Guidable Local Hamiltonian Problems with Implications to Heuristic Ansatz State Preparation and the Quantum PCP Conjecture

Jordi Weggemans[1,2], Marten Folkertsma[1], and Chris Cade[2,3]

[1]QuSoft & CWI, Amsterdam, the Netherlands
[2]Fermioniq, Amsterdam, the Netherlands
[3]QuSoft & University of Amsterdam (UvA), Amsterdam, the Netherlands

September 29, 2023

Abstract

We study 'Merlinized' versions of the recently defined Guided Local Hamiltonian problem, which we call '*Guidable* Local Hamiltonian' problems. Unlike their guided counterparts, these problems do not have a guiding state provided as a part of the input, but merely come with the promise that one *exists*. We consider in particular two classes of guiding states: those that can be prepared efficiently by a quantum circuit; and those belonging to a class of quantum states we call *classically evaluatable*, for which it is possible to efficiently compute expectation values of local observables classically. We show that guidable local Hamiltonian problems for both classes of guiding states are QCMA-complete in the inverse-polynomial precision setting, but lie within NP (or NqP) in the constant precision regime when the guiding state is classically evaluatable.

Our completeness results show that, from a complexity-theoretic perspective, classical Ansätze selected by classical heuristics are just as powerful as quantum Ansätze prepared by quantum heuristics, as long as one has access to quantum phase estimation. In relation to the quantum PCP conjecture, we (i) define a complexity class capturing quantum-classical probabilistically checkable proof systems and show that it is contained in $\mathsf{BQP}^{\mathsf{NP}[1]}$ for constant proof queries; (ii) give a no-go result on 'dequantizing' the known quantum reduction which maps a QPCP-verification circuit to a local Hamiltonian with constant promise gap; (iii) give several no-go results for the existence of quantum gap amplification procedures that preserve certain ground state properties; and (iv) propose two conjectures that can be viewed as stronger versions of the NLTS theorem. Finally, we show that many of our results can be directly modified to obtain similar results for the class MA.

Contents

1 Introduction

Quantum chemistry and quantum many-body physics are generally regarded as two of the most promising application areas of quantum computing [Aar09, BBMC20]. Whilst perhaps the original vision of the early pioneers of quantum computing was to simulate the *time-dynamics* of quantum systems [Ben80, Fey18], for many applications one is interested in *stationary* properties. One particularly noteworthy quantity is the *ground state energy* (which corresponds to the smallest eigenvalue) of a local Hamiltonian describing a quantum mechanical system of interest, say a small molecule or segment of material. The precision to which one can estimate the ground state energy plays a crucial role in practice: for instance, in chemistry the relative energies of molecular configurations enter into the exponent of the term computing reaction rates, making the latter exceptionally sensitive to small (non-systematic) errors in energy calculations. Indeed, to match the accuracy obtained by experimentation for such values one aims for an accuracy that is smaller than so-called *chemical accuracy*, which is about 1.6 millihartree.[1] This quantity – which reads as a constant – is defined with respect to a (physical) Hamiltonian whose norm grows polynomially in the system size and particle dimension, and thus chemical accuracy is in fact a quantity that scales inverse polynomially in the system size when one considers (sub-)normalized Hamiltonians, which is often the case in the quantum computing / Hamiltonian complexity literature.

The problem of estimating the smallest eigenvalue of a local Hamiltonian up to some additive error (the decision variant of which is known as the *local Hamiltonian problem*) is well-known to be QMA-hard when the required accuracy scales inversely with a polynomial, where QMA is the quantum analogue of the class NP, also known as Quantum Merlin Arthur. Therefore, it is generally believed that, without any additional help or structure, quantum computers are not able to accurately estimate the smallest eigenvalues of general local Hamiltonians, and there is some evidence that this hardness carries over to those Hamiltonians relevant to chemistry and materials science [OIWF22]. A natural question to ask is then the following: how much 'extra help' needs to be provided in order to accurately estimate ground state energies using a quantum computer?

In the quantum chemistry community, it is often suggested that this extra help could come from a classical heuristic that first finds some form of *guiding state*: a classical description of a quantum state that can be used as an input to a quantum algorithm to compute the ground state energy accurately [LLS+22]. Concretely, this comes down to the following two-step procedure [GHLGM22]:

- Step 1 (Guiding state preparation): A classical heuristic algorithm is applied to obtain a *guiding state* $|\psi\rangle$, which is hoped to have 'good'[2] fidelity with the ground space.
- Step 2: (Ground state energy approximation): The guiding state $|\psi\rangle$ is used as input to Quantum Phase Estimation (QPE) to efficiently and accurately compute the corresponding ground state energy.

Step 2 of the above procedure can be formalised by the *Guided k-local Hamiltonian problem (k-GLH)*, which was introduced in [GLG22] and shown to be BQP-complete under certain parameter regimes that were subsequently improved and tightened in [GHLGM22, CFW22, CFG+23]. The problem k-GLH is stated informally as follows: given a k-local Hamiltonian H, an appropriate classical 'representation' of a guiding state $|u\rangle$ promised to have ζ-fidelity with the ground space of H, and real thresholds $b > a$, decide if the ground state energy of H lies above or below the interval $[a, b]$. In a series of works [GLG22, GHLGM22, CFW22, CFG+23],

[1] This quantity, which is ≈1 kcal/mol, is chosen to match the accuracy achieved by thermochemical experiments.

[2] 'Good' here means at least inverse polynomial in the number of qubits the Hamiltonian acts on.

it was shown that 2-GLH is BQP-complete for *inverse polynomial* precision and fidelity, i.e. $b - a \geq 1/poly(n)$ and $\zeta = 1 - \Omega(1/poly(n))$ respectively. In contrast, when $b - a \in \Theta(1)$ and $\zeta = \Omega(1)$, k-GLH can be efficiently solved *classically* by using a dequantised version of the quantum singular value transformation.

