

Theorem 1. *Let $q \in \mathbb{N}$ be some constant and p_1,p_2,p_3 be polynomials. Let V_x be a $[k,q,p_1,p_2,p_3]$-QPCP-verifier as in Definition 4, taking input x, $|x| = n$ for some $n \in \mathbb{N}$, and a kp_2-qubit quantum proof ξ. For all $\epsilon > 0$, there exists a quantum reduction from V_x to a q-local Hamiltonian $\hat{H}_x = \sum_{i\in[m]} \hat{H}_{x,i}$, with $m = \mathrm{poly}(n)$, $\hat{H}_{x,i}$ PSD for each $i \in [m]$, and $\left\|\hat{H}_x\right\| \leq 1$, such that, given a proof ξ,*

$$\left|\Pr[V_x \text{ accepts } \xi] - \left(1 - \mathrm{tr}\left[\hat{H}_x\xi\right]\right)\right| \leq \epsilon. \tag{9}$$

The quantum reduction succeeds with probability $1-\delta$ and runs in time $\mathrm{poly}(n, 1/\epsilon, \log(1/\delta))$.

Proof. The reduction is specified in Algorithm 1 below. We proceed to show its correctness by arguing it has the required precision, success probability, and time complexity.

Precision: Step 1*a* of Algorithm 1 produces estimates $\tilde{z}^{\alpha,\beta,j}_{(i_1,\ldots,i_q)}$ for the parameters $z^{\alpha,\beta,j}_{(i_1,\ldots,i_q)}$ using Lemma 6. We have that

$$\left|z^{\alpha,\beta,j}_{(i_1,\ldots,i_q)} - \tilde{z}^{\alpha,\beta,j}_{(i_1,\ldots,i_q)}\right| \leq 2\epsilon',$$

since we estimated both the real and imaginary parts up to precision ϵ'. By the triangle inequality (and $a_j \in \{1, -1\}$) we now have

$$\left| \frac{\tilde{h}^{\alpha,\beta}_{i_1,\ldots,i_q}}{\Gamma} - \langle\alpha| H_{x,i_1,\ldots,i_q} |\beta\rangle \right| \leq 2q!\epsilon'.$$

Since $\tilde{H}_{x,i_1,\ldots,i_q} = \sum_{\alpha,\beta\in\{0,1\}^q} \tilde{h}^{\alpha,\beta}_{i_1,\ldots,i_q} |\alpha\rangle \langle\beta|$, we have that

$$\begin{aligned}\left\| \tilde{H}_{x,i_1,\ldots,i_q} - H_{x,i_1,\ldots,i_q} \right\| &\leq 2^q \max_{\alpha,\beta} \left| \langle\alpha| H_{x,i_1,\ldots,i_q} |\beta\rangle - \langle\alpha| \tilde{H}_{x,i_1,\ldots,i_q} |\beta\rangle \right| \\ &\leq 2^{q+1} q!\epsilon',\end{aligned}$$

which follows from the bound on the operator norm by the max-norm. Now suppose that $\tilde{H}_{x,i_1,\ldots,i_q}$ is not PSD. Since $H_{x,i_1,\ldots,i_q}$ is PSD, we have that $\lambda_{\min}(\tilde{H}_{x,(i_1,\ldots,i_q)}) \geq -2^{q+1}q!\epsilon'$, so we have that adding the identity term can only double the error, making it at most $2^{q+2}q!\epsilon'$. By another triangle inequality

$$\left\| \tilde{H}_x - H_x \right\| \leq |\Omega| 2^{q+2} q!\epsilon' \leq \epsilon/4,$$

for our choice of ϵ'. Finally, the error introduced by step 3 of the protocol can be bounded by

$$\begin{aligned}\left\| \hat{H}_x - H_x \right\| &\leq \left\| \hat{H}_x - \tilde{H}_x \right\| + \left\| \tilde{H}_x - H_x \right\| \\ &\leq \left| \frac{1}{1+\epsilon/4} - 1 \right| \left\| \tilde{H}_x \right\| + \frac{\epsilon}{4} \\ &\leq \frac{\epsilon}{4}(1 + \frac{\epsilon}{4}) + \frac{\epsilon}{4} \\ &= \frac{\epsilon}{2} + \left(\frac{\epsilon}{4}\right)^2 \\ &\leq \epsilon.\end{aligned}$$

Hence, for any state $\xi = \sum_i p_i |\psi_i\rangle\langle\psi_i|$, with $\sum_i p_i = 1$, we have

$$\begin{aligned}\left| \Pr[V_x \text{ accepts } \xi] - \left(1 - \operatorname{tr}\left[\hat{H}_x \xi\right]\right) \right| &= \left| \operatorname{tr}\left[\hat{H}_x \xi\right] - \operatorname{tr}[H_x \xi] \right| \\ &= \left| \operatorname{tr}\left[(\hat{H}_x - H_x)\xi \right] \right| \\ &= \left| \operatorname{tr}\left[(\hat{H}_x - H_x) \sum_i p_i |\psi_i\rangle\langle\psi_i| \right] \right| \\ &= \sum_i p_i \left| \langle\psi_i| (\hat{H}_x - H_x) |\psi_i\rangle \right| \\ &\leq \epsilon,\end{aligned}$$

as desired.

Success probability: We have to count the number of times we run the Hadamard test of Lemma 6, each of which succeeds with success probability $\geq 1 - \delta'$. Recall that Ω is the set of all possible indices (when order does not matter), which is given by $\Omega = \binom{[kp_2(n)]}{q}$ for proofs of length p_2. We run the Hadamard test for a total of $|\Omega| q! 4^q \Gamma$ times (see the number of iterations in Algorithm 1), and thus

$$(1 - \delta')^{|\Omega| q! 4^{q+1} \Gamma} \geq 1 - \delta' |\Omega| q! 4^{q+1} \Gamma = 1 - \delta,$$

using the inequality $(1 - x)^T \geq 1 - Tx$ for all $x \in [0, 1]$. The extra factor of two again accounts for estimating both the real and imaginary parts.

Time complexity: By definition of QPCP$[q]$, we have that V has gate complexity poly(n). Using Lemma 7, we have that V, U_ϕ, and U_ψ always have polynomially bounded gate complexities. Filling in our choice of δ' and ϵ', we have that the total number of (controlled) applications of V, U_ϕ, and U_ψ can be upper bounded as

$$\mathcal{O}\left(q!4^{q+1}\Gamma\frac{(|\Omega|2^{q+4}q!)^2\log\left(\frac{q!4^{q+1}\Gamma}{\delta}\right)}{\epsilon^2}\right)=\text{poly}(n,1/\epsilon,\log(1/\delta)),$$

since for $k=\text{poly}(n)$ and $q=\mathcal{O}(1)$ we have $|\Omega|=\text{poly}(n)$ and $\Gamma=\text{poly}(n)$. □

Algorithm 1: Quantum reduction from QPCP$[k,q]$ verification to a local Hamiltonian

Input: A (k,q,p_1,p_2,p_3)-QPCP verifier V_x with hardcoded input x, a precision parameter ϵ (variation: η), maximum error probability δ.

Set: $\Omega:=\binom{[kp_2(n)]}{q}$, $\Gamma:=2^{q+1}(kp_2(n))^q$, $\epsilon':=\frac{\epsilon}{|\Omega|2^{q+4}q!}$, $\delta':=\frac{\delta}{|\Omega|q!4^{q+1}\Gamma}$.

Algorithm:

1. For $\{i_1,\dots,i_q\}\in\Omega$
   (a) For $(i_1,\dots,i_q)\in\{(i_1,\dots,i_q)\}!$
      i. For $\alpha\in\{0,1\}^q$, $\beta\in\{0,1\}^q$:
         - For $j\in[\Gamma]$ compute $a_j,U_{j,x}$ and estimate $z^{\alpha,\beta,j}_{(i_1,\dots,i_q)}$ as $\tilde{z}^{\alpha,\beta,j}_{(i_1,\dots,i_q)}$ using Lemma 6 for both the real and imaginary part with $V=U_{j,x}$, $U_\theta=\hat{V}(\otimes_{i=1}^q(X_i)^{\alpha_i}\otimes\mathbb{I})$ and $U_\phi=\hat{V}(\otimes_{i=1}^q(X_i)^{\beta_i}\otimes\mathbb{I})$, ϵ' and δ'.

      ii. Set $\tilde{h}^{\alpha,\beta}_{i_1,\dots,i_q}=\sum_{(i_1,\dots,i_q)\in S(\{(i_1,\dots,i_q)\}}\sum_{j\in[\Gamma]}a_j\tilde{z}^{\alpha,\beta,j}_{(i_1,\dots,i_q)}$.

   (b) Set $\tilde{H}_{x,(i_1,\dots,i_q)}=\sum_{\alpha,\beta\in\{0,1\}^q}\tilde{h}^{\alpha,\beta}_{i_1,\dots,i_q}|\alpha\rangle\langle\beta|$.

   (c) Compute $\lambda_{\min}(\tilde{H}_{x,i_1,\dots,i_q})$. If $\lambda_{\min}(\tilde{H}_{x,i_1,\dots,i_q})<0$, let $\tilde{H}_{x,i_1,\dots,i_q})\leftarrow\tilde{H}_{x,i_1,\dots,i_q})-\lambda_{\min}(\tilde{H}_{x,i_1,\dots,i_q})\mathbb{I}$, else continue.

2. Let $\tilde{H}_x=\sum_{\{i_1,\dots,i_q\}\in\Omega}\tilde{H}_{x,i_1,\dots,i_q}$. Output

$$\hat{H}_x=\frac{\tilde{H}_x}{\text{argmax}\{\left\|\tilde{H}_x\right\|,1\}}.$$

Variation: To learn $\tilde{H}_x$ up to η bits of precision, use $\epsilon':=\frac{2}{(2^\eta+1)}$ and keep only the first η bits for the estimates of $z^{\alpha,\beta,j}_{(i_1,\dots,i_q)}$ as $\tilde{z}^{\alpha,\beta,j}_{(i_1,\dots,i_q)}$ in Step 1(a)i.

One can use a simple trick that exploits the freedom we have in the reduction to set the precision. Since every local term is bounded in the operator norm, we must also have that for each of these terms the matrix entries are bounded by 1. Hence, we can convert a bound in terms of approximation in entry-wise error to one in having a least a certain number of bits from a certain bit-wise representation of the value being correct. The advantage of the latter is that it allows us to make the reduction deterministic in the sense that if it succeeds, it always produces *exactly* the same Hamiltonian. This can alternatively be

viewed as having a "rounding scheme" which, with high probability, always rounds to the same operator in operator space. Though not strictly needed, this allows one to always reason about the same Hamiltonian when the reduction is used as a subroutine.

Corollary 1. *For any $\epsilon, \delta > 0$, under the same setup as in Theorem 1, there exists a quantum reduction which produces a fixed Hamiltonian $\tilde{H}_x$ with probability $1-\delta$ which satisfies Eq. (9) and runs in time* $\mathrm{poly}(n, 1/\epsilon, \log(1/\delta))$.

We have put the proof of Corollary 1 in Appendix C.

Remark 2. *Unlike these general quantum PCPs, where the probabilities of taking a certain query path can depend on the proof, we have that for non-adaptive quantum PCPs the probability distribution over which the proof is accessed only depends on the input. Generalising the proof of Lemma 6.3 in [19], one can easily verify that for non-adaptive quantum PCPs the obtained Hamiltonian can be assumed to be of the form $H_x = \sum_{i\in[m]} p_{x,i} H_{x,i}$, where the $p_{x,i}$'s form a probability distribution and $0 \preceq H_{x,i} \preceq 1$ for all q-local terms. For these general (adaptive) quantum PCPs, the key difference is that each term of the Hamiltonian encodes both the probability that certain parts of the proof are accessed as well as the eventual probability of acceptance. We have provided sufficient conditions on the error parameters of the reduction in Appendix B.*

4.2 Local smoothings of local Hamiltonians

In this subsection we show that one can always transform any Hamiltonian in the form of H_x of Eq. (3) to match the form usually adopted in the literature; specifically, we want to transform a local Hamiltonian with PSD terms, operator norm at most 1 and constant promise gap to another Hamiltonian such that the promise gap becomes $\Omega(m)$, where m is the number of terms in the new Hamiltonian, whilst still having that each term H_i is local and satisfies $0 \preceq H_i \preceq 1$. This way, we have that at least a constant fraction of all the m terms contribute to the energy in the NO-case, which intuitively implies that only a constant number of terms have to be checked to find out whether the energy of the Hamiltonian is high or low (see Section 4.3). This turns out to be possible at the cost of increasing the locality by a factor of two and the promise gap by a constant factor.[8]

Lemma 8 (Hamiltonian local smoothing lemma)**.** *Let H be a q-local Hamiltonian on n-qubits, $q = \mathcal{O}(1)$, such that $H = \sum_{i\in[m]} H_i$, where each H_i is PSD and $0 \preceq H \preceq 1$, and $\lambda_{\min}(H) \leq a$ or $\geq b$ for $b - a = \gamma(m)$. Then there exists a polynomial time transformation to another $2q$-local Hamiltonian H' on n qubits such that $H' = \frac{1}{m}\sum_{i\in[m]} H'_i$, $0 \preceq H'_i \preceq 1$ for all $i \in [m]$ and $\lambda_{\min}(H) \leq \alpha$ or $\geq \beta$ for some β, α such that $\beta - \alpha \in \Omega(\gamma(m))$.*

Proof. Write $\rho_d = \frac{\mathbb{I}_d}{d}$ for the d-dimensional maximally mixed state. We have

$$\mathrm{tr}\left[H\rho_{2^n}\right] = \mathrm{tr}\left[\sum_{i\in[m]} H_i \,\mathrm{tr}_{\bar{C}_i}\left[\rho_{2^n}\right]\right] = \mathrm{tr}\left[\sum_{i\in[m]} H_i \rho_{2^q}\right] = \frac{1}{2^q}\sum_{i\in[m]} \mathrm{tr}[H_i].$$

The variational principle tells us that

$$\mathrm{tr}\left[H\rho_{2^n}\right] \leq \max_{\xi} \mathrm{tr}[H\xi] \leq \|H\| \leq 1.$$

[8] We believe such a result might be known in existing literature, but as we could not find a reference we provide a proof for completeness.