The GLH problem forms the starting point of this work. We study '*Merlinized*' versions of GLH – in which guiding states are no longer given as part of the input but instead are only promised to exist – and use these as a way to gain some insight into important theoretical questions in quantum chemistry and complexity theory. In the subsequent paragraphs, we introduce some of the motivating questions guiding the study of the complexity of these so-called 'guidable' local Hamiltonian problems.

Ansätze[3] for state preparation. Step 1 of the aforementioned two-step procedure generally requires one to have access to classical heuristics capable of finding guiding states whose energies can be estimated classically (as a metric to test whether candidate states are expected to be close to the actual ground state or not). Furthermore, these 'trial states' should also be preparable as quantum states on a quantum computer, so that they can be used as input to phase estimation in Step 2. In [GLG22], inspired by a line of works that focused on the dequantization of quantum machine learning algorithms [Tan19, CGL+20, JGS20], a particular notion of 'sampling-access' to the guiding state u is assumed. Specifically, it is assumed that one can both query the amplitude of arbitrary basis states, and additionally that one can sample basis states according to their l_2 norm with respect to the overall state u.[4] Whilst this can be a somewhat powerful model [CHM21], it is closely related to the assumption of QRAM access to classical data, and thus in the context of quantum machine learning (where such access is commonly assumed), it makes sense to compare quantum machine learning algorithms to classical algorithms with sampling access to rule out quantum speed-ups that come merely from having access to quantum states that are constructed from exponential-size classical data.

However, for quantum chemistry and quantum many-body applications, this type of access to quantum states seems to be somewhat artificial. From a theoretical perspective, one might wonder to what extent this sampling access model 'hides' some complexity, allowing classical algorithms to perform well on the problem when they otherwise would not.

Finally, one may ask whether the fact that the ground state preparation in Step 1 considers only *classical* heuristics might be too restrictive. *Quantum* heuristics for state preparation, such as variational quantum eigensolvers [TCC+22] and adiabatic state preparation techniques [AL18], have so far mostly been considered as quantum approaches within the NISQ era. However, one can argue that even in the fault-tolerant setting, such heuristics will likely still be viable approaches to state preparation, in particular when used in conjunction with Quantum Phase Estimation.

The quantum PCP conjecture. Arguably the most fundamental result in classical complexity theory is the Cook-Levin Theorem [Coo71, Lev73], which states that constraint satisfaction problems (CSPs) are NP-complete. The PCP theorem [ALM+98, AS98], which originated from a long line of research on the complexity of interactive proof systems, can be viewed as

[3] An Ansatz (plural Ansätze) is a German word often used in physics and mathematics for an assumption about the form of an unknown function or solution which is made in order to facilitate the solution of some problem. An Ansatz for state preparation refers to an assumption (restriction) on the states that are prepared on a quantum computer, for example to matrix product states or stabilizer states.

[4] In this work we slightly abuse notation by making a distinction between the vector representing a quantum state, which we will denote as 'u', and that same vector instantiated as a quantum state (e.g. living on a quantum computer), which we will denote by '$|u\rangle$'. Of course, these are the same mathematical object ($u = |u\rangle \in \mathcal{C}^{2^n}$), and we only use the different notation to make our theorem statements and proofs clearer.

a 'strengthening' of the Cook-Levin theorem. In its proof-checking form, it states that all decision problems in NP can be decided, with a constant probability of error, by only checking a constant number of bits of a polynomially long proof string y (selected randomly from the entries of y). There are also alternative equivalent formulations of the PCP theorem. One, due to Dinur [Din07], is in terms of *gap amplification*: it states that it remains NP-hard to decide whether an instance of CSP is either completely satisfiable, or whether no more than a constant fraction of its constraints can be satisfied. It is straightforward to show that this formulation is equivalent to the aforementioned proof-checking version.

Naturally, quantum complexity theorists have proposed proof-checking and gap amplification versions of PCP in the quantum setting. Given the close relationship between QMA and the local Hamiltonian problem, the most natural formulation is in terms of gap amplification: in this context, the *quantum* PCP conjecture roughly states that energy estimation of a (normalized) local Hamiltonian up to *constant* precision remains QMA-hard. This conjecture is arguably one of the most important open problems in quantum complexity theory and has remained unsolved for nearly two decades. In a recent breakthrough result, the NLTS conjecture was proven to be true, which (amongst other things) means that an important class of Ansätze (constant depth quantum circuits) are not expressive enough to estimate the ground state energies of all Hamiltonians up to even constant precision, which is a prerequisite for quantum PCP to hold [ABN22]. However, there have also been some no-go results: for example, a quantum PCP statement cannot hold for local Hamiltonians defined on a grid, nor on high-degree or expander graphs [BH13].

One way to shed light on the validity of the quantum PCP conjecture can be to study PCP-type conjectures for other 'Merlinized' complexity classes. Up until this point, PCP-type conjectures have not been considered for other classes besides NP and QMA.[5] However, there is the beautiful result of [AG19], which studies the possibility of a gap amplification procedure for the class MA by considering a particular type of Hamiltonian: uniform stoquastic local Hamiltonians. The authors show that deciding whether the energy of such a Hamiltonian is exactly zero or inverse polynomially bounded away from zero is MA-hard, but that the problem is in NP when this interval is increased to be some constant. Consequently, this implies that there can exist a gap-amplification procedure for uniform stoquastic Local Hamiltonians (in analogy to the gap amplification procedure for constraint satisfaction problems in the original PCP theorem) if and only if $\mathsf{MA} = \mathsf{NP}$ – i.e. if MA can be derandomized. Since $\mathsf{MA} \subseteq \mathsf{QMA}$, this result also shows that if a gap amplification procedure for the general local Hamiltonian problem would exist that 'preserves stoquasticity', then it could also be used to derandomize MA.