Combining the two, this implies

$$2^q \geq \sum_{i\in[m]} \mathrm{tr}[H_i] \geq \sum_{i\in[m]} \|H_i\|$$

using that $0 \preceq H_i \preceq 1$ implies that $\mathrm{tr}[H_i] \geq \|H_i\|$. Now consider $\hat{H} = \sum_{i\in[m]} \hat{H}_i$ where $\hat{H}_i = \frac{1}{2^{q+3}} H_i$. We have that $\lambda_0(\hat{H}) \geq \frac{b}{2^{q+3}}$ or $\lambda_0(\hat{H}) \leq \frac{a}{2^{q+3}}$. Let $\alpha_i = \left\|\hat{H}_i\right\|$, for which we know that $\sum_{i\in[m]} \alpha_i \leq \frac{1}{8}$. Define index sets:

$$L = \{i \in [m] : \alpha_i \leq \frac{1}{2m}\},$$
$$U = [m] \setminus L = \{i \in [m] : \alpha_i > \frac{1}{2m}\}.$$

We have

$$|U|\frac{1}{2m} \leq \sum_{i\in U} \alpha_i \leq \sum_{i\in[m]} \alpha_i \leq \frac{1}{8},$$

which implies $|U| \leq \frac{m}{4}$ and hence $|L| \geq \frac{3m}{4}$. We now construct a new Hamiltonian $H' = \sum_{i\in[m']} H'_i$, where each H'_i is $2q$-local and satisfies $0 \preceq H'_i \preceq 1$, by redistributing each high-norm term $\hat{H}_j$ from U into smaller pieces and assigning them to low-norm terms from L. This way, we are guaranteed that all new terms are positive semi-definite and have operator norm at most $1/m$, so we can simply scale with a factor m to obtain an operator norm bound of 1. For each α_j, assuming $\alpha_j > 1/m$ with $j \in U$, we want to find the largest possible integer t_j such that:

$$\frac{1}{2m} \leq \alpha_j - t_j\frac{1}{2m} \leq \frac{1}{m}.$$

Which implies that $t_j \geq 2m\alpha_j - 2$ and $t_j \leq 2m\alpha_j - 1$, so we can take $t_j = \lfloor 2m\alpha_j - 1 \rfloor$.

Consider the following procedure:

1. Initialise $L' \coloneqq L$.
2. For each $j \in U$, check if $\alpha_j > 1/m$. If this is not the case, continue the loop (or exit after the last j). If this is the case, set $t_j = \lfloor 2m\alpha_j - 1 \rfloor$. For the first t_j indices $i \in L'$, define
$$Q_i \coloneqq \hat{H}_i + \frac{1}{2m}\hat{H}_j.$$
These Q_i are at most $2q$-local and have operator norm at most $1/m$. Remove these t_j indices i from L'. Define the q-local leftover term
$$Q_j \coloneqq \left(1 - \frac{t_j}{2m}\right)\hat{H}_j.$$
3. For each remaining $j \in U$ that was not used in any redistribution, simply set $Q_j \coloneqq \hat{H}_j$.
4. Let $H'_i \coloneqq mQ_i$ for each term. Return $H' \coloneqq \frac{1}{m}\sum H'_i$.

We only have to check whether the above procedure does not run out of terms in L to redistribute to. We obtain:

$$\sum_{j\in U} t_j \leq \sum_{j\in U} \lfloor 2m\alpha_j - 1 \rfloor \leq \sum_{j\in U} 2m\alpha_j \leq \frac{1}{4}m \leq |L|,$$

using that $\sum_{j\in[m]} \alpha_j \leq 1/8$. By construction, each term in H' is at most $2q$-local and satisfies $0 \preceq H'_i \preceq 1$, which means that $\|H\| \leq 1$. Since $\hat{H} = H/2^{q+3}$ and H' has the same spectrum as $\hat{H}$ (we only combined different terms together and rescaled), we have:

- If $\lambda_0(H) \leq a$, then $\lambda_0(H') \leq \alpha$ with $\alpha \coloneqq \frac{a}{2^{q+3}}$
- If $\lambda_0(H) \geq b$, then $\lambda_0(H') \geq \beta$ with $\beta \coloneqq \frac{b}{2^{q+3}}$.

Hence, $\beta - \alpha = \Omega(\gamma(m))$ when $q = \mathcal{O}(1)$. □

4.3 Kitaev's energy estimation protocol

In this section we saw that general quantum PCP can be transformed into a local Hamiltonian with a constant promise gap (and a specific form). In the following sections it will prove useful to have a verifier for the local Hamiltonian problem suitable for Hamiltonians from quantum PCP verifiers. For this we will use Kitaev's QMA-protocol for the local Hamiltonian problem [3], albeit with a small modification to allow for sampling the terms according to some probability weights associated with the Hamiltonian. We formulate the energy protocol in the pure setting as the mixed state case can simply be viewed as a convex combination of acceptance probabilities given by the pure states.

Protocol 1: Kitaev's energy estimation protocol

Input: A classical description of a n-qubit, q-local Hamiltonian of the form $H = \sum_{i\in[m]} p_i H_i$, $0 \preceq H_i \preceq 1$ with weights $\{p_i\}$ such that $\sum_i p_i = 1$.

Protocol:

1. The prover sends the state $|\psi\rangle$.
2. For each H_i, $i \in [m]$, let $H_i = \sum_j \lambda_{i,j} |\lambda_{i,j}\rangle\langle\lambda_{i,j}|$ be its spectral decomposition. Define the $(q+1)$-local operator W_i such that W_i acts on a $(n+1)$-qubit space as

$$W_i |\lambda_{i,j}\rangle |b\rangle = |\lambda_{i,j}\rangle \left(\sqrt{\lambda_{i,j}} |b\rangle + \sqrt{1-\lambda_{i,j}} |b \oplus 1\rangle\right). \tag{10}$$

   The verifier picks a $i \in [m]$ with probability p_i, and applies W_i on $|\psi\rangle |0\rangle$, and measures the final qubit.
3. The verifier accepts if and only if the outcome is $|1\rangle$.

Lemma 9 (Kitaev's energy estimation protocol (weighted))**.** *Consider a q-local, n-qubit Hamiltonian $H = \sum_{i\in[m]} p_i H_i$ with $\sum_{i\in[m]} p_i = 1$, $p_i \geq 0$, and $0 \preceq H_i \preceq 1$ for all $i \in [m]$. Then, given an n-qubit quantum state $|\psi\rangle$, there exists a measurement on q qubits of $|\psi\rangle$ that outputs 1 with probability*

$$1 - \langle\psi| H |\psi\rangle. \tag{11}$$

Proof. This follows from a simple generalisation of Kitaev's original QMA verification protocol, which can be found in Chapter 14 of [3]. Let us now show the correctness of Protocol 1. For any i, let $|\psi\rangle = \sum_j \alpha_{i,j} |\lambda_{i,j}\rangle$ be the decomposition of $|\psi\rangle$ in the eigenbasis of H_i. The probability that this protocol accepts, conditioned on picking term i, is given by

$$\begin{aligned}
\Pr[\mathcal{V} \text{ accepts } |\psi\rangle \mid i] &= \|(\mathbb{I} \otimes |1\rangle\langle 1|) W_i |\psi\rangle |0\rangle\|_2^2 \\
&= \left(\sum_j \bar{\alpha}_{i,j} \langle \lambda_{i,j}| \langle 0| \right) W_i^\dagger (\mathbb{I} \otimes |1\rangle\langle 1|) W_i \left(\sum_j \alpha_{i,j} |\lambda_{i,j}\rangle |0\rangle \right) \\
&= \left(\sum_j \bar{\alpha}_{i,j} \sqrt{1-\lambda_{i,j}} \langle \lambda_{i,j}| \right) \left(\sum_j \alpha_{i,j} \sqrt{1-\lambda_{i,j}} |\lambda_{i,j}\rangle \right) \\
&= \sum_j (1-\lambda_{i,j}) \bar{\alpha}_{i,j} \alpha_{i,j} \\
&= 1 - \langle \psi| H_i |\psi\rangle .
\end{aligned}$$

The overall acceptance probability is then given by the expectation value over all choices of i, which is

$$\sum_{i \in [m]} p_i \Pr[\mathcal{V} \text{ accepts } |\psi\rangle \mid i] = 1 - \langle \psi| \sum_{i \in [m]} p_i H_i |\psi\rangle = 1 - \langle \psi| H |\psi\rangle .$$

□

Kitaev's energy estimation protocol (Protocol 1) can be viewed as a $(1, q)$-$\mathsf{QPCP}_{\mathrm{NA}}$ verifier, where the completeness and soundness bounds correspond to one minus the promised upper and lower bounds on the ground state energy in the YES- and NO-cases, respectively. When $q = \mathcal{O}(\log n)$, each W_i acts non-trivially only on $q+1$ qubits and can thus be implemented efficiently. By applying Lemma 8 in combination with weak error reduction (Lemma 2), we can correctly decide any local Hamiltonian problem with positive semidefinite terms and a constant promise gap using Kitaev's protocol, taking $p_i = 1/m$ for each $i \in [m]$.

Corollary 2. *For any constant $q \in \mathbb{N}$ and constant $\delta > 0$, we have that q-$\mathsf{LH}[\delta]$ is contained in $\mathsf{QPCP}_{\mathrm{NA}}[q']$ with $q' \in \mathcal{O}(1)$.*

Proof. Let a, b with $b - a = \delta$ be the completeness and soundness parameters of the q-$\mathsf{LH}[\delta]$ instance. By Lemma 8, we can transform this into another instance of $2q$-$\mathsf{LH}[\delta']$ with a $2q$-qubit Hamiltonian $H' = \frac{1}{m} \sum_{i \in [m]} H'_i$, $0 \preceq H'_i \preceq 1$, completeness a' and soundness b', with $b' - a' = \delta' > 0$ being some other constant that depends on q. We have that H' is in the desired form for Protocol 1, which has completeness $1 - a'$ and soundness $1 - b'$, and thus also promise gap δ'. Using weak error reduction from Lemma 2, we can then boost the promise gap back so that it satisfies completeness $2/3$ and soundness $1/3$. □

5 Applications

This section discusses some applications of the ideas presented earlier in the paper, all of which are derived (with varying degrees of overhead) by using the reduction as a subroutine. The section is divided into four subsections, each of which can be read independently to some extent.

5.1 Reduction to the average particle energy formulation

As a first application, we consider a specific formulation of the quantum PCP conjecture, stated in terms of an error constant relative to the number of sites (i.e., qubits or qudits), rather than the total number of terms. For this, we rely on the following lemma due to Tropp.

Lemma 10 ([40])**.** *Consider a finite sequence $\{X_k\}$ of independent, random, Hermitian matrices with dimension d, and let $\{A_k\}$ be a sequence of fixed Hermitian matrices. Assume that each random matrix satisfies*

$$\mathbb{E}[X_k] = 0 \quad \textit{and} \quad X_k^2 \preceq A_k^2 \quad \textit{almost surely.}$$

Then for any $t \geq 0$,

$$\Pr\left[\lambda_{\max}\left(\sum_k X_k\right) \geq t\right] \leq d \cdot e^{-t^2/8\sigma^2},$$

where

$$\sigma^2 := \left\|\sum_k A_k^2\right\|.$$

Here, $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue.

Proposition 1. *Consider a q-local Hamiltonian problem with Hamiltonian $H = \frac{1}{m}\sum_{i\in[m]} H_i$, $0 \preceq H_i \preceq 1$, with completeness and soundness parameters $a, b \geq 0$ such that $b - a = \gamma$ for some $\gamma = \Omega(1)$. Then for any $\delta = \Omega(2^{-n})$, there exists a randomized polynomial-time reduction to a $2q$-local Hamiltonian problem with Hamiltonian $G' = \frac{1}{l}\sum_{j\in[l]} G'_j$, $0 \preceq G'_j \preceq 1$, where $l = \mathcal{O}(n)$, with completeness and soundness parameters c, d such that $c - d \geq \gamma' = \Omega(1)$, which succeeds with probability $1 - \delta$.*

Proof. For any positive integer k, let X_k be a random variable defined as $H_i - H$ with probability $1/m$. Let $l \in \mathbb{N}$ be the length of the sequence $\{X_k\}_{k\in[l]}$, which is to be determined later. Clearly, for all $k \in [l]$,

$$\mathbb{E}[X_k] = \frac{1}{m}\sum_{i\in[m]} (H_i - H) = H - H = 0.$$

Since $0 \preceq H_i \preceq 1$ for all $i \in [m]$, it follows that $0 \preceq H \preceq 1$, and therefore $-1 \preceq X_k \preceq 1$, which implies $X_k^2 \preceq 1$. Now, for all $k \in [l]$, set $A_k = \mathbb{I}$, where $A_k = A_k^2$, so that $X_k^2 \preceq A_k^2$ holds. We then have

$$\sigma^2 := \left\|\sum_{k\in[l]} A_k^2\right\| = \|l\mathbb{I}\| = l.$$

Given the sequence $\{X_k\}_{k\in[l]}$, define $G_k = X_k + H$ and let $G = \frac{1}{l}\sum_{k\in[l]} G_k$. We can then express

$$\begin{aligned}\Pr\left[\lambda_{\max}\left(\frac{1}{l}\sum_{k\in[l]} X_k\right) \geq \frac{t}{l}\right] &= \Pr\left[\lambda_{\max}\left(H - \frac{1}{l}\sum_{k\in[l]} G_k\right) \geq \frac{t}{l}\right] \\ &= \Pr\left[\|H - G\| \geq \frac{t}{l}\right].\end{aligned}$$

By applying Lemma 10, we get

$$\Pr\left[\|H - G\| \geq \epsilon\right] \leq 2^n \cdot e^{-\epsilon^2 l/8}.$$

Now, set $\epsilon = \gamma/4$. In this case, if $\|H - G\| \leq \epsilon$, then it must hold that:

- if $\lambda_{\min}(H) \leq a$, then $\lambda_{\min}(G) \leq a' := a + \epsilon$;
- if $\lambda_{\min}(H) \geq b$, then $\lambda_{\min}(G) \geq b' := b - \epsilon$,

where $b' - a' \geq \gamma/2 = \Omega(1)$. To achieve a success probability of at least $1 - \delta$, we require

$$2^n \cdot e^{-\gamma^2 l/128} \leq \delta,$$

which implies a condition on l of

$$l \geq \frac{128}{\gamma^2}\left(n\ln(2) + \ln\left(\frac{1}{\delta}\right)\right).$$

For $\gamma = \Omega(1)$ and $\delta = \Omega(2^{-n})$, this yields $l = \Theta(n)$. Finally, we must ensure that each G_k satisfies $0 \preceq G_k \preceq 1$. The condition $0 \preceq G_k$ is trivially satisfied, but $G_k \preceq 1$ might not hold if we sample the same term H_i multiple times. To resolve this, we apply the deterministic transformation from Lemma 8 to obtain a Hamiltonian $G' = \frac{1}{l}\sum_{j\in[l]} G'_j$, $0 \preceq G'_j \preceq 1$ at most $2q$-local, which maintains the constant relative promise gap while increasing the locality by a factor of two. □

It is now straightforward to combine the reductions of Theorem 1, Lemma 8, and Proposition 1 to obtain the following corollary:

Corollary 3. *For any $q = \mathcal{O}(1)$, there exist a $q' = \mathcal{O}(q)$ and a constant $\delta > 0$ such that the problem q'-LH$[\delta]$, restricted to Hamiltonians with $m = \Theta(n)$ positive semidefinite terms, is* QPCP$[q]$*-hard under quantum reductions.*

5.2 Proof checking versus local Hamiltonian formulations of quantum PCP

The goal of this subsection is proving Theorem 2. We will utilize the fact that any QCMA-verifier would able to perform the reduction in Algorithm 1 (up to a polynomial number of bits of precision). This can be used to show that a QCMA upper bound on the local Hamiltonian problem with constant promise gap implies a QCMA upper bound on a quantum PCP system with the same locality.

Protocol 2: QCMA protocol for QPCP$[\mathcal{O}(1)]$ assuming that q-LH$[\Theta(1)] \in$ QCMA.

Input: A classical description of a $(1, q, p_1, p_2, p_3)$-QPCP verifier V_x with input x hardcoded into it, with completeness c and soundness s.

Set: $\epsilon := (c-s)/4$, $\Gamma' := (q+1)2p_2(n)$, $\eta := \left\lceil q\log\left(\frac{4\Gamma'(\Gamma'+1)}{\epsilon} - 1\right)\right\rceil$, $\delta := 1 - \sqrt{\frac{2}{3}}$, $a := c + \epsilon/4$ and $b := c - \epsilon/4$.

Protocol:

1. The prover sends the witness y.
2. The verifier performs the **variation** of Algorithm 1 with precision η and maximum error probability δ to obtain a q-local Hamiltonian $\tilde{H}_x = \sum_{i\in[m]} \tilde{H}_{x,i}$ up to η bits of precision.
3. It accepts if and only if the QCMA protocol, having completeness $\sqrt{\frac{2}{3}}$ and soundness $1 - \sqrt{\frac{2}{3}}$, with witness y for $(\hat{H}, a, b)$ accepts.

Theorem 2 (Local Hamiltonian versus proof verification)**.** *If the q-local Hamiltonian problem with a constant promise gap can be decided in* QCMA*, we have that*

$$\mathsf{QPCP}[q] \subseteq \mathsf{QCMA},$$

for all constant $q \in \mathbb{N}$.

Proof. Let $A = (A_{\text{yes}}, A_{\text{no}})$ be any promise problem in $\mathsf{QPCP}[q]$, x be an input, V_x the QPCP verifier with x hardcoded into it. Let $\tilde{H}_x$ be the corresponding local Hamiltonian obtained via the reductions in Step 2, conditioned on Step 2 succeeding. By Corollary 1, we have that in this case the same Hamiltonian is always produced with probability $\geq 1 - \delta = \sqrt{2/3}$, so the prover can provide the proof without knowing the outcome of the reduction. Moreover, we note that for our choice of parameters, Step 2 can be performed in quantum polynomial-time.

By assumption, the q-local Hamiltonian problem with constant promise gap is in QCMA. Hence, there exists a QCMA verifier Q such that:

- if $\lambda_0(H) \leq a$, then there exists $y \in \{0,1\}^{p(n)}$ such that

$$\Pr[Q \text{ accepts } ((H, a, b), y)] \geq \sqrt{\frac{2}{3}}.$$

- if $\lambda_0(H) \geq b$, then for all $y \in \{0,1\}^{p(n)}$;

$$\Pr[Q \text{ rejects } ((H, a, b), y)] \geq \sqrt{\frac{2}{3}},$$

since QCMA allows for strong error reduction. Now consider Protocol 2. Let Q' be the QCMA verifier that first perform the reduction from Step 2 to obtain $\tilde{H}_x$ to η bits of precision, and then runs $Q(\tilde{H}_x, a, b, |y\rangle)$. Since Step 2 succeeds with probability $1-\delta = \sqrt{\frac{2}{3}}$, we have:

- If $x \in A_{\text{YES}}$, there exists $y \in \{0,1\}^{p(n)}$ such that $\Pr[Q' \text{ accepts } (x, y)] \geq 2/3$.
- If $x \in A_{\text{NO}}$, then for all $y \in \{0,1\}^{p(n)}$, $\Pr[Q' \text{ accepts } (x, y)] \leq 1/3$,

for our choice of δ. This shows that $\mathsf{QPCP}[q] \subseteq \mathsf{QCMA}$. □

5.3 Adaptive versus non-adaptive quantum PCPs

Since adaptive quantum PCPs generalise non-adaptive quantum PCPs, we have that the inclusion $\mathsf{QPCP}_{\text{A}}[q] \supseteq \mathsf{QPCP}_{\text{NA}}[q]$ is immediate. In this subsection, we will prove that non-adaptive QPCPs can also simulate adaptive QPCPs with only constant overhead, by using our Lemma 8 and weak error reduction lemma (Lemma 2) for non-adaptive QPCPs.

Protocol 3: Non-adaptive simulation of an adaptive quantum PCP

Input: A classical description of a $(1, q, p_1, p_2, p_3)$-QPCP verifier V_x with input x hardcoded into it, with completeness c and soundness s.

Set: $\epsilon := (c-s)/4$, $\delta := 1 - \sqrt{\frac{2}{3}}$, $\Gamma' := (q+1)2p_2(n)$, $\eta := \left\lceil q \log\left(\frac{4\Gamma'(\Gamma'+1)}{\epsilon} - 1\right)\right\rceil$, and $R := \left\lceil 2\left(\frac{2^{q+4}}{c-s}\right)^2\right\rceil$.

Protocol:

1. The prover sends a quantum state $|\psi\rangle$.
2. The verifier runs the **variation** of Algorithm 1, with precision η and maximum error probability δ, to obtain a q-local Hamiltonian $\tilde{H}_x$.
3. The verifier applies the transformation of Lemma 8 to obtain a $2q$-local Hamiltonian $\hat{H}_x = \frac{1}{m}\sum_i \hat{H}_{x,i}$.
4. The verifier runs Protocol 1 R times for $\hat{H}_x$, and accepts if and only if at least a $\frac{(c-s)}{2^{q+4}}$-fraction of the outcomes equal $|1\rangle$.

Theorem 3 (Adaptive versus non-adaptive)**.** *For any $c - s = \Omega(1)$ and $q = \mathcal{O}(1)$, we have that*

$$\mathsf{QPCP}_{c,s}[q] \subseteq \mathsf{QPCP}_{\mathrm{NA}}[q']$$

with

$$q' = \mathcal{O}\left(q\left(\frac{4^q}{c-s}\right)^2\right).$$

Proof. We verify the correctness of Protocol 3, which defines a (q')-$\mathsf{QPCP}_{\mathrm{NA}}$ verifier for some q', which we will show to be $\mathcal{O}(q)$. Let $A = (A_{\mathrm{YES}}, A_{\mathrm{NO}})$ be a promise problem in $\mathsf{QPCP}_{\mathrm{A}}[q]$ for an arbitrary constant q, and let $x \in \{0,1\}^n$ be an input.

Correctness. Step 2 of the protocol produces a q-local Hamiltonian $\tilde{H}_x = \sum_{i\in[m]} \tilde{H}_i$ satisfying:

- If $x \in A_{\mathrm{YES}}$, then $\lambda_0(\tilde{H}_x) \leq s + \epsilon$,
- If $x \in A_{\mathrm{NO}}$, then $\lambda_0(\tilde{H}_x) \geq c - \epsilon$,

with probability at least $1-\delta$. After applying the transformation from Lemma 8 to obtain $\hat{H}_x$ from $\tilde{H}_x$, we get:

- If $x \in A_{\mathrm{YES}}$, then $\lambda_0(\hat{H}_x) \leq \frac{s+\epsilon}{2^{q+3}}$,
- If $x \in A_{\mathrm{NO}}$, then $\lambda_0(\hat{H}_x) \geq \frac{c-\epsilon}{2^{q+3}}$,

Applying Kitaev's energy estimation protocol (Protocol 1) to $\hat{Q}_x = \sum_i \frac{1}{m}\hat{H}_{x,i}$ yields a $(2q)$-$\mathsf{QPCP}_{\mathrm{NA}}$ verifier with a promise gap of at least

$$\gamma := \frac{(c-s)-2\epsilon}{2^{q+3}} = \frac{(c-s)}{2^{q+4}} = \Theta(1),$$

since q was constant and $c - s = \Omega(1)$. Since Step 4 performs R parallel runs of a $(2q)$-$\mathsf{QPCP}_{\mathrm{NA}}$ verifier, accepting if and only if a $\frac{c-s}{2^{q+4}}$-fraction of the verifiers accept, it forms an (Rq)-$\mathsf{QPCP}_{\mathrm{NA}}$ verifier with amplified completeness and soundness. By Lemma 2, and our choice of R (see Eq. (1) in the proof of Lemma 2), it follows that, conditioned on Step 2 succeeding, the acceptance (resp. rejection) probability in Step 4 is at least $\sqrt{2/3}$ in the YES-case (resp. NO-case). Moreover, we have that

$$q' = qR = q\left\lceil 2\left(\frac{2^{q+4}}{c-s}\right)^2\right\rceil = \mathcal{O}\left(q\left(\frac{4^q}{c-s}\right)^2\right).$$

Since Step 2 succeeds with probability at least $\sqrt{2/3}$, the overall completeness (soundness) is at least $2/3$ (at most $1/3$). □

We conclude this subsection with the following observation.

Remark 3. *The proof of Theorem 3 actually demonstrates that using a uniform distribution (over a subset of all proof qubits) to decide which proof qubits to access suffices. Indeed, Kitaev's energy estimation protocol can be applied with $p_i = 1/m$ for all $i \in [m]$, where m is the number of terms in the Hamiltonian. Since we still apply weak error reduction (Lemma 2) to boost the completeness and soundness parameters back to their original values, the resulting distribution over queries becomes uniform over a subset of all possible index combinations. This is because parallel repetition restricts us to combinations where each supposed copy is used only once. This resolves an open question posed in [20], and shows that the definition of a quantum PCP given in [21] is in fact fully general.*

5.4 A multi-prover quantum PCP for QMA[2] implies QMA[2] = QMA

In this section we prove Theorem 5. Of course, if QMA[2] would allow a single-prover quantum PCP, i.e., $\mathsf{QMA}[2] \subseteq \mathsf{QPCP}[1, \mathcal{O}(1)]$, then we would trivially have that $\mathsf{QMA}[2] = \mathsf{QMA}$ as $\mathsf{QPCP}[1, \mathcal{O}(1)] \subseteq \mathsf{QMA}$. The point will be that for any polynomial number of unentangled provers, each providing a polynomially-sized proof, a quantum PCP that checks a constant number of qubits across the tensor product of these proofs can be simulated in QMA.

The proof of this result is based on the ideas in [31] where it is proven that the 2-separable local Hamiltonian problem is in QMA. Since we need a slight generalisation of their result (namely from 2-separable to k-separable), we include a full proof for completeness. The proof relies on the consistency of local density matrices problem (CLDM), which will be introduced first. Moreover, for completeness and future reference, we give a new proof of QMA-containment which holds for a larger range of problem parameter settings than previous proofs [32, 33].

5.4.1 The consistency of local density matrices problem

We start by stating the definition of the consistency of local density matrices problem, first defined in [32].

Definition 6 (Consistency of local density matrices, CLDM(q, α, β))**.**

Input: *A classical description of a collection of local density matrices* $\{\rho_i\}_{i\in[m]}$ *on* n *qubits,* $m = \text{poly}(n)$*, where each* ρ_i *is a density matrix over qubits* $C_i \subseteq [n]$ *with* $|C_i| \leq q$*. For each* $i \in [m]$*, write* $\overline{C}_i = [n] \setminus C_i$ *for the complementary subset. Additionally, we are given two efficiently computable real numbers* α, β *such that* $\beta - \alpha > 0$*.*

Promise: *One of the following two cases holds:*

(i) There exists an n*-qubit mixed state* σ *such that for all* $i \in [m]$*,*

$$\left\| \text{tr}_{\overline{C}_i}[\sigma] - \rho_i \right\|_1 \leq \alpha.$$

(ii) For all n*-qubit mixed states* σ*, there exists an* $i \in [m]$ *such that*

$$\left\| \text{tr}_{\overline{C}_i}[\sigma] - \rho_i \right\|_1 \geq \beta.$$

Output: YES *if (i) holds, and* NO *if (ii) holds.*

If $\alpha = 0$*, we write* CLDM(q, β)*.*

Liu showed that CLDM is in QMA for $\beta/4^q =: \epsilon = \Omega(1/\text{poly}(n))$ and $\alpha = 0$. Broadbent and Grilo [33] provided a proof for arbitrary α (and $\epsilon = \beta - \alpha$), but their proof contains a small error.[9] Liu's proof can easily be modified to also incorporate the case for general α, however, this still leads to a condition on β which depends on q, leaving open as to whether CLDM(q, β, α) is in QMA for any $\beta - \alpha = \Omega(1/\text{poly}(n))$.[10]

We will now demonstrate that a different proof technique can be used to lift this restriction on β. Though not strictly needed to obtain our results, we include it for completeness and future reference. Our protocol is based on two ideas: (i) with a sufficient number of copies, one can learn the local marginals of any given state, and (ii) a specific formulation of a quantum de Finetti theorem under local measurements.