1.1 Summary of main results

1.1.1 Completeness results for the guidable local Hamiltonian problem

Inspired by classical heuristics that work with Ansätze to approximate the ground states of local Hamiltonians, we define a general class of states that we call *classically evaluatable and quantumly preparable*.

Definition 1.1 (Informal) (Classically evaluatable and quantumly preparable states, from Definition 3.2)**.** *We say that an n-qubit state u is classically evaluatable if*

(i) it has an efficient classical description which requires at most a polynomial number of bits to write down and

[5] This is barring a result by Drucker which proves a PCP theorem for the class AM [Dru11]; though there is no direct relationship between QMA and AM and hence it is not clear whether this gives any intuition about the likely validity of the quantum PCP conjecture.

(ii) one can, given such a description, classically efficiently compute expectation values of $\mathcal{O}(\log n)$-local observables of u.

In addition, we say that the state is also quantumly preparable if (iii) there exists a quantum circuit that prepares u as a quantum state $|u\rangle$ using only a polynomial number of two-qubit gates.

In the main text we consider a more general version of the definition above, which also allows for probabilistic estimation of expectation values, and we provide four concrete examples of Ansätze that satisfy all three conditions: matrix product states (MPS), stabilizer states, constant-depth quantum circuits and IQP circuits. We also relate classically evaluatable states to the samplable states of [GLG22], and show that if one allows for an error in the estimation of local observables, it forms in fact a larger class of quantum states (Theorem 3.1).

Our main focus is on a new family of local Hamiltonian problems, which we call *Guidable local Hamiltonian problems*, in which we are promised that the ground state is close (with respect to fidelity) to a certain state from a particular class of states.

Definition 1.2 (Informal) (Guidable Local Hamiltonian problems, from Definition 3.3)**.** *Guidable Local Hamiltonian Problems are a class of problems defined by having the following input, promise, either extra promise 1 or extra promise 2, and output:*
Input: *A k-local Hamiltonian H with $\|H\| \leq 1$ acting on n qubits, threshold parameters $a, b \in \mathbb{R}$ such that $b - a \geq \delta > 0$ and a fidelity parameter $\zeta \in (0, 1]$.*
Promise: *We have that either $\lambda_0(H) \leq a$ or $\lambda_0(H) \geq b$ holds, where $\lambda_0(H)$ denotes the ground state energy of H.*
Extra promises: *Let Π_{gs} be the projection on the subspace spanned by the ground states of H. Then for each problem class, we have that either one of the following promises hold:*

1. ***Classically Guidable and Quantumly Preparable k-LH*** *($\mathsf{CGaLH}^*(k, \delta, \zeta)$): there exists a classically evaluatable and quantumly preparable state $u \in \mathbb{C}^{2^n}$ for which $\|\Pi_{gs} u\|^2 \geq \zeta$.*
2. ***Quantumly Guidable k-LH*** *($\mathsf{QGaLH}(k, \delta, \zeta)$): There exists a quantum circuit of polynomially many two-qubit gates that produces the state $|\phi\rangle$ for which $\|\Pi_{gs} |\phi\rangle\|^2 \geq \zeta$.*

Output:

- *If $\lambda_0(H) \leq a$, output* YES.
- *If $\lambda_0(H) \geq b$, output* NO.

A guidable local Hamiltonian problem variant for a different class of guiding states was already introduced in [GLG22] without giving any hardness results. Using techniques from Hamiltonian complexity we obtain the following completeness results.[6]

Theorem 1.1 (Informal) (Complexity of guidable local Hamiltonian problems, from Corollary 4.1 and Theorem 4.2)**.** *For constant k and $\delta = 1/\,poly(n)$, we have that both $\mathsf{CGaLH}^*(k, \delta, \zeta)$ and $\mathsf{QGaLH}(k, \delta, \zeta)$ are QCMA-complete when $\zeta \in (\Omega(1/\,poly(n)), 1 - \Omega(1/\,poly(n)))$.*

We also obtain similar complexity results for a guidable version of the quantum satisfiability problem (see Appendix D). A direct corollary of the above theorem is the following.

Corollary 1.1 (Classical versus quantum state preparation)**.** *When one has access to a quantum computer (and in particular quantum phase estimation), then having the ability to prepare any quantum state preparable by a polynomially-sized quantum circuit is no more powerful than the ability to prepare states from the family of classically evaluatable and quantumly preparable states, when the task is to decide the local Hamiltonian problem with precision $\Theta(1/\,poly(n))$.*

[6] In fact $\mathsf{QGaLH}(k, \delta, \zeta)$ remains QCMA-hard all the way up to $\zeta = 1$.

It should be noted that our result does *not* imply that all Hamiltonians which have efficiently quantumly preparable guiding states also necessarily have guiding states that are classically evaluatable. All this result says is that for any instance of the guidable local Hamiltonian problem with the promise that there exist guiding states that can be efficiently prepared by a quantum computer, there exists an (efficient) *mapping* to another instance of the guidable local Hamiltonian problem with the promise that there exist guiding states that are classically evaluatable and quantumly preparable. Whilst this reduction is efficient in the complexity-theoretic sense, it might not be for practical purposes, as it would likely remove all the structure present in the original Hamiltonian. Hence, the main implication of our result is not that these kinds of reductions are of practical merit, but that they provide some theoretical evidence as to why the aforementioned classical-quantum hybrid approach of guiding state selection through *classical* heuristics combined with *quantum* energy estimation might indeed be a promising quantum approach to quantum chemistry and quantum many-body physics.