5.4.2 Full state tomography of marginals

The task of full state tomography is, given access to copies of an unknown quantum state ρ, to learn an ϵ-approximation $\tilde{\rho}$ (with respect to some distance measure on quantum states), while minimising the number of copies required and, potentially, also the total processing time. Since we only care about overall efficiency, we will use the most basic state tomography protocol, as it allows for the simplest analysis. For an n-qubit system, a *Pauli word* (also called a *Pauli string*) is any operator of the form $P = \sigma_1 \otimes \sigma_2 \otimes \cdots \otimes \sigma_n$, where each $\sigma_i \in \{\mathbb{I}_2, X, Y, Z\}$ is a Pauli operator. For a Pauli word P_j, write M_j for the measurement $M_j = \{(P_j + \mathbb{I})/2, (\mathbb{I} - P_j)/2\}$ with corresponding outcomes $\{+1, -1\}$. Hence, given a state ρ, we have that the random variable $X_j \in \{+1, -1\}$ corresponding to the measurement of ρ using M_j, satisfies $\mathbb{E}[X_j] = \text{tr}\left[\frac{1}{2}(P_j + \mathbb{I})\rho\right](+1) + \text{tr}\left[\frac{1}{2}(\mathbb{I} - P_j)\rho\right](-1) = \text{tr}[P_j\rho]$. Since the Pauli words $\{P_j\}_{j\in[d^2]}$ form a basis for the space of d-dimensional Hermitian matrices, we can write any density operator $\rho = \sum_{j\in[d^2]} c_j P_j$ where $c_j = \text{tr}[P_j\rho]$. If instead we have estimates $\tilde{c}_j$ such that $|\tilde{c}_j - c_j| \leq \epsilon$, then we have that $\tilde{\rho} = \sum_{j\in[d^2]} \tilde{c}_j P_j$ satisfies $\|\tilde{\rho} - \rho\|_1 \leq d^2\epsilon$.

[9] The proof relies on Hoeffding's inequality to show soundness. However, the random variables it is used on are generally not independent in this setting, since the proof can be highly entangled.

[10] The initial arXiv version of the present paper contains this modified proof.

Lemma 11. *Let $\rho \in \mathcal{D}(\mathbb{C}^{2^n})$ and let $\{C_i\}_{i\in[m]}$ be a collection of subsets of qubit indices, each satisfying $|C_i| \leq q$. Write $\rho_i = \mathrm{tr}_{\overline{C}_i}[\rho]$. Then, there exists a measurement $M = \left\{ M_{a_1}^{(1)} \otimes \cdots \otimes M_{a_l}^{(l)} \;\middle|\; a \in \{\pm 1\}^l \right\}$ on the state $\rho^{\otimes l}$ and a classical algorithm running in $\mathrm{poly}(l)$ time that, given the measurement outcomes, outputs a classical description of $\{\tilde{\rho}_i\}_{i\in[m]}$ satisfying*

$$\|\tilde{\rho}_i - \rho_i\|_1 \leq \epsilon \quad \textit{for all } i \in [m],$$

with probability at least $1-\delta$. This uses

$$l = \mathcal{O}\left(mq16^q \log(m/\delta)/\epsilon^2\right)$$

copies of ρ.

Proof. The measurement consists of applying a tensor product of different Pauli measurements M_j across the many copies of ρ to obtain estimates $\tilde{c}_{j,i} \approx \mathrm{tr}[P_j\sigma_i]$, from which all the marginals ρ_i can be approximately reconstructed. Since each measurement only acts on a single density matrix of a single copy of ρ, this ensures the tensor product structure of the overall measurement. Let $d_i = 2^{|C_i|}$, so that $\rho_i = \mathrm{tr}_{\overline{C}_i}[\rho]$ is the d_i-dimensional density matrix corresponding to the reduced density matrix of ρ that has indices from C_i. We have $\rho_i = \sum_{j\in[d_i^2]} c_{j,i}P_j$, where $c_{j,i} = \mathrm{tr}[P_j\rho_i]$. Using the measurement M_j corresponding to P_j, where each measurement outcome is bounded, standard mean estimation (see for example [41]) gives an estimate $\tilde{c}_{j,i}$ such that $|\tilde{c}_{j,i} - c_{j,i}| \leq \epsilon/d_i^2$ with probability $1-\delta/(md_i^2)$, using $\mathcal{O}(d_i^4 \log(md_i^2/\delta)/\epsilon^2)$ copies of ρ. We set $\tilde{\rho}_i = \sum_{j\in[d_i^2]} \tilde{c}_{j,i}P_j$. By a union bound and converting a bound on the max norm to the trace norm, we must have that for each $i \in [m]$, $\|\tilde{\rho}_i - \rho_i\|_1 \leq \epsilon$ holds with probability at least $1-\delta/m$ for each i. Another union bound shows that the probability of $\|\tilde{\rho}_i - \rho_i\|_1 \leq \epsilon$ holding for all $i \in [m]$ simultaneously is at least $1-\delta$. Using the upper bound $d_i \leq 2^q$ for all $i \in [m]$, the total number of copies (and thus measurements in the tensor product) can be upper bounded as

$$\mathcal{O}\left(mq16^q \log(m/\delta)/\epsilon^2\right).$$

As the classical post-processing consists primarily of the addition of all obtained measurement outcomes, it clearly can be done in time $\mathrm{poly}(l)$. □

5.4.3 Quantum de Finetti under local measurements

We first need to introduce some additional notation. For bipartite states ρ^{XY}, we use the convention that omitting subscripts corresponds to taking the partial trace over those systems; for example, $\rho^X = \mathrm{tr}_Y[\rho^{XY}]$. We say that $\rho^{A_1\ldots A_k}$ is permutation symmetric if $\rho^{A_{\pi(1)}\ldots A_{\pi(k)}} = \rho^{A_1\ldots A_k}$ for any permutation $\pi \in \mathrm{S}_k$ (with S_k is the symmetric group of order k). We associate to any POVM $\{M_k\}$ a map $\Lambda(X) = \sum_k \mathrm{Tr}(M_kX)\,|k\rangle\langle k|$, where $\{|k\rangle\}$ is an orthonormal basis. Thus, the $\Lambda(X)$ are so-called quantum-classical channels, as they map density operators to other density matrices that are diagonal in the basis $\{|k\rangle\}$. This implies that for two states ρ and σ, we have

$$\frac{1}{2}\|\Lambda(\rho-\sigma)\|_1 = D_{\mathrm{TV}}\left(\{p_k\},\{q_k\}\right)$$

where $p_k = \mathrm{tr}\{M_k\rho\}$, $q_k = \mathrm{tr}\{M_k\sigma\}$ and $D_{\mathrm{TV}}(\cdot,\cdot)$ is the total variation distance.

Informally, the classical de Finetti Theorem states that if the joint probability distribution of a sequence of random variables is invariant under any permutation of the variables, then the marginal probability distribution of a subset of $l \ll k$ variables from such a k-variable sequence will be close to a convex combination of i.i.d. variables [42]. Quantum versions of the de Finetti Theorem posit that an l-partite quantum state $\rho^{A_1 \dots A_l}$, which is the reduced state of a permutation-symmetric state $\rho^{A_1 \dots A_k}$ on $k \gg l$ subsystems, is close to a convex combination of i.i.d. quantum states, i.e., $\rho^{A_1 \dots A_l} \approx \int d\mu(\sigma)\, \sigma^{\otimes l}$, with μ a probability measure on quantum states. Different quantum de Finetti theorems consider different notions of "closeness", e.g., [43], [44], and [45]. We will focus on closeness with respect to local measurements performed on the subsystems, for which a quantum de Finetti theorem was proven by Brandão and Harrow [46].

Lemma 12 ([46])**.** *Let $\rho^{A_1 \dots A_k} \in \mathcal{D}(A^{\otimes k})$ be a permutation-invariant state. Then for every $0 \leq l \leq k$, there is a measure ν (that depends on ρ) on $\mathcal{D}(A)$ such that*

$$\max_{\Lambda_2, \dots, \Lambda_l} \left\| (\mathbb{I} \otimes \Lambda_2 \otimes \cdots \otimes \Lambda_l) \left(\rho^{A_1 \dots A_l} - \int \nu(d\sigma) \sigma^{\otimes l} \right) \right\|_1 \leq \sqrt{\frac{2l^2 \ln |A|}{k - l}}.$$

Given any input state on k registers, the permutation-invariant assumption can always be enforced by randomly permuting all the input registers (this does of course alter the input state if the state was not already permutation-invariant).

5.4.4 The quantum marginal problem is in QMA

We now have all ingredients to give our QMA protocol for CLDM(q, α, β), which is given in Protocol 4.

The core idea is that, with enough copies of a state, the verifier can estimate its local marginals via tomography. Then, using Lemma 12, we argue that any permutation-invariant state (which includes the verifier's post-processed state) is close to a separable state. Hence, from the verifier's perspective, the tomography is effectively performed on a state that is nearly separable, even if the prover were to send a highly entangled state. From this, we can show that the acceptance probability of the protocol is very close to that of an idealised version in which the prover sends a mixture of actual tensor-product copies, from which soundness follows.

Protocol 4: QMA protocol for CLDM(q, α, β)

Input: Classical descriptions of the density matrices $\{\rho_i\}_{i\in[m]}$ and the indices $\{C_i\}_{i\in[m]}$ of the qubits on which they are supported, problem parameters q, β and α.

Set: $\epsilon \coloneqq (\beta-\alpha)/4$, $\delta \coloneqq 1/6$, $l \coloneqq \mathcal{O}\left(mq16^q \log(m/\delta)/\epsilon^2\right)$,
$k \coloneqq \frac{2l\delta^2+l^2 n \ln 2}{2\delta^2}$.

Protocol:

1. The prover sends a state $\hat{\rho}^{A_1\dots A_k} \in \mathcal{D}(A^{\otimes k})$.
2. The verifier randomly permutes the index labels and traces out the last $k-l$ registers, creating the state $\rho^{A_1\dots A_l}$.
3. The verifier performs the measurement of Lemma 11, with desired precision ϵ and success probability $1-\delta$, on the registers $A_1,\dots,A_l$ and uses the measurement outcome to construct $\{\tilde{\rho}_i\}_{i\in[m]}$.
4. Accept if $\|\tilde{\rho}_i - \rho_i\|_1 \leq \alpha+\epsilon$ for all $i\in[m]$, and reject otherwise.

Theorem 4. CLDM(q,α,β) *is in* QMA *for any* $q=\mathcal{O}(\log n)$ *and* $\beta-\alpha=\Omega(1/\mathrm{poly}(n))$.

Proof. We will show the correctness of Protocol 4.

Completeness. In this case, the prover sends the state $\sigma^{\otimes k}$, which is already permutation-invariant. After Step 2, the resulting state is $\sigma^{\otimes l}$. Hence, we have access to l perfect copies of σ, meaning that by Lemma 11, the estimates $\tilde{\sigma}_i$ of all m density matrices σ_i will be retrieved up to precision ϵ with high probability. Since we are in a YES-instance, this means that, conditioned on the estimation procedure succeeding, we have $\|\tilde{\sigma}_i-\rho_i\|_1 \leq \alpha+\epsilon$ for all $i\in[m]$, so we will accept with probability 1. Thus, the overall success probability is lower bounded by the success probability of the estimation procedure, which is $\geq 1-\delta \geq 2/3$ for our choice of δ.

Soundness. Let the prover send any arbitrary state $\hat{\rho}^{A_1\dots A_k}$. Step 2 includes a random permutation of the subsystems, so the *expected* state on the remaining l systems (after tracing out $k-l$ registers) is the symmetrised marginal

$$\frac{1}{k!}\sum_{\pi\in S_k} \mathrm{tr}_{A_{k-l+1}\dots A_k}\left[\pi\hat{\rho}^{A_1\dots A_k}\pi^\dagger\right].$$

Since the acceptance probability is linear in the input state, we may analyse it as though the prover had sent a symmetric state from the start. Thus, from here onward, we assume without loss of generality that $\rho^{A_1\dots A_l}$ is permutation-invariant.

Let $P(\rho^{A_1\dots A_l})$ be the probability that the overall protocol in Protocol 4 accepts. We will first argue that if the prover provides a tensor product of multiple copies of some state

we will reject with high probability. Next, we will argue that by Lemma 12, Step 2 ensures that any arbitrary (even entangled) state must be close to such a state, and thus will also be rejected with high probability. We now formalise this.

Just as in the completeness case, suppose the state after Step 2 is of the form

$$\rho^{A_1\dots A_l} = \sigma^{\otimes l}$$

for some arbitrary $\sigma \in \mathcal{D}(A)$. Conditioned on the success of Step 3, which occurs with probability $1-\delta$, we will find a set $\{\tilde{\sigma}_i\}_{i\in[m]}$ such that $\left\|\tilde{\sigma}_i - \mathrm{tr}_{\overline{C}_i}\,\sigma\right\|_1 \leq \epsilon$ for all $i \in [m]$. However, the promise implies that for any such σ, there must be some $i \in [m]$ such that $\|\tilde{\sigma}_i - \rho_i\|_1 \geq \beta - \epsilon > \alpha + \epsilon$. Hence, we have $P(\sigma^{\otimes l}) \leq \delta$. By convexity, the same argument implies that $P\left(\int \nu(d\sigma)\sigma^{\otimes l}\right) \leq \delta$ for any measure ν on $\mathcal{D}(A)$.

Recall that the $\Lambda_j(\cdot)$ are quantum-classical channels, mapping quantum states to discrete probability distributions, where each channel implements a single measurement $M^{(j)} = \{M^{(j)}_{+1}, M^{(j)}_{-1}\}$ corresponding to the one in Lemma 11 (note that the index j here does not represent the Pauli word, but labels the measurement instead). Step 4 can potentially distinguish between the two probability distributions described by the following two density matrices, which are diagonal in some orthonormal basis:

$$\rho_1 \coloneqq (\Lambda_1 \otimes \cdots \otimes \Lambda_l)\left(\rho^{A_1\dots A_l}\right),$$

and

$$\rho_2 \coloneqq (\Lambda_1 \otimes \cdots \otimes \Lambda_l)\left(\int \nu(d\sigma)\sigma^{\otimes l}\right),$$

for some measure ν as per Lemma 12. Write p_1, p_2 for the distributions associated with measuring ρ_1, ρ_2 in their eigenbasis, respectively. By Lemma 12, and using the fact that the total variation distance upper bounds the maximum bias in the output by a single-sample distinguisher, we have

$$\left|P(\rho^{A_1\dots A_l}) - P\left(\int \nu(d\sigma)\sigma^{\otimes l}\right)\right| \leq D_{\mathrm{TV}}(p_1, p_2),$$

where

$$\begin{aligned}
D_{\mathrm{TV}}(p_1,p_2) &= D(\rho_1,\rho_2)\\
&= \frac{1}{2}\left\|(\Lambda_1\otimes\cdots\otimes\Lambda_l)\left(\rho^{A_1\dots A_l} - \int \nu(d\sigma)\sigma^{\otimes l}\right)\right\|_1\\
&\leq \max_{\Lambda_2',\dots,\Lambda_l'}\left\|(\mathbb{I}\otimes\Lambda_2'\otimes\cdots\otimes\Lambda_l')\left(\rho^{A_1\dots A_l} - \int \nu(d\sigma)\sigma^{\otimes l}\right)\right\|_1\\
&\leq \frac{1}{2}\sqrt{\frac{2l^2\ln|A|}{k-l}}\\
&\leq \delta,
\end{aligned}$$

for our choice of parameters in Protocol 4. Here we also used that the trace distance is non-increasing under the application of a quantum channel. Thus, for our choice of δ, we have $P(\rho^{A_1\dots A_l}) \leq 2\delta = 1/3$, as desired.

Complexity. Step 2 clearly takes time polynomial in n when $k = \mathrm{poly}(n)$. Step 3 runs in polynomial time when l is polynomial in n, and Step 4 runs in polynomial time when the sizes of the density matrices are at most polynomial in n and $m = \mathrm{poly}(n)$. All conditions hold when $q = \mathcal{O}(\log n)$, $\beta - \alpha \geq 1/\mathrm{poly}(n)$, and $m = \mathrm{poly}(n)$. □

5.4.5 Simulating a multi-prover constant-query quantum PCP in QMA

Before we prove the main result of this subsection, we need to show the following simple property of QMA, which shows that one can compute the AND-function of multiple promise problems in QMA with a single QMA-verifier.

Lemma 13. *Let $1 \leq l \leq \text{poly}(n)$. Suppose that $A_1, A_2, \ldots, A_l$ are promise problems in* QMA. *Then the promise problem $B = (B_{\text{YES}}, B_{\text{NO}})$ defined as*

- $x = (x_1, \ldots, x_l) \in B_{\text{YES}}$, *if $x_i \in A_{i,\text{YES}}$ for all $i \in [l]$;*
- $x = (x_1, \ldots, x_l) \in B_{\text{NO}}$, *if there exists an $j \in [l]$ such that $x_j \in A_{j,\text{NO}}$, with $x_i \in \{A_{i,\text{YES}}, A_{i,\text{NO}}\}$ for all $i \in [l]$;*

is in QMA.

Proof. Since $A_i \in$ QMA for all $i \in [l]$, we have that for each A_i, and for every polynomial $p_2(n) \geq 1$, there exists a uniform family of quantum circuits $\{U_n^i \mid n \in \mathbb{N}\}$ such that:

1. If $x_i \in A_{i,\text{yes}}$, then there exists a proof $|\psi_i\rangle$ such that $\Pr[U_n^i \text{ accepts } |\psi_i\rangle] \geq 1 - 2^{-p_2(n)}$.
2. If $x_i \in A_{i,\text{no}}$, then for all quantum proofs $|\psi_i\rangle$, we have $\Pr[U_n^i \text{ accepts } |\psi_i\rangle] \leq 2^{-p_2(n)}$,

by standard error reduction for QMA. We will set p_2 later. We now define the verifier U_n as follows: it expects a quantum proof $\bigotimes_{i\in[l]} |\psi_i\rangle$, runs all U_n^i in parallel, measures all l designated output qubits in the computational basis, and accepts if and only if all measurement outcomes are $|1\rangle$.

Case (i): If $x_i \in A_{i,\text{yes}}$ for all $i \in [l]$, then there exists a state $\bigotimes_{i\in[l]} |\psi_i\rangle$ such that

$$\Pr\left[U_n \text{ accepts } \otimes_{i\in[l]} |\psi_i\rangle\right] \geq (1 - 2^{-p_2(n)})^l \geq 1 - 2^{-p_2(n)} l \geq 1 - 2^{-p(n)}$$

whenever $p_2(n) \geq p(n) + \log(l)$.

Case (ii): Suppose there exists $j \in [l]$ such that $x_j \in A_{j,\text{no}}$. We must argue that it does not help the prover to provide a highly entangled state. This follows from the same reasoning as in the proof that QMA admits weak error reduction. Let γ be the total quantum proof and C_i, $i \in [l]$ be the index sets corresponding to the proof qubits to be used for input x_i. For all $i < j$, suppose the verifier has already executed the subprotocols and all of them accepted (otherwise we are done). Let the resulting density matrix on qubits from C_j of the remaining state be γ'_{C_j}. By convexity of acceptance probability, it follows that all mixed states γ'_{C_j} have acceptance probability at most $2^{-p_2(n)}$ for U_n^j, since this holds for all pure states. Hence, the acceptance probability at step j is at most $2^{-p_2(n)}$, regardless of previous outcomes. Since we set $p_2(n) \geq p(n)$ for case (i), this proves case (ii).

Since $1 \leq l \leq \text{poly}(n)$, we can just set $p_2(n) \coloneqq p(n) + l$ so that both cases are satisfied Thus, B is in QMA. □

Next we show that the k-separable q-local Hamiltonian problem (Definition Definition 3) is in QMA. This relies directly on the QMA containment of the CLDM problem. We first state a QMA protocol.

Protocol 5: QMA protocol for the k-separable q-local Hamiltonian problem

Input: Classical descriptions of m q-local terms $\{H_i\}$, completeness and soundness parameters b, a, and a number k.

Set $a' := a + \frac{b-a}{4}$, $b' := b - \frac{b-a}{4}$, $\beta := \frac{b-a}{2qm}$, $\alpha := \beta/8^q$, and $\delta := \frac{k}{3}$.

1. The prover sends a quantum witness γ and a classical witness $\{\rho_i^j\}$.
2. The verifier performs the following three checks:
   - **Check 1:** It checks that each reduced density matrix ρ_i^j is positive semidefinite and has trace one.
   - **Check 2:** It checks that $\mathrm{tr}\left[H \otimes_{j\in[k]} \rho_i^j\right] \leq a'$.
   - **Check 3:** It splits up the indices of the quantum witness γ in k disjoint sets C_j of equal size, $j \in [k]$. It runs k parallel executions of the QMA protocol given by Protocol 4, each with completeness $1-\delta$ and soundness δ, on the k instances of the CLDM(q, α, β) with respective input $\{\rho_i^j\}$ and proof $\mathrm{tr}_{\bar{C}_j}[\gamma]$. The check is passed if and only if all CLDM(q, α, β) verifications accept.
3. The verifier accepts if and only if all three checks are passed.

Lemma 14. *The k-separable q-local Hamiltonian problem (*LH(k, q, a, b)*) is in* QMA *for any $1 \leq k \leq \mathrm{poly}(n)$, $q = \mathcal{O}(\log n)$ and $b - a \geq 1/\mathrm{poly}(n)$.*

Proof. This follows the same ideas as in [31], but now for a k- instead of 2-separable state.

First, we observe that Check 3 defines the logical AND of multiple promise problems in QMA, so by the proof of Lemma 13, it follows that for our choice of δ, the overall procedure defined by Check 3 has completeness at least 2/3 and soundness at most 1/3.

Completeness. By the promise, there exists a state $\xi = \otimes_{j\in[k]}\xi^j$ such that $\mathrm{tr}[H\xi] \leq a$. The prover sends as a quantum witness $\gamma = \bigotimes_{j\in[k]} \gamma^j$, where each γ^j is a witness for the CLDM instance corresponding to ξ^j. For the classical witness, the prover sends descriptions of $\{\tilde{\xi}_i^j\}$ of the q-local reduced density matrices ξ_i^j of ξ, each specified up to trace distance $\epsilon = 2^{-p(n)}$ for some large polynomial $p(n)$, such that the second and third checks are satisfied (see below). This is possible, as the entries of each density matrix can be described using a polynomial number of bits. In this case, the protocol proceeds as follows:

- The first check of Protocol 5 is passed directly.
- For the second check, we observe that
$$\sum_{i\in[m]} \mathrm{tr}\left[H_i \otimes_{j\in[k]} \tilde{\xi}_i^j\right] \leq a + mk\epsilon \leq a',$$
provided ϵ is chosen sufficiently small, which means that it also passes.
- For Check 3, by Lemma 13, this check also succeeds with probability at least 2/3. Hence, the overall acceptance probability is at least 2/3.

Soundness. If Check 1 or Check 2 fails, we are done. If Check 1 succeeds, we can be certain that each matrix ρ_i^j is a density matrix up to an exponentially small correction. If Check 2 succeeds, we must have that

$$\sum_{i\in[m]} \mathrm{tr}\Big[H_i \otimes_{j\in[k]} \rho_i^j\Big] \leq a'.$$

According to the promise, we have that for any state $\xi = \otimes_{j\in[k]}\xi^j$

$$\sum_{i\in[m]} \mathrm{tr}\Big[H_i \otimes_{j\in[k]} \xi_i^j\Big] \geq b'.$$

This means that for our choice of a', b' we have

$$\begin{aligned}\frac{b-a}{2} &\leq \sum_{i\in[m]} \Big(\mathrm{tr}\Big[H_i \otimes_{j\in[k]} \xi_i^j\Big] - \mathrm{tr}\Big[H_i \otimes_{j\in[k]} \rho_i^j\Big]\Big) \\ &= \sum_{i\in[m]} \mathrm{tr}\Big[H_i \left(\otimes_{j\in[k]}\xi_i^j - \otimes_{j\in[k]}\rho_i^j\right)\Big] \\ &\leq \sum_{i\in[m]} \left\|\otimes_{j\in[k]}\xi_i^j] - \otimes_{j\in[k]}\rho_i^j\right\|_1\end{aligned}$$

which implies that there must exist an i such that $\frac{b-a}{2m} \leq \left\|\otimes_{j\in[k]}\xi_i^j] - \otimes_{j\in[k]}\rho_i^j\right\|_1$. In the worst case, we have that $k \geq q$ and the qubits comprising this density matrix may be distributed across q distinct proof registers. However, by the subadditivity property of the trace distance with respect to tensor products, we have that for any subset $S \subseteq [k]$ with $|S| \leq q$:

$$\left\|\otimes_{j\in S}\xi_i^j - \otimes_{j\in S}\rho_i^j\right\|_1 \leq \sum_{j\in S} \left\|\xi_i^j - \rho_i^j\right\|_1,$$

which implies that there must exist a i, j pair such that $\beta := \frac{b-a}{2qm} \leq \left\|\xi_i^j - \rho_i^j\right\|_1$, which satisfies the promise of a NO-instance for CLDM(q, α, β). As we have already argued, Step 3 has a success probability of at least $2/3$ for detecting a single NO-instance, by the proof of Lemma 13. Hence, we have that the overall acceptance probability is at most $1/3$. □

Finally, we arrive at the main result for this section. We will prove it by arguing that QPCP$[k, q]$ is contained in QMA for any $k = \mathrm{poly}(n)$.

Protocol 6: QMA protocol for QPCP$[k, q]$.

Input: A classical description of a (k, q, p_1, p_2, p_3)-QPCP verifier V_x with input x hardcoded into it, with completeness c and soundness s.

Set: $\epsilon \coloneqq (c-s)/4$, $\delta = \delta' \coloneqq 1 - \sqrt{\frac{2}{3}}$, $\Gamma' \coloneqq (q+1)2kp_2(n)$, $\eta \coloneqq \left\lceil q \log\left(\frac{4\Gamma'(\Gamma'+1)}{\epsilon} - 1\right)\right\rceil$.

Protocol:

1. The prover sends the witness ξ.
2. The verifier runs the **variation** of Algorithm 1, with precision η and maximum error probability δ, to obtain a q-local Hamiltonian $\tilde{H}_x$.
3. The verifier runs Protocol 5 for Hamiltonian $\tilde{H}_x$ with completeness $1-\delta'$ and soundness δ', with $a = s + \epsilon$ and $b = c - \epsilon$, and accepts if and only if Protocol 5 accepts.

Theorem 5 (PCPs for QMA[2])**.** *If there exists a* $2 \leq k' \leq \mathrm{poly}(n)$ *and* $q = \mathcal{O}(1)$ *such that* QMA[2] $\subseteq$ QPCP$[k', q]$*, then* QMA[2] = QMA*.*

Proof. Suppose that such a k' and q indeed exist. Let $A = (A_{\mathrm{YES}}, A_{\mathrm{NO}})$ be any problem in QPCP$[k', q]$, with verifier V and input x. We have already verified in the proof of Theorem 3 that Step 2 of Protocol 6 can be made to with probability at least $1-\delta' \geq \sqrt{2/3}$, producing a fixed Hamiltonian when succeeding, and Step 3 can be made to succeed with probability at least $1 - \delta \geq \sqrt{2/3}$ by standard error reduction for QMA (the arguments holds for any $1 \leq k \leq \mathrm{poly}(n)$). Thus, we have:

- If $x \in A_{\mathrm{YES}}$, then Pr[Protocol 6 accepts] $\geq 2/3$,
- If $x \in A_{\mathrm{NO}}$, then Pr[Protocol 6 accepts] $\leq 1/3$,

which implies $A \in$ QMA, and hence QPCP$[k', q] \subseteq$ QMA. Hence, since the assumption implies QMA[2] $\subseteq$ QPCP$[k', q] \subseteq$ QMA and QMA $\subseteq$ QMA[2] holds trivially, the result follows. □

6 Proofs of quantum PCP theorems do not (quantumly) relativize

6.1 Quantum proof discretisation

We start by recalling the definition of h-nets on pure states.

Definition 7 (h-net)**.** *Let* $\mathcal{H}$ *be a* d*-dimensional Hilbert space. We say a set* $P_h^d = \{|\psi_i\rangle\} \subseteq \mathcal{P}(\mathcal{H})$ *forms an* h*-net for* $\mathcal{P}(\mathcal{H})$ *if for all* $|\phi\rangle \in \mathcal{P}(\mathcal{H})$ *there exists a* $|\psi_i\rangle \in P_h^d$ *such that* $|\langle\phi|\psi_i\rangle| \geq h$*.*

It is known that for all h there exists such a h-net of size that scales roughly exponentially in the dimension, as illustrated by the following lemma.

Lemma 15 ([47])**.** *Let $\mathcal{H}$ be a d-dimensional Hilbert space. For all $0 < h < 1$, there exists a h-net P_h^d for $\mathcal{P}(\mathcal{H})$ of size smaller than*

$$C' \left(\frac{d^{3/2} \log(2 + dh^2)}{(1 - h^2)^d} \right),$$

for some universal constant $C' > 0$.

To the best of our knowledge, no explicit construction has yet been found that achieves this scaling in d.

We give an analogous definition of h-nets for mixed states with respect to trace distance, which we call *ϵ-covering sets of density matrices*.

Definition 8 (ϵ-covering set of density matrices)**.** *Let $\mathcal{H}$ be some d-dimensional Hilbert space. We say a set of density matrices $D_\epsilon^d = \{\rho_i\} \subseteq \mathcal{D}(\mathcal{H})$ is ϵ-covering for $\mathcal{D}(\mathcal{H})$ if for all $\sigma \in \mathcal{D}(\mathcal{H})$ there exists a $\rho \in D_\epsilon^d$ such that $\frac{1}{2}\|\rho - \sigma\|_1 \leq \epsilon$.*

It is easy to show that the existence of an h-net implies the existence of a ϵ-covering set of density matrices with $\epsilon = \sqrt{1 - h^2}$.