We complement our quantum hardness results with classical containment results (of the classically guidable local Hamiltonian problem), obtained through a deterministic dequantized version of Lin and Tong's ground state energy estimation algorithm [LT20]. Here CGaLH is just as CGaLH* but without the promise of the guiding state being classically preparable (see Definition 3.3 in the main text).

Theorem 1.2 (Informal) (Classical containment of the classically guidable local Hamiltonian problem, from Theorem 5.2.)**.** *Let $k = \mathcal{O}(\log n)$. When δ is constant, we have that* CGaLH(k, δ, ζ) *is in* NP *when when ζ is constant and is in* NqP *when $\zeta = \Omega(1/\,poly(n))$. Here* NqP *is just as* NP *but with the verifier circuit being allowed to run in quasi-polynomial time.*

Through a more careful analysis of when exactly the quantum hardness vanishes, the picture of Figure 1 emerges, which characterises the complexity of CGaLH$^*(k, \delta, \zeta)$ for relevant parameter settings in the desired precision and promise on the fidelity.

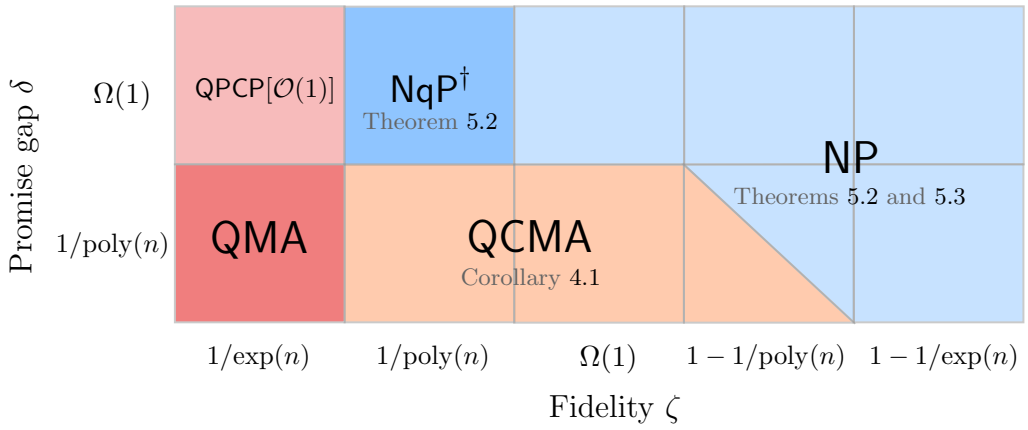

Figure 1: Complexity characterization of CGaLH$^*(k, \delta, \zeta)$ over parameter regime δ and ζ, for $k = \mathcal{O}(1)$. Any classification indicates completeness for the respective complexity class, except for NqP, for which we only know containment (indicated by the '†'). Here completeness for certain parameter combinations means that for all functions of the indicated form, the problem is contained in the complexity class, and for a subset of these functions the problem is also hard. The results for QPCP$[\mathcal{O}(1)]$ and QMA follow directly from [AALV09] and [KSV02].

1.1.2 Quantum-classical probabilistically checkable proofs

We introduce the notion of a *quantum-classical probabilistically checkable proof system* in the following way.

Definition 1.3 (Informal) (Quantum Classical PCP, from Definition 6.3)**.** *A* $\mathsf{QCPCP}[q]$ *protocol consists of a polynomial-time quantum verifier* V *that uses* $poly(n)$ *ancilla qubits and is given an input* $x \in \{0,1\}^n$ *and a classical proof* $y \in \{0,1\}^{p(n)}$*, where* $p(n) \leq poly(n)$*, from which it queries at most* q *bits non-adaptively. The verifier measures the first qubit and accepts only if the outcome is* $|1\rangle$*. A promise problem* $A = (A_{yes}, A_{no})$ *belongs to the class* $\mathsf{QCPCP}[q]$ *if it has a* $\mathsf{QCPCP}[q]$ *verifier system with the following properties*

> ***Completeness.*** *If* $x \in A_{yes}$*, then there exists a classical proof* y *such that the verifier accepts with probability at least* $2/3$*.*
>
> ***Soundness.*** *If* $x \in A_{no}$*, then for all classical proofs* y *the verifier accepts with probability at most* $1/3$*.*

Note that we have that $\mathsf{QCPCP}[\mathcal{O}(1)]$ trivially contains CGaLH with any constant promise gap and any constant fidelity,[7] since we have shown that this problem is in NP (Theorem 5.2) and therefore admits a classical PCP system to solve the problem. Note that replacing the classically evaluatable states with samplable states in the promise, as in [GLG22], does not necessarily mean that the problem is in $\mathsf{QCPCP}[\mathcal{O}(1)]$ as we do not know whether MA admits a (quantum-classical) PCP.

We first prove two basic facts about QCPCP: we show that it allows for weak error reduction (Proposition 6.1) and that the non-adaptiveness restriction does not limit the power of the class when the number of proof queries is constant (Theorem 6.1). Our 'quantum-classical PCP conjecture' then posits that $\mathsf{QCPCP}[\mathcal{O}(1)] = \mathsf{QCMA}$, analogously to the quantum PCP conjecture which states that $\mathsf{QPCP}[\mathcal{O}(1)] = \mathsf{QMA}$ (here $\mathsf{QPCP}[q]$ denotes the complexity class associated with *quantum* probabilistically checkable proof systems).[8]

Our main result regarding $\mathsf{QCPCP}[\mathcal{O}(1)]$ is that we can provide a non-trivial upper bound on the complexity of the class.