Proposition 2. *For all $0 < \epsilon < 1$, there exists a ϵ-covering set of density matrices D_ϵ^d of size smaller than*

$$C \left(\frac{1}{\epsilon} \right)^{5d},$$

for some universal constant $C > 0$.

Proof. One can view the h-net P_ϵ^{2d} of Lemma 15 on a $2d$-dimensional Hilbert space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$, both of equal dimension d, as containing purifications of density matrices from the set $D_\epsilon^d \subseteq \mathcal{D}(\mathcal{H}_A)$. Let ρ be some arbitrary mixed state in $\mathcal{D}(\mathcal{H})$ and let $|\psi\rangle \in \mathcal{P}(\mathcal{H} \otimes \mathcal{H})$ be a purification of ρ. Then there must exists a state $|\phi\rangle \in P_\epsilon^{2d}$, with reduced density matrix σ, such that

$$|\langle\phi|\psi\rangle| \geq h \iff 1 - |\langle\phi|\psi\rangle|^2 \leq 1 - h^2,$$

which means that setting $h := \sqrt{1 - \epsilon^2}$ should have that

$$\epsilon \geq \sqrt{1 - |\langle\phi|\psi\rangle|^2} = \frac{1}{2} \| |\psi\rangle\langle\psi| - |\phi\rangle\langle\phi| \|_1 \geq \frac{1}{2} \|\rho - \sigma\|_1,$$

since the trace distance can only decrease under partial trace. Hence, setting

$$D_\epsilon^d := \{ \mathrm{tr}_B[|\phi\rangle\langle\phi|] : |\phi\rangle \in P_\epsilon^{2d} \}$$

satisfies conditions of Definition 8. We can now upper bound the cardinality of D_ϵ^d using Lemma 15 as

$$\begin{aligned} |D_\epsilon^d| &\leq C' \left(\frac{(2d)^{3/2} \log(2 + 2d(1 - \epsilon^2))}{(\epsilon^2)^{2d}} \right) \\ &\leq C \left(\frac{1}{\epsilon} \right)^{5d} \end{aligned}$$

for some universal constant $C > 0$. □

Let us now formally define the proof-discretised version of QPCP, denoted as QPCP_ϵ. We will use a non-adaptive formulation of the quantum PCP, since our proof of $\mathsf{QPCP}[\mathcal{O}(1)] = \mathsf{QPCP}_{\mathrm{NA}}[\mathcal{O}(1)]$ in Theorem 3 relativizes.

Definition 9 (Proof-discretised QPCP)**.** *Let $q \in \mathbb{N}$, $\epsilon \in [0,1]$ be some constant, and let $\mathcal{H} = \mathbb{C}^{2^q}$. Let $D_\epsilon^{2^q}$ be a ϵ-covering density matrix set as in Section 6.1. Then we define $\mathsf{QPCP}_\epsilon[k,q,c,s]$ as the complexity class which is just as $\mathsf{QPCP}[k,q,c,s]$, but where the reduced density matrix $\rho \in \mathcal{D}(\mathcal{H})$ obtained from the proof is projected (and subsequently re-normalised) on the density matrix $\rho' \in D_\epsilon^{2^q}$ closest to ρ in trace distance. If $c = 2/3$ and $s = 1/3$, we simply write $\mathsf{QPCP}_\epsilon[q]$.*

Whilst QPCP_ϵ is a highly artificial class, it will be helpful for us to prove some results since we can now use counting arguments on the exact size of the proof space that Merlin sees. However, note that the definition of this does *not* put any limitations on the proofs that Merlin can send, as the projection on a state from the set $D_\epsilon^{2^q}$ (and following re-normalisation) happens only after the proof has been received by the verifier. Therefore, the verifier has the guarantee that Merlin sends a valid quantum state, circumventing the problem of checking the consistency of the local density matrices with a global quantum state (which is QMA-hard by the results of [32, 33]).

Let us now justify that, with respect to all (quantum) oracles, we have that putting this extra restriction does not limit the power of QPCP_ϵ as long as the ϵ-covering set is large enough.

Lemma 16. *Let $q \in \mathbb{N}$, $\epsilon \in (0, 1/6)$ constant and let $D_\epsilon^{2^q}$ be a ϵ-covering density matrix set as in Section 6.1. We have that for all quantum oracles U that*

$$\mathsf{QPCP}^U[k,q] \subseteq \mathsf{QPCP}_\epsilon^U[k,q,2/3-\epsilon,1/3+\epsilon],$$

Proof. Since the proof of $\mathsf{QPCP}[\mathcal{O}(1)] = \mathsf{QPCP}_{\mathrm{NA}}[\mathcal{O}(1)]$ relativizes (Theorem 3), our assumptions that the PCP is non-adaptive is w.l.o.g. even in a relativized setting. Let V_x^U be a $\mathsf{QPCP}_{\mathrm{NA}}^U$ verification circuit as in Definition 4 (but with the input x hardcoded into it) that uses $p_1(n)$ ancilla qubits and has access to the quantum oracle U, and $x \in \{0,1\}^n$. We define

$$p_x(i_1,\ldots,i_q) := \left\| \Pi_{i_1,\ldots,i_q} V_{x,0}^U |x\rangle |0\rangle^{\otimes p_1(n)} \right\|^2.$$

Let $\xi_{i_1,\ldots,i_q} = \mathrm{tr}_{C_{i_1,\ldots,i_q}}[\xi]$ and $\hat{\xi}_{i_1,\ldots,i_q}$ be the reduced density matrix that is accessed by QPCP_ϵ^U instead. Next, define $\sigma_{i_1,\ldots,i_q} = \xi_{i_1,\ldots,i_q} \otimes \rho_{i_1,\ldots,i_q}$ where $\rho_{i_1,\ldots,i_q}$ is the $p_1(n)+n$-qubit post-measurement state after the measurement outcome to decide the indices of the proof to be queried returned $i_1,\ldots,i_q$. In similar fashion, define $\tilde{\sigma}_{i_1,\ldots,i_q}$ using $\hat{\xi}_{i_1,\ldots,i_q}$. We

have

$$\begin{aligned}
&\left|\Pr\left[V_x^U \text{accepts } \xi\right] - \Pr\left[\hat{V}_x^U \text{accepts } \xi\right]\right| \\
&\quad= \sum_{\{i_1,\ldots,i_q\}\in\Omega} p_x(i_1,\ldots,i_q)\left(\operatorname{tr}\left[\Pi^1_{\text{output}} V^{U\dagger}_{x,1}(\sigma i_1,\ldots,i_q - \hat{\sigma}_{i_1,\ldots,i_q}) V^U_{x,1}\Pi^1_{\text{output}}\right]\right) \\
&\quad= \sum_{\{i_1,\ldots,i_q\}\in\Omega} p_x(i_1,\ldots,i_q)\left(\operatorname{tr}\left[V^U_{x,1}\Pi^1_{\text{output}} V^{U\dagger}_{x,1}(\sigma_{i_1,\ldots,i_q} - \hat{\sigma}_{i_1,\ldots,i_q})\right]\right) \\
&\quad\leq \sum_{\{i_1,\ldots,i_q\}\in\Omega} p_x(i_1,\ldots,i_q)\left\|\sigma_{i_1,\ldots,i_q} - \hat{\sigma}_{i_1,\ldots,i_q}\right\|_1 \\
&\quad= \sum_{\{i_1,\ldots,i_q\}\in\Omega} p_x(i_1,\ldots,i_q)\left\|\xi_{i_1,\ldots,i_q} - \hat{\xi}_{i_1,\ldots,i_q}\right\|_1 \\
&\quad\leq \epsilon,
\end{aligned}$$

where we used the cyclic property of the trace in going from line 1 to line 2 and that the trace distance is preserved under tensor products (with the same state) in going from line 3 to line 4. □

6.2 Oracle separations

We will now follow the proof of the oracle separation in [34] to show that the same oracle U that shows $\mathsf{QCMA}^U \neq \mathsf{QMA}^U$ also shows that $\mathsf{QPCP}[\mathcal{O}(1)]^U \neq \mathsf{QMA}^U$. We will omit most of the details, focusing on the parts where we depart from [34].

Let μ be the uniform probability measure over $\mathcal{P}(\mathcal{H})$. We will need the notion of p-uniform probability measures, which have the following definition.

Definition 10 (p-uniform probability measures)**.** *For all $p \in [0,1]$, a probability measure σ over $\mathcal{P}(\mathcal{H})$ is called p-uniform if $p\sigma \leq \mu$. Equivalently, σ is p-uniform if it can be obtained by starting from μ, and then conditioning on an event that occurs with probability at least p.*

We also state the following lemma from [34].

Lemma 17 (Expected fidelity p-uniform random states [34])**.** *Let $\mathcal{H}$ be a d-dimensional Hilbert space and let σ be a p-uniform probability measure over $\mathcal{P}(\mathcal{H})$. Then for all mixed states ρ, we have that*

$$\mathbb{E}_{|\Psi\rangle\in\sigma} \langle\Psi|\,\rho\,|\Psi\rangle = \mathcal{O}\left(\frac{1+\log 1/p}{d}\right).$$

The key idea behind the oracle separation in [34] is to provide a lower bound on a decision version of "quantum search", where one is given a black-box unitary which either marks a single n-qubit quantum state or applies the identity operation. This is reminiscent of the lower bound on search by using the OR-function, where one is given a black-box Boolean function which "marks" a single bit string (i.e. $f(x) = 1$ for a single x) or for all bit strings outputs 0 ($f(x)$ for all x). By exploiting the fact that there are a doubly exponential number of quantum states that have small pairwise overlap, [34] can show that a polynomial number of classical bits to point you towards the right quantum state does not help much, and still requires one to make an exponential number of queries to the oracle.

Lemma 18. *Let $q \in \mathbb{N}$ and let $D_\epsilon^{2^q}$ be a ϵ-covering set of mixed states on q qubits as per Section 6.1. Suppose we are given oracle access to an n-qubit unitary U, and want to decide which of the following holds:*

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

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

Then even if we have a witness tuple $(i_1, \ldots, i_q, \rho)$, where $i_{i_1}, \ldots, i_{i_q}$ is a set of the q indices and a quantum witness $\rho \in D_\epsilon^{2^q}$ in support of case (i), we still need

$$\Omega\left(\sqrt{\frac{2^n}{5 \cdot 2^q \log\left(\frac{1}{\epsilon}\right) + q \cdot \mathrm{polylog}(n,q) + C}}\right).$$

queries to verify the witness, with a bounded probability of error. Here $C > 0$ is some universal constant.

Proof. This follows the same structure as Theorem 3.3 in [34], but considers q-qubit proofs from $D_\epsilon^{2^q}$ instead of classical proofs. Let $\mathcal{H}$ be a 2^n-dimensional Hilbert space. Let A be a quantum algorithm that queries the oracle U a total of T times, with the goal to output 1 in case (i) and output 0 in case (ii). For each n-qubit state $|\Psi\rangle$, we fix both the indices $(i_1, \ldots, i_q), i_j \in [p(n)]$ for all $j \in [q]$ *and* corresponding quantum witness $\rho \in D_\epsilon^{2^q}$ that maximises the difference in the probability that A accepts as compared to the case where U is the identity. This is allowed, since by Remark 3 the choice of distribution over indices does not have to depend on the oracle nor the input. The expected difference in accepting given $U_{|\Psi\rangle}$ or I satisfies

$$\begin{aligned} g &\leq \max_\xi \mathbb{E}_{i_1,\ldots,i_q} \Pr\left[V^{U_{|\Psi\rangle}} \text{ accepts querying } \xi_{i_1,\ldots,i_q}\right] - \mathbb{E}_{i_1,\ldots,i_1} \Pr\left[V^I \text{ accepts querying } \xi_{i_1,\ldots,i_q}\right] \\ &= \max_\xi \mathbb{E}_{i_1,\ldots,i_q} [\Pr\left[V^{U_{|\Psi\rangle}} \text{ accepts querying } \xi_{i_1,\ldots,i_q}\right] - \Pr\left[V^I \text{ accepts querying } \xi_{i_1,\ldots,i_q}\right]] \\ &\leq \max_\xi \max_{i_1,\ldots,i_q} \Pr\left[V^{U_{|\Psi\rangle}} \text{ accepts querying } \xi_{i_1,\ldots,i_q}\right] - \Pr\left[V^I \text{ accepts querying } \xi_{i_1,\ldots,i_q}\right], \end{aligned}$$

so we can set $\rho = \xi_{i_1,\ldots,i_q}$ to be the density matrix as in the last line, and our lower bound would also apply in the setting where the indices are chosen probabilistically. Here $g = \Omega(1)$ is the promise gap between both cases. If $g > \frac{1}{2}$, then we have that the completeness satisfies $c \geq g$ and the soundness $s \leq 1 - g$ with $c - s = \Omega(1)$. Let $S(i_1, \ldots, i_q, \rho)$ be the set of $|\Psi\rangle$'s that belong to the index/state tuple $(i_1, \ldots, i_q, \rho)$. Then we have that the $S(i_1, \ldots, i_q, \rho)$'s form a partition of $\mathcal{P}(\mathcal{H})$, and hence we must have that there exists a $(i_1^*, \ldots, i_q^*, \rho^*)$ such that

$$\begin{aligned} \Pr_{|\Psi\rangle\in\mu}\left[|\Psi\rangle \in S(i_1^*, \ldots, i_q^*, \rho^*)\right] &\geq \left(\binom{p(n)}{q} C \left(\frac{1}{\epsilon}\right)^{5\cdot 2^q}\right)^{-1} \\ &\geq \left(\left(\frac{ep(n)}{q}\right)^q C \left(\frac{1}{\epsilon}\right)^{5\cdot 2^q}\right)^{-1} =: p \end{aligned}$$

Since ρ is a q-qubit quantum state, there exists a unitary U of circuit depth $\mathcal{O}(4^q \mathrm{poly}(n))$ such that $U|0\rangle^{\otimes 2q} = |\phi\rangle$ and $\|\mathrm{tr}_B[|\phi\rangle\langle\phi|] - \rho\|_1 \leq 1/\exp(n)$ (by the Solovay-Kitaev Theorem). We hardcode U and the information $i_q, \ldots, i_q$ into A. Now let σ be the probability

measure over $S(i_1^*, \ldots, i_q^*, \rho^*)$. Note that it is p-uniform. Following the rest of the steps of the proof in [34, Theorem 3.3], this yields the desired bound of

$$T \geq \Omega\left(\sqrt{\frac{2^n}{5 \cdot 2^q \log\left(\frac{1}{\epsilon}\right) + q \cdot \text{polylog}(n,q) + C}}\right).$$

□

Theorem 6. *There exists a quantum oracle U relative to which $\mathsf{QCMA}^U \neq \mathsf{QMA}^U$ and $\mathsf{QPCP}[k,q]^U \neq \mathsf{QMA}^U$, for all $q \in \mathcal{O}(\log n)$, $1 \leq k \leq \text{poly}(n)$.*

Proof. This follows from using the same proof as [34], Theorem 1.1, using the lower bound of Lemma 18 to show that $\mathsf{QPCP}_{P^{k'}_{\epsilon}}[q]^U \neq \mathsf{QMA}^U$ for all constant $q \in \mathcal{O}(\log n)$, from which the separation $\mathsf{QPCP}[q]^U \neq \mathsf{QMA}^U$ follows by Lemma 16. □

We remark that our proof technique should work for any oracle separation between QCMA and QMA that uses counting arguments exploiting the doubly exponentially large number of quantum states and the fact that QCMA has only access to an exponential number of proofs. This shows that any proof of the quantum PCP conjecture (via the proof verification formulation) requires (as expected) quantumly non-relativising techniques.