Theorem 1.3 (Informal) (Upper bound on QCPCP, from Theorem 6.2)**.**

$$\mathsf{QCPCP}[\mathcal{O}(1)] \subseteq \mathsf{BQP}^{\mathsf{NP}[1]}.$$

Here $\mathsf{BQP}^{\mathsf{NP}[1]}$ is the class of all problems that can be solved by a BQP-verifier that makes a single query to an NP-oracle. The key idea behind the proof is that a quantum reduction can be used to transform a QCPCP verification circuit to a local Hamiltonian that is *diagonal* in the computational basis, and thus can be solved with a single query to an NP oracle. Using this upper bound, we then show that if our quantum-classical PCP conjecture is true, then $\mathsf{NP}^{\mathsf{BQP}} \subseteq \mathsf{BQP}^{\mathsf{NP}}$. This inclusion would have the effect that one could show that $\mathsf{NP} \subseteq \mathsf{BQP} \implies \mathsf{PH} \subseteq \mathsf{BQP}$, i.e. that if NP is contained in BQP then so is the entire polynomial hierarchy.[9] Such a result would provide strong evidence that quantum computers are indeed not capable of solving NP-hard problems.

1.1.3 Three implications for the quantum PCP conjecture

Finally, we use our obtained results on QCPCP and CGaLH to obtain two interesting results and a new conjecture with respect to the quantum PCP conjecture. First, we give evidence

[7] Recall that the problem is QCMA-hard when the promise gap is inverse polynomial in the number of Hamiltonian terms instead, even when the fidelity is constant (but < 1).

[8] The question of whether a PCP can be shown for QCMA was also raised briefly in [BGK22].

[9] This would contrast the result by [AIK21] which shows that $\mathsf{NP}^{\mathsf{BQP}} \not\subset \mathsf{BQP}^{\mathsf{NP}}$ relative to an oracle. However, since our inclusion goes via $\mathsf{QCMA} \subseteq \mathsf{QCPCP}[\mathcal{O}(1)]$, which would likely require non-relativizing techniques just as was the case for the classical PCP Theorem, the conjecture and this result could simultaneously be true.

that it is unlikely that there exists a *classical* reduction from a QPCP-system (see Definition 6.2 for a formal definition) to a local Hamiltonian problem with a constant promise gap having the same properties as the known *quantum* reduction (see for example [Gri18]), unless $\mathsf{BQP} \subseteq \mathsf{QCPCP}[\mathcal{O}(1)] \subseteq \mathsf{NP}$, something that is not expected to hold [Aar10, RT22].

Theorem 1.4 (Informal) (No-go for classical polynomial-time reductions, from Theorem 7.1)**.** *For all constant $\epsilon < 1/6$ there cannot exist a classical polynomial-time reduction from a $\mathsf{QPCP}[\mathcal{O}(1)]$ verification circuit V to a local Hamiltonian H such that, given a proof $|\psi\rangle$,*

$$|\mathbb{P}[V \text{ accepts } |\psi\rangle] - (1 - \langle\psi| H |\psi\rangle)| \leq \epsilon,$$

unless $\mathsf{QCPCP}[\mathcal{O}(1)] \subseteq \mathsf{NP}$ (which would imply $\mathsf{BQP} \subseteq \mathsf{NP}$).

This provides strong evidence that allowing for reductions to be quantum is indeed necessary to show equivalence between the gap amplification and proof verification formulations of the quantum PCP conjecture [AAV13].

Second, our classical containment results of CGaLH^* with constant promise gap can be viewed as no-go theorems for a gap amplification procedure for QPCP having certain properties, as illustrated by the following result.

Theorem 1.5 (Informal) (No-go results for Hamiltonian gap amplification, from Theorem 7.2)**.** *There cannot exist a gap amplification procedure for the local Hamiltonian problem that preserves the fidelity between the ground space of the Hamiltonian and any classically evaluatable state up to a*

- *multiplicative constant, unless $\mathsf{QCMA} = \mathsf{NP}$, or*
- *multiplicative inverse polynomial, unless $\mathsf{QCMA} \subseteq \mathsf{NqP}$.*

This result is analogous to the result of [AG19], which rules out a gap amplification procedure that preserves stoquasticity under the assumption that $\mathsf{MA} \neq \mathsf{NP}$.[10] Moreover, we point out that many Hamiltonian gadget constructions *do* satisfy such fidelity-preserving conditions, and indeed are precisely those that were used in [GHLGM22] and [CFW22] to improve the hardness results for the guided local Hamiltonian problem.[11] We obtain similar results for the class MA by considering a variant of CGaLH that restricts the Hamiltonian to be stoquastic (Appendix C).

Third, we can use our results to formulate a stronger version of the NLTS theorem (and an alternative to the NLSS conjecture [GLG22]), which we will call the *No Low-energy Classically evaluatable States conjecture*. This conjecture can hopefully provide a new stepping stone towards proving the quantum PCP conjecture.

Conjecture 1.1 (Informal) (NLCES conjecture, from Conjecture 7.1)**.** *There exists a family of local Hamiltonians $\{H_n\}_{n\in\mathbb{N}}$ on n qubits, and a constant $\beta > 0$, such that for every classically describable state $u \in \mathbb{C}^{2^n}$ as per Definition 3.2, we have for sufficiently large n that*

$$\langle u| H |u\rangle \geq \lambda_0(H_n) + \beta\,.$$

[10] Or taking a different view, proving the existence of such gap amplifications would allow one to simultaneously prove that MA can be derandomized (or even RP if it exhibits some additional properties) [AG19].

[11] For a quantum version of gap amplification one would typically expect locality-reducing Hamiltonian gadgets as part of the procedure, to compensate for a "powering step" which consists of taking powers of the Hamiltonian (which therefore increases locality). It is already known that the current best-known locality-reducing gadgets [BDLT08] cannot be used because they increase the norm of the Hamiltonian by a constant factor, which results in an unmanageable decrease of the relative promise gap. Our result shows that even if one would find better constructions that don't have this effect, they would still have to satisfy the additional constraints as described in Theorem 7.2.

Just as is the case for the NLSS conjecture and the NLTS theorem, the NLCES conjecture would, if proven to be true, not necessarily imply the quantum PCP conjecture. For example, it might be that there exist states that can be efficiently described classically but for which computing expectation values is hard (just as, for example, tensor network contraction is #P-hard in the worst case [SWVC07, BMT15]). Furthermore, as we have shown in this work, states with high energy but also a large fidelity with the ground state suffice as witnesses to decision problems on Hamiltonian energies, and these would not be excluded by a proof of the NLCES conjecture above. To make this more concrete, we also formulate an even stronger version of the NLCES conjecture, which states that there must be a family of Hamiltonians, for which no classically evaluatable state has good fidelity with the low energy spectrum (Conjecture 7.2).

1.2 Overview of techniques

QCMA-hardness proof for guidable local Hamiltonian problems. We follow a similar proof structure as used in the BQP-hardness proofs of the Guided Local Hamiltonian problem [GLG22, GHLGM22, CFW22, CFG+23]. This construction begins with a BQP circuit $V = V_{T-1} \dots V_0$ and applies Kitaev's circuit-to-Hamiltonian construction, which transforms V to a 5-local Hamiltonian $H = H_{\text{in}} + H_{\text{clock}} + H_{\text{prop}} + H_{\text{out}}$. By applying several tricks one is able to ensure that the ground state of the Hamiltonian has a non-negligible fidelity with a certain classical guiding state. However, at this point, there are several obstacles which prevent one from directly adopting the same proof in the QCMA setting, i.e. when starting with a QCMA verification circuit U. This mostly comes down to the fact that U, unlike a BQP-circuit, has an additional input register for the witness. This creates many valid 'history states' (which are 0-eigenvectors of the Hamiltonian $H_{\text{in}} + H_{\text{clock}} + H_{\text{prop}}$), giving us less control over and knowledge about the actual ground state of the Hamiltonian generated by the circuit-to-Hamiltonian construction.

To work around this, we use several tricks in our new construction. First, we use the CNOT-trick, introduced in [WJB03], to 'force' all witnesses to be classical. Second, we show that there exists a randomized reduction from a QCMA protocol with verification circuit U to one in which the verification circuit $\tilde{U}$ satisfies both (i) that there exists a unique accepting witness in the YES-case and (ii) this witness is accepted with probability 1. By applying a 'majority vote'-type of error reduction, we can ensure that the acceptance probabilities of all other witnesses are suppressed exponentially close to zero. Next, we apply the small-penalty circuit-to-Hamiltonian mapping of [DGF22], which allows us to control the bounds on the energies in the low-energy subspace of the Hamiltonian. Combining this with the aforementioned randomized reduction, we find that the ground space of H is now one-dimensional and consists only of the history state corresponding to the uniquely accepting witness in the YES-case. This allows us to construct a corresponding semi-classical subset state (which we show to be classically evaluatable and quantumly preparable) that has good fidelity with this history state, and use this as our guiding state. We also apply the pre-idling and block-encoding tricks from [GLG22] to increase fidelity with the guiding state and to handle the NO-case, respectively. Finally, by using locality-reducing Hamiltonian gadget constructions that preserve the 'classical evaluatibility' of the guiding state, we arrive at our final result.

A deterministic spectral amplification algorithm. Our classical algorithm is inspired by the techniques developed in [GLG22], which in a more general setting dequantizes the quantum singular value transformation [GSLW19] for sparse matrices. Our technique can essentially be viewed as a dequantization of Lin and Tong's ground state energy estimation algorithm [LT20]: here one assumes access to a unitary U_H that implements a block-encoding of H. Since H is Hermitian, a polynomial function applied to H can be viewed as acting on the eigenvalues of H. An approximate low-energy subspace projector can then be constructed using a polynomial which approximates the sign function, using a result from [HC17]. We construct a similar

algorithm, but this time in a classical deterministic setting, assuming the input states of being of the form of classically evaluatable states (see Section 1.1). We measure the complexity of our dequantization algorithm by counting the number of expectation values of local observables that have to be computed, which follows straightforwardly from applying the multinomial theorem to the polynomial approximation we consider. Finally, we derive the complexity of the algorithm when it is applied to solving the Hamiltonian energy decision problems considered in this paper.

1.2.1 Relation to previous work

The starting point of this paper is the *guided* local Hamiltonian problem, introduced in [GLG22]. Our work diverges from theirs in two principle directions: 1) whereas their work focuses predominantly on the case in which a guiding state is given as a part of the input, we focus here on the 'guidable' version of the problem – i.e. when a guiding state is only *promised* to exist[12]; and 2) we consider a more general and natural family of guiding states, namely what we term *classically evaluatable* states.

Direction 1) allows us to introduce and consider the idea of a quantum-classical PCP (QCPCP) conjecture and, combined with our results on the hardness of the guidable local Hamiltonian problem, obtain results about the relationship between problems admitting QCPCPs and other computational complexity classes. The move from 'guided' to 'guidable' here is necessary: without the notion of a witness, a PCP framework cannot be considered.

Direction 2) puts our results in a more general setting, and in particular one that is somewhat more relevant to questions surrounding the application of quantum computers to hard problems in chemistry and physics. In particular, as we show in Section 3, our notion of classically evaluatable states captures many Ansätze commonly used for estimating ground state energies of physically relevant Hamiltonians. Moreover, the set of ϵ-classically evaluatable states captures a larger set of states than those that are sampleable, as we discuss in Section 3. Moving to this class of states allows us to weaken the fairly stringent assumptions of the original guided local Hamiltonian problem.