The previous quantum oracle separations crucially exploit the doubly exponentially large number of quantum states that have low mutual fidelities. However, one may wonder whether a similar idea might be used to also show that for low-complexity states a similar separation holds? It turns out that this is surprisingly easy, by simply using a similar oracle to the one that separates BQP from NP [48]. Observe this oracle is classical, and can alternatively be viewed as a state 1-design over the set of all quantum states used in the previous separation. We state the following result, for which the proof is given in Appendix D.

Corollary 4. *There exists a classical oracle $\mathcal{O}$ relative to which $\mathsf{QPCP}[k,q]^{\mathcal{O}} \neq \mathsf{QMA}^{\mathcal{O}}$ (in fact $\neq \mathsf{NP}^{\mathcal{O}}$) for all $q \in \mathcal{O}(1)$.*

Note that a much stronger statement also holds, we can even combine QCMA and QPCP (so a classical proof and quantum proof which can be accessed locally) and the same separation would still hold.

References

[1] Stephen A. Cook. The complexity of theorem-proving procedures. In *Proceedings of the Third Annual ACM Symposium on Theory of Computing*, STOC '71, page 151–158, 1971. ISBN 9781450374644. 3

[2] Leonid A. Levin. Universal sequential search problems. *Problemy peredachi informatsii*, 9(3):115–116, 1973. 3

[3] Alexei Y. Kitaev, Alexander Shen, and Mikhail N. Vyalyi. *Classical and quantum computation*. American Mathematical Society, 2002. ISBN 978-0-8218-3229-5. 3, 25, 26

[4] Sevag Gharibian. Guest Column: The 7 faces of quantum NP. *SIGACT News*, 54(4): 54–91, January 2024. ISSN 0163-5700. arXiv: 2310.18010. 3

[5] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof Verification and the Hardness of Approximation Problems. *Journal of the ACM*, 45(3):501–555, May 1998. ISSN 0004-5411. 3

[6] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: a new characterization of NP. *J. ACM*, 45(1):70–122, January 1998. ISSN 0004-5411. 3

[7] Irit Dinur. The PCP Theorem by Gap Amplification. *Journal of the ACM*, 54(3): 12–es, June 2007. ISSN 0004-5411. 3

[8] Anurag Anshu, Nikolas P. Breuckmann, and Chinmay Nirkhe. NLTS Hamiltonians from Good Quantum Codes. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, STOC 2023, page 1090–1096, 2023. ISBN 9781450399135. arXiv: 2206.13228. 3

[9] Nolan J. Coble, Matthew Coudron, Jon Nelson, and Seyed Sajjad Nezhadi. Local Hamiltonians with No Low-Energy Stabilizer States. In *18th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2023)*, volume 266, pages 14:1–14:21, 2023. ISBN 978-3-95977-283-9. arXiv: 2302.14755. 3

[10] Eric R. Anschuetz, David Gamarnik, and Bobak Kiani. Combinatorial NLTS From the Overlap Gap Property. *Quantum*, 8:1527, November 2024. ISSN 2521-327X. arXiv:2304.00643. 3

[11] Nolan J. Coble, Matthew Coudron, Jon Nelson, and Seyed Sajjad Nezhadi. Hamiltonians whose low-energy states require $\Omega(n)$ T gates, 2023. arXiv: 2310.01347. 3

[12] Yaroslav Herasymenko, Anurag Anshu, Barbara M. Terhal, and Jonas Helsen. Fermionic Hamiltonians without trivial low-energy states. *Phys. Rev. A*, 109:052431, May 2024. arXiv: 2307.13730. 3, 11

[13] Nikhil Bansal, Sergey Bravyi, and Barbara M. Terhal. Classical approximation schemes for the ground-state energy of quantum and classical ising spin Hamiltonians on planar graphs. *Quantum Information & Computation*, 9(7):701–720, July 2009. ISSN 1533-7146. arXiv: 0705.1115. 3

[14] Fernando G.S.L. Brandao and Aram W. Harrow. Product-state approximations to quantum ground states. In *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '13, page 871–880, 2013. ISBN 9781450320290. arXiv: 1310.0017. 3, 9

[15] Dorit Aharonov and Alex Bredariol Grilo. Stoquastic PCP vs. Randomness. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1000–1023, 2019. arXiv: 1901.05270. 3

[16] Sevag Gharibian and François Le Gall. Dequantizing the Quantum singular value transformation: hardness and applications to Quantum chemistry and the Quantum PCP conjecture. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2022, page 19–32, 2022. ISBN 9781450392648. arXiv: 2111.09079. 3

[17] Chris Cade, Marten Folkertsma, and Jordi Weggemans. Complexity of the guided local Hamiltonian problem: improved parameters and extension to excited states, 2022. arXiv: 2207.10097. 3

[18] Chris Cade, Marten Folkertsma, Sevag Gharibian, Ryu Hayakawa, François Le Gall, Tomoyuki Morimae, and Jordi Weggemans. Improved Hardness Results for the Guided Local Hamiltonian Problem. In *50th International Colloquium on Automata, Languages, and Programming (ICALP 2023)*, volume 261, pages 32:1–32:19, 2023. ISBN 978-3-95977-278-5. arXiv: 2207.10250. 3

[19] Jordi Weggemans, Marten Folkertsma, and Chris Cade. Guidable Local Hamiltonian Problems with Implications to Heuristic Ansatz State Preparation and the Quantum PCP Conjecture. In *19th Conference on the Theory of Quantum Computation, Com-*

munication and Cryptography (TQC 2024), volume 310, pages 10:1–10:24, 2024. ISBN 978-3-95977-328-7. arXiv: 2302.11578. 3, 6, 23, 50

[20] Dorit Aharonov, Itai Arad, Zeph Landau, and Umesh Vazirani. The detectability lemma and quantum gap amplification. In *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*, STOC '09, page 417–426, 2009. ISBN 9781605585062. arXiv: 0811.3412. 3, 4, 5, 7, 31

[21] Alex B. Grilo. *Quantum proofs, the local Hamiltonian problem and applications*. PhD thesis, Université Sorbonne Paris Cité, April 2018. 3, 4, 7, 11, 14, 19, 31

[22] Dorit Aharonov and Tomer Naveh. Quantum NP—a survey, October 2002. arXiv: quant-ph/0210077. 4, 13

[23] Anand Natarajan and Chinmay Nirkhe. The status of the quantum PCP conjecture (games version), 2024. arXiv: 2403.13084. 6

[24] Anurag Anshu, Jonas Haferkamp, Yeongwoo Hwang, and Quynh T Nguyen. Unique-QMA vs QMA: oracle separation and eigenstate thermalization hypothesis, 2024. arXiv: 2410.23811. 6

[25] Scott Aaronson, DeVon Ingram, and William Kretschmer. The Acrobatics of BQP. In *Proceedings of the 37th Computational Complexity Conference (CCC 2022)*, volume 234 of *LIPIcs*, pages 20:1–20:17, 2022. DOI: 10.4230/LIPIcs.CCC.2022.20. arXiv: 2111.10409. 7

[26] Aram W. Harrow and Ashley Montanaro. Testing Product States, Quantum Merlin-Arthur Games and Tensor Optimization. *J. ACM*, 60(1), February 2013. ISSN 0004-5411. arXiv: 1001.0017. 7, 11, 14

[27] Yi-Kai Liu, Matthias Christandl, and Frank Verstraete. Quantum Computational Complexity of the N-Representability Problem: QMA Complete. *Physical review letters*, 98:110503, Mar 2007. arXiv: quant-ph/0609125. 7

[28] Scott Aaronson, Salman Beigi, Andrew Drucker, Bill Fefferman, and Peter Shor. The power of unentanglement. In *2008 23rd Annual IEEE Conference on Computational Complexity*, pages 223–236, 2008. arXiv: 0804.0802. 7

[29] Salman Beigi. NP vs $\mathrm{QMA}_{\log}(2)$. *Quantum Information & Computation*, 10(1):141–151, 2010. arXiv: 0810.5109. 7

[30] Hugue Blier and Alain Tapp. All Languages in NP Have Very Short Quantum Proofs. In *2009 Third International Conference on Quantum, Nano and Micro Technologies*, pages 34–37, 2009. arXiv: 0709.0738. 7

[31] André Chailloux and Or Sattath. The Complexity of the Separable Hamiltonian Problem. In *2012 IEEE 27th Conference on Computational Complexity*, pages 32–41, 2012. arXiv: 1111.5247. 8, 11, 31, 38

[32] Yi-Kai Liu. Consistency of local density matrices is QMA-complete. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques: 9th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, APPROX 2006 and 10th International Workshop on Randomization and Computation, RANDOM 2006*, pages 438–449, 2006. arXiv: quant-ph/0604166. 8, 31, 42

[33] Anne Broadbent and Alex Bredariol Grilo. QMA-hardness of consistency of local density matrices with applications to quantum zero-knowledge. *SIAM Journal on Computing*, 51(4):1400–1450, 2022. arXiv: 1911.07782. 8, 31, 32, 42

[34] Scott Aaronson and Greg Kuperberg. Quantum versus classical proofs and advice. In *Twenty-Second Annual IEEE Conference on Computational Complexity (CCC'07)*, pages 115–128, 2007. arXiv: quant-ph/0604056. 8, 43, 44, 45

[35] Lance Fortnow. The role of relativization in complexity theory. *Bulletin of the EATCS*, 52:229–243, 1994. 8

[36] Sergey Bravyi, David P. DiVincenzo, Daniel Loss, and Barbara M. Terhal. Quantum Simulation of Many-Body Hamiltonians Using Perturbation Theory with Bounded-Strength Interactions. *Physical Review Letters*, 101:070503, August 2008. arXiv: 0803.2686. 9

[37] Chris Marriott and John Watrous. Quantum Arthur-Merlin games. *Computational Complexity*, 14(2):122–152, 2005. DOI: 10.1007/s00037-005-0194-x. arXiv: cs/0506068. 9

[38] Dorit Aharonov, Itai Arad, and Thomas Vidick. Guest column: the quantum PCP conjecture. *SIGACT News*, 44(2):47–79, June 2013. ISSN 0163-5700. arXiv: 1309.7495. 11

[39] Dorit Aharonov, Vaughan Jones, and Zeph Landau. A polynomial quantum algorithm for approximating the Jones polynomial. In *Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '06, page 427–436, 2006. ISBN 1595931341. arXiv: quant-ph/0511096. 19

[40] Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12:389–434, 2012. 27

[41] Jasper Lee. Lecture 3: Concentration inequalities and mean estimation. Lecture Notes, CSCI 1951-W Sublinear Algorithms for Big Data, Fall 2020, 2020. URL https://cs.brown.edu/courses/csci1951-w/lec/lec%203%20notes.pdf. 33

[42] Persi Diaconis and David Freedman. Finite Exchangeable Sequences. *The Annals of Probability*, pages 745–764, 1980. 34

[43] Renato Renner. Symmetry of large physical systems implies independence of subsystems. *Nature Physics*, 3(9):645–649, 2007. arXiv: quant-ph/0703069. 34

[44] Matthias Christandl, Robert König, Graeme Mitchison, and Renato Renner. One-and-a-half quantum de Finetti theorems. *Communications in Mathematical Physics*, 273(2):473–498, 2007. arXiv: quant-ph/0602130. 34

[45] Fernando GSL Brandao, Matthias Christandl, and Jon Yard. Faithful squashed entanglement. *Communications in Mathematical Physics*, 306:805–830, 2011. arXiv: 1010.1750. 34

[46] Fernando G.S.L. Brandao and Aram W. Harrow. Quantum de Finetti Theorems under Local Measurements with Applications. In *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '13, page 861–870, 2013. ISBN 9781450320290. arXiv: 1210.6367. 34

[47] Károly Böröczky and Gergely Wintsche. *Covering the Sphere by Equal Spherical Balls*, pages 235–251. Springer, Berlin, Heidelberg, 2003. ISBN 978-3-642-55566-4. 41

[48] Charles H. Bennett, Ethan Bernstein, Gilles Brassard, and Umesh Vazirani. Strengths and Weaknesses of Quantum Computing. *SIAM Journal on Computing*, 26(5):1510–1523, 1997. arXiv: quant-ph/9701001. 45

A Strong error reduction for non-adaptive quantum PCPs with near-perfect completeness

In this appendix we argue that non-adaptive quantum PCPs allow for strong error reduction, so long as they are nearly perfectly complete. We will use the well-known gentle measurement lemma.

Lemma 19 (Gentle measurement lemma)**.** *Consider a state ρ and a measurement operator Λ where $0 \preceq \Lambda \preceq 1$. Now suppose that $\mathrm{tr}[\Lambda\rho] \geq 1-\epsilon$, where $0 < \epsilon \leq 1$. Then the post-measurement state*

$$\rho' = \frac{\sqrt{\Lambda}\rho\sqrt{\Lambda}}{\mathrm{tr}[\Lambda\rho]}$$

satisfies $\|\rho - \rho'\|_1 \leq 2\sqrt{\epsilon}$.

The error reduction protocol will be a variant of the parallel repetition protocol but now we will use the *same* quantum proof in all repetitions.

Claim 1 (Strong error reduction for non-adaptive QPCPs with near-perfect completeness)**.** *For $l \in \mathcal{O}(1)$ it holds that*

$$\mathsf{QPCP}_{\mathrm{NA}}[1, q, c, s] = \mathsf{QPCP}_{\mathrm{NA}}[1, lq, c', s']$$

with $c = 1 - 2^{-\Omega(n)}$, $s = 1/2$ and $c' = 1 - 2^{-\Omega(n)}$, $s' = 2^{-\mathcal{O}(l)}$.

Proof. Let $\mathcal{A}$ be a (q)-$\mathsf{QPCP}_{\mathrm{NA}}$ verifier consisting of circuits V_0 and V_1, and let ξ be the provided quantum proof. Let $\rho^0_{i_1,\ldots,i_q} |i_1\rangle\langle i_1| \ldots |i_q\rangle\langle i_q| \otimes \rho_{\mathrm{rest}}$ be the post-measurement state after PVM Π_1 is performed to obtain indices $i_1, \ldots, i_q$. Now let $\rho^1_{i_1,\ldots,i_q}$ be the state after V_1 acts on $|i_1\rangle\langle i_1| \ldots |i_q\rangle\langle i_q| \otimes \rho_1$ and qubits $i_1, \ldots, i_q$ of ξ. In the NO-case, we have that the soundness property holds with respect to all proof states. So for every reduced density matrix ξ' after the measurement is performed, we have

$$\mathbb{E}_{i_1,\ldots,i_q}\left[\mathrm{tr}\left[\Pi^1_{\mathrm{output}}\rho^0_{i_1,\ldots,i_q}\right]\right] \leq 1/2.$$

Hence, the bound follows from the same argument as in the weak error reduction case (Lemma 2).

In the YES-case, we have that Π_0^{output} satisfies

$$\mathbb{E}_{i_1,\ldots,i_q}\left[\mathrm{tr}\left[\Pi^1_{\mathrm{output}}\rho^0_{i_1,\ldots,i_q}\right]\right] \geq 1 - 2^{-c_0 n}$$

which implies that

$$\mathbb{P}_{i_1,\ldots,i_q}[\mathrm{tr}\left[\Pi^1_{\mathrm{output}}\rho^0_{i_1,\ldots,i_q}\right] \geq 1 - 2^{-c_0 n/2}] \geq 1 - 2^{c_0 n/2}.$$

By Lemma 19, we have that the post-measurement state $\rho'_{i_1,\ldots,i_q}$ satisfies

$$\frac{1}{2}\left\|\rho^1_{i_1,\ldots,i_q} - \rho'_{i_1,\ldots,i_q}\right\|_1 \leq 2^{-nc_0/4}.$$

with probability $\geq 1 - 2^{c_0 n/2}$. Hence, we can simply apply $V_1^\dagger$ such that the density matrix ξ' in the proof register satisfies

$$\frac{1}{2}\|\xi' - \xi\|_1 \leq 2^{-c_0 n/4},$$

using the fact that the trace distance can only decrease under the partial trace. Therefore, our new acceptance probability satisfies

$$\mathbb{E}_{i_1,\ldots,i_q}\left[\mathrm{tr}\left[\Pi^1_{\mathrm{output}}\rho^0_{i_1,\ldots,i_q}\right]\right] \geq 1 - 2^{-c_0 n} - 2^{-c_0 n/4} \geq 1 - 2^{-c_1 n}$$

for some constant $c_1 > c_0$. Hence, after $l = \mathcal{O}(1)$ of such steps, we have that with probability at least

$$\prod_{i\in[l]} 1 - 2^{-c_i n/2} \geq \left(1 - 2^{-c_l n/2}\right)^l$$

each step had an acceptance probability $\geq 1 - 2^{-c_1 n}$. Hence, the majority vote definitely accepts with probability at least

$$\left(1 - 2^{-c_l n/2}\right)^l \left(1 - 2^{-c_l n}\right)^l \geq 1 - 2^{-\Omega(n)},$$

when $l \in \mathcal{O}(1)$, as desired. □

B Sufficient error bounds for learning weighted Hamiltonians from non-adaptive quantum PCPs

In the lemma proven in this appendix we give sufficient parameters to adopt the quantum reduction of [19] to the $\mathsf{QPCP}_{\mathrm{NA}}$ setting.

Lemma 20. *Let $H = \sum_{i\in[m]} p_i H_i$ be a k-local Hamiltonian consisting of weights $p_i \in [0,1]$ such that $\sum_{i\in[m]} p_i = 1$, and k-local terms H_i for which $\|H_i\| \leq 1$ for all $i \in [m]$. Suppose $\tilde{H} = \sum_{i\in[m]} \tilde{p}_i \tilde{H}_i$ is another Hamiltonian such that, for all $i \in [m]$, we have $\tilde{H}_i$ PSD, $|\tilde{p}_i - p_i| \leq \epsilon_0$ and $\left\|H_i - \tilde{H}_i\right\| \leq \epsilon_1$. Let $\tilde{W} = \sum_{i\in[m]} \tilde{p}_i$ and,*

$$\hat{p}_i = \frac{\tilde{p}_i}{\tilde{W}}, \quad \hat{H}_i = \frac{\tilde{H}_i}{\mathrm{argmax}_{i\in[m]}\{\tilde{H}_i, 1\}}$$

for all $i \in [m]$. Then we have that setting

$$\epsilon_0 \leq \frac{\epsilon}{8m^2} \quad \epsilon_1 \leq \frac{\epsilon}{6},$$

suffices to have

$$\left\|H_x - \tilde{H}_x\right\| \leq \epsilon.$$

Proof. Suppose $\epsilon_0 < 1/2m$. Then

$$\begin{aligned}
|\hat{p}_i - p_i| &= \left|\frac{\tilde{p}_i}{\tilde{W}} - p_i\right| \\
&\leq \frac{p_i + \epsilon_0}{1 - m\epsilon_0} - p_i \\
&= \frac{\epsilon_0(mp_i + 1)}{1 - m\epsilon_0} \\
&\leq \frac{\epsilon_0(m+1)}{1 - m\epsilon_0} \\
&\leq 4\epsilon_0 m.
\end{aligned}$$

For the local terms we have

$$\begin{aligned}\left\|\hat{H}_i - H_i\right\| &\leq \left\|\hat{H}_i - \tilde{H}_i\right\| + \left\|\tilde{H}_i - H_i\right\| \\ &\leq \left|\frac{1}{1+\epsilon_1} - 1\right|\left\|\tilde{H}_i\right\| + \epsilon_1 \\ &\leq \epsilon_1(1+\epsilon_1) + \epsilon \\ &= 2\epsilon_1 + \epsilon_1^2\end{aligned}$$

So finally,

$$\begin{aligned}\left\|H - \tilde{H}\right\| &= \left\|\sum_{i\in[m]]} \hat{p}_i\tilde{H}_i - \sum_{i\in[m]} p_i H_i\right\| \\ &\leq \sum_{i\in[m]} \left\|\hat{p}_i\tilde{H}_i - p_i H_i\right\| \\ &\leq \sum_{i\in[m]} \left\|\hat{p}_i\left(\tilde{H}_i - H_i\right) + H_i\left(\hat{p}_i - p_i\right)\right\| \\ &\leq \sum_{i\in[m]} \left\|\hat{p}_i\left(\tilde{H}_i - H_i\right)\right\| + \left\|H_i\left(\hat{p}_i - p_i\right)\right\| \\ &\leq \sum_{i\in[m]} \hat{p}_i\left\|\tilde{H}_i - H_i\right\| + \|H_i\||\hat{p}_i - p_i| \\ &\leq 2\epsilon_1 + \epsilon_1^2 + 4m^2\epsilon_0\end{aligned}$$

which can be made $\leq \epsilon$ if for example

$$\epsilon_0 := \frac{\epsilon}{8m^2}, \quad \epsilon_1 := \frac{\epsilon}{6}.$$

□

C Proof of Corollary 1

Corollary 1. *For any $\epsilon, \delta > 0$, under the same setup as in Theorem 1, there exists a quantum reduction which produces a fixed Hamiltonian $\tilde{H}_x$ with probability $1 - \delta$ which satisfies Eq. (9) and runs in time* $\mathrm{poly}(n, 1/\epsilon, \log(1/\delta))$.

Proof. We have that H_x (resp. $\tilde{H}_x$ are specified by the $4^q\binom{kp_2(n)}{q}$ complex numbers $\langle\alpha| H_{x,i_1,\ldots,i_q} |\beta\rangle$ (resp. $\langle\alpha| \tilde{H}_{x,(i_1,\ldots,i_q)} |\beta\rangle$). Note that $\left|\langle\alpha| H_{x,i_1,\ldots,i_q} |\beta\rangle\right| \leq 1$ since

$$\left|\langle\alpha| H_{x,i_1,\ldots,i_q} |\beta\rangle\right| \leq \max_{\alpha,\beta}\left|\langle\alpha| H_{x,i_1,\ldots,i_q} |\beta\rangle\right| \leq \left\|H_{x,i_1,\ldots,i_q}\right\| \leq 1.$$

Therefore, we can adopt the following binary notation to specify the values of $\langle\alpha| H_{x,i_1,\ldots,i_q} |\beta\rangle \in \mathbb{C}$: we use the most-significant bit to indicate whether it is the real or complex part, all remaining bits to specify a value in $[-1, 1]$ in evenly spaced intervals. We have that

$$\begin{gathered}\mathrm{Re}(\langle\alpha| \tilde{H}_{x,(i_1,\ldots,i_q)} |\beta\rangle) \text{ is correct up to } \eta \text{ bits} \\ \Updownarrow \\ \left|\mathrm{Re}(\langle\alpha| \tilde{H}_{x,(i_1,\ldots,i_q)} |\beta\rangle) - \mathrm{Re}(\langle\alpha| \tilde{H}_{x,(i_1,\ldots,i_q)} |\beta\rangle)\right| \leq \frac{2}{(2^\eta + 1)}.\end{gathered}$$

The same argument holds for the imaginary part of $\langle\alpha|\,\tilde{H}_{x,(i_1,\ldots,i_q)}\,|\beta\rangle$. By the triangle inequality, we have that in this case

$$\left|\langle\alpha|\,\tilde{H}_{x,(i_1,\ldots,i_q)}\,|\beta\rangle - \langle\alpha|\,\tilde{H}_{x,(i_1,\ldots,i_q)}\,|\beta\rangle\right| \leq \frac{4}{2^\eta+1} \coloneqq \epsilon'.$$

To achieve

$$\frac{4}{2^\eta+1} \leq \frac{\epsilon}{|\Omega|2^{q+4}q!},$$

it suffices to set

$$\eta \coloneqq \lceil \log\left(\frac{4|\Omega|2^{q+4}q!}{\epsilon} - 1\right)\rceil.$$

Since the Hadamard test can learn η bits of precision in $\mathcal{O}(2^\eta)$ time, we have that the time complexity of the reduction in Algorithm 1 still runs in time $\mathrm{poly}(n, 1/\epsilon, \log(1/\delta))$ as $\eta = \mathcal{O}(\log(\mathrm{poly}(n, 1/\epsilon, \log(1/\delta))))$. □

D A classical oracle separation

Here we show that the same idea behind the quantum oracle separation of Section 6 can be used to give a classical oracle separation. We first prove a lower bound analogous to the one in Lemma 18 but now for "classical search".

Lemma 21. *Let $q \in \mathbb{N}$ and let $D_\epsilon^{2^q}$ be a ϵ-covering set of mixed states on q qubits as per Section 6.1. Suppose we are given oracle access to an n-qubit phase oracle O_f, and want to decide which of the following holds:*

(i) There exists an n-bit string x^* such that $\mathcal{O}_f\,|x^*\rangle = -\,|x^*\rangle$

(ii) $\mathcal{O}_f\,|x\rangle = |x\rangle$ for all x.

Then even if we have a witness tuple $(i_1,\ldots,i_q,\rho)$, where $i_{i_1},\ldots,i_{i_q}$ is a set of the q indices and a quantum witness $\rho \in D_\epsilon^k$ in support of case (i), we still need

$$\Omega\left(\sqrt{\frac{2^n}{\left(\frac{ep(n)}{q}\right)^q C\left(\frac{1}{\epsilon}\right)^{5\cdot 2^q}}}\right).$$

queries to verify the witness, with bounded probability of error. Here $C > 0$ is some universal constant.

Proof. Let $\Omega = \{0,1\}^n$, and let σ be any p-uniform distribution over Ω. Then we have that for all $\rho \in \mathcal{D}(\mathcal{H})$

$$\begin{aligned}\max_{\sigma:p\text{-uniform}} \mathbb{E}_{|x\rangle\in\sigma}\left[\langle x|\,\rho\,|x\rangle\right] &\leq \max_{\sigma:p\text{-uniform}} \max_{z\in\Omega} \mathbb{E}_{|x\rangle\in\sigma}\left[\langle x|z\rangle\,\langle z|x\rangle\right] \\ &= \max_{\sigma:p\text{-uniform}} \mathbb{E}_{|x\rangle\in\sigma}\left[\langle x|0\rangle\,\langle 0|x\rangle\right] \\ &= \max_{\sigma:p\text{-uniform}} \mathbb{E}_{|x\rangle\in\sigma}\left[|\langle 0|x\rangle|^2\right],\end{aligned}$$

by the fact that the expected fidelity is maximised for ρ's that are basis states and then using the maximisation over p-uniform ρ to choose the maximal z to be 0. Clearly, this

is maximised by conditioning on any $\log_2(1/p)$ of bits being in 0, which happens with probability $(1/2)^{\log_2(1/p)} = p$. Hence, for any σ we have

$$\mathbb{E}_{|x\rangle\in\sigma}\left[|\langle 0|x\rangle|^2\right] \leq \frac{2^{\log_2(\frac{1}{p})}}{2^n} = \frac{1/p}{2^n},$$

and thus

$$\mathbb{E}_{|x\rangle\in\sigma}\left[\langle x|\,\rho\,|x\rangle\right] \leq \frac{1/p}{2^n},$$

for all mixed states ρ and all p-uniform distributions σ. Using the value of p as in Lemma 18 , we obtain the lower bound of

$$T \geq \left(\sqrt{\frac{2^n}{\left(\frac{ep(n)}{q}\right)^q C\left(\frac{1}{\epsilon}\right)^{5\cdot 2^q}}}\right).$$

□

Given Lemma 21, the following corollary follows directly from the same proof of Theorem 6.

Corollary 4. *There exists a classical oracle $\mathcal{O}$ relative to which $\mathsf{QPCP}[k,q]^{\mathcal{O}} \neq \mathsf{QMA}^{\mathcal{O}}$ (in fact $\neq \mathsf{NP}^{\mathcal{O}}$) for all $q \in \mathcal{O}(1)$.*