In [WJB03], the authors consider a QCMA-complete problem of a similar flavour to ours, namely deciding whether a 3-local Hamiltonian has low energy states that can be prepared using a polynomially bounded number of elementary quantum gates. Apart from the specificity of the requirement (preparable via polynomial-time quantum circuits vs. classically evaluatable), this differs in an important way from the type of constraint that we consider in this work: that the requirement on the ground states of the Hamiltonian is regarding their *fidelity* with a particular class of quantum states, not that they themselves belong to that class. We elaborate more on these differences at the end of Section 3.

1.3 Open questions and future work

A non-trivial lower bound on quantum PCP. A trivial lower bound on the computational power of quantum PCP is NP, which follows from the standard PCP theorem. Our formulation of the QCPCP provides a way to prove the first non-trivial lower bound on quantum PCP. Since the proofs in QCPCP are always (or can be forced to be) classical, one might hope to do this by using some of the techniques used to prove (formulations of the) PCP theorem, like exponentially long PCPs, PCPs of proximity, alphabet reduction etc., which could carry over more easily to the QCPCP setting as compared to QPCP.

Proof checking versus the local Hamiltonian formulations of quantum PCPs. Another obvious lower bound to QPCP (or QCPCP) comes from BQP $\subseteq$ QPCP, since the verifier

[12]This was briefly touched upon in [GLG22]), where it was shown that the local Hamiltonian problem for all Hamiltonians whose ground space has constant fidelity with samplable states can be estimated up to constant precision in MA, but without any hardness results.

can simply ignore the proof. However, the relationship between BQP and NP is very much unclear: it is generally believed that for both classes there exist problems that are exclusively contained in only one of them. In this work we show that it is unlikely that there exists a classical reduction from a QPCP verifier circuit to a local Hamiltonian problem with a constant promise gap that has the same properties as the known quantum reduction. This means that it is entirely possible that the generic local Hamiltonian problem with constant promise gap is contained NP, whilst the *proof checking* version of QPCP is not, provided that the quantum reduction from the proof checking formulation to the local Hamiltonian problem can indeed not be 'dequantized'.[13] That is, despite results that show 'equivalence' of the proof-checking and local Hamiltonian variants of the quantum PCP conjecture, the two variants could actually have quite different computational power since equivalence is shown only under quantum reductions. It would be interesting to explore the possibility of different complexities for the proof checking and local Hamiltonian variants of QPCP further.

The (strong) NLCES conjecture. It would be interesting to see whether the family of Hamiltonians used to prove the well-known NLTS conjecture, or constructions inspired by the proof thereof (in particular Hamiltonians that arise from error-correcting codes), can also be used to prove (weaker versions of) our NLCES conjecture (see Conjecture 7.1). Note that our NLCES conjecture is strictly stronger than NLTS, since it includes all states that can be prepared by constant depth quantum circuits (i.e. those states covered by the NLTS conjecture), but also includes states that require super-constant quantum depth, for example arbitrary Clifford circuits[14], matrix-product states, etc.

MA **containment of guidable stoquastic LH.** It is well-known that for stoquastic Hamiltonians, deciding if the ground state energy is $\leq a$ or $\geq b$ with $b - a = \Omega(1/\,poly(n))$ is StoqMA-complete, for arbitrary $b \geq a$ inverse polynomially separated, but MA-complete when $a = 0$ and $b = 1/\,poly(n)$ [AGL20]. In [Bra15], it is shown that for a much stronger type of assumption on the existence of a guiding state than what we consider, the problem is also MA-complete for arbitrary $b \geq a$ inverse polynomially separated. Showing MA-containment for our definition of *guidable* stoquastic local Hamiltonian problems (with arbitrary a, b, inverse polynomially separated) could provide a way to study the exact relationship between StoqMA and MA.

The classical guiding state existence assumption. As discussed in [GHLGM22, CFW22], the existence of practical quantum advantage based on the previously mentioned two-step procedure is only expected if there exist guiding states, quantum or classical, that have not too much (exponentially close) but also not too little (exponentially small) fidelity with the ground space of the Hamiltonian under study. Whilst there is some literature that (partially) explores this direction [BLH+21, TME+18, LLZ+22], it would be useful and interesting to study this assumption in the special case of Ansätze that describe classically evaluatable and quantumly preparable states. This could provide numerical evidence to support the results that we have shown from a complexity-theoretic perspective: that classical heuristics combined with quantum phase estimation is indeed the right way to approach fault-tolerant quantum advantage in chemistry.

[13] Indeed, it could be that the local Hamiltonian problem with constant promise gap is contained in some complexity class C. Then so long as BQP $\subsetneq$ C, it is possible that the proof checking version of QCPCP is strictly more powerful than the local Hamiltonian version (i.e. QCPCP $\subsetneq$ C), since the quantum reduction cannot necessarily be performed 'inside' C.

[14] This has in fact recently been proven for Clifford circuits, see [CCNN23].

2 Preliminaries

2.1 Notation

We write $\lambda_i(A)$ to denote the ith eigenvalue of a Hermitian matrix A, ordered in non-decreasing order, with $\lambda_0(A)$ denoting the smallest eigenvalue (ground state energy). When we write $\|\cdot\|$ we refer to the operator norm when its input is a matrix, and Euclidean norm for a vector.

2.2 Some basic definitions and results from complexity theory

Let us first recall a couple of basic definitions and results from (quantum) complexity theory, which is central to this work. All complexity classes will be defined with respect to promise problems (and not languages). To this end, we take a (promise) problem $A = (A_{\text{yes}}, A_{\text{no}})$ to consist of two non-intersecting sets $A_{\text{yes}}, A_{\text{no}} \subseteq \{0,1\}^*$ (the YES and NO instances, respectively). We have that $A_{\text{inv}} = \{0,1\}^* \setminus A_{\text{yes}} \cup A_{\text{no}}$ is the set of all invalid instances, and we do not care how a class behaves on problem instances $x \in A_{\text{inv}}$ (it can accept or reject arbitrarily, see the paragraph 'oracle access' for a more elaborate discussion on what this entails).

Definition 2.1 (P)**.** *A promise problem $A = (A_{yes}, A_{no})$ is in* P *if and only if there exists a polynomial-time verifier circuit V which takes as input a string $x \in \{0,1\}^*$ and decides on acceptance or rejection of x such that*

- *if $x \in A_{yes}$ then $V(x)$ accepts.*
- *if $x \in A_{no}$ then $V(x)$ rejects.*

Definition 2.2 (NP)**.** *A promise problem $A = (A_{yes}, A_{no})$ is in* NP *if and only if there exists a polynomial-time verifier circuit V and a polynomial p, where V takes as input a string $x \in \{0,1\}^*$ and a $p(|x|)$-bit witness y and decides on acceptance or rejection of x such that*

- *if $x \in A_{yes}$ then there exists a y such that $V(x,y)$ accepts.*
- *if $x \in A_{no}$ then for every y we have that $V(x,y)$ rejects.*

Definition 2.2a (NqP)**.** *If the above verifier circuit V is instead allowed to run in quasi-polynomial time, i.e. $2^{\mathcal{O}(\log^c(n))}$ for some constant $c > 0$, we talk about the complexity class* NqP *(Non-deterministic quasi-Polynomial time).*

Definition 2.3 (MA)**.** *A promise problem $A = (A_{yes}, A_{no})$ is in* MA$[c,s]$ *if and only if there exists a probabilistic polynomial-time verifier circuit V and a polynomial p, where V takes as input a string $x \in \{0,1\}^*$ and a $p(|x|)$-bit witness y and decides on acceptance or rejection of x such that*

- *if $x \in A_{yes}$ then there exists a y such that $V(x,y)$ accepts with probability $\geq c$,*
- *if $x \in A_{no}$ then for every y we have that $V(x,y)$ accepts with probability $\leq s$,*

where $c - s = \Omega(1/\,poly(n))$. When $c = 2/3$ and $s = 1/3$ we omit the $[c,s]$ notation and call the class MA*.*

Definition 2.3a (UMA)**.** *The class* UMA$[c,s]$ *has the same definition as* MA *but with the extra constraint that if $x \in A_{yes}$ then there exists only a single y^* such that $V(x,y^*)$ accepts with probability $\geq c(= 2/3)$, and otherwise for all $y \neq y^*$ we have that $V(x,y)$ accepts with probability $\leq s(= 1/3)$.*

Definition 2.4 (QMA)**.** *A promise problem $A = (A_{yes}, A_{no})$ is in* QMA*[c,s] if and only if there exists a quantum polynomial-time verifier circuit V and a polynomial $p : \mathbb{N} \to \mathbb{N}$, where V takes as input a string $x \in \{0,1\}^*$ and a $p(|x|)$-qubit witness quantum state $|\psi\rangle$ and decides on acceptance or rejection of x such that*

- *if $x \in A_{yes}$ then there exists a poly(n)-qubit witness state $|\psi_x\rangle$ such that $V(x, |\psi_x\rangle)$ accepts with probability $\geq c$,*
- *if $x \in A_{no}$ then for every purported poly(n)-qubit witness state $|\psi\rangle$, $V(x, |\psi\rangle)$ accepts with probability $\leq s$,*

where $c - s = \Omega(1/\,poly(n))$. If $c = 2/3$ and $s = 1/3$, we abbreviate to QMA.

Definition 2.5 (QCMA)**.** *A promise problem $A = (A_{yes}, A_{no})$ is in* QCMA*[c,s] if and only if there exists a quantum polynomial-time verifier circuit V and a polynomial $p : \mathbb{N} \to \mathbb{N}$, where V takes as input a string $x \in \{0,1\}^*$ and a $p(|x|)$-qubit witness computational basis state $|y\rangle$ and decides on acceptance or rejection of x such that*

- *if $x \in A_{yes}$ then there exists a poly(n)-qubit computational basis state $|y\rangle$ such that $V(x, |y\rangle)$ accepts with probability $\geq c$,*
- *if $x \in A_{no}$ then for every purported poly(n)-qubit computational basis state $|y\rangle$, $V(x, |y\rangle)$ accepts with probability $\leq s$,*

where $c - s = \Omega(1/\,poly(n))$. If $c = 2/3$ and $s = 1/3$, we abbreviate to QCMA.

Definition 2.5a (UQCMA)**.** *The class* UQCMA$[c, s]$ *has the same definition as* QCMA *but with the extra constraint that if $x \in A_{yes}$ then there exists only a single poly(n)-qubit witness computational basis state $|y^*\rangle$ such that $V(x, |y^*\rangle)$ accepts with probability $\geq c(= 2/3)$, and otherwise for all $|y\rangle \neq |y^*\rangle$ $V(x, |y\rangle)$ accepts with probability $\leq s(= 1/3)$.*

Unlike QMA, it is known that a lot of the behaviours exhibited by the classical complexity classes NP and MA hold for QCMA as well. An example of this, and one that we use later, is a result from [ABOBS22] stating that there exists a randomized reduction from QCMA to UQCMA.

Lemma 2.1 (Randomized reduction from QCMA to UQCMA [ABOBS22])**.** *Let $\langle U, p_1, p_2\rangle$ describe a promise problem in* QCMA*, where U is the description of a quantum circuit which takes an input x of length $|x| = n$, a witness y with length $|y| = poly(n)$. Denote p_1 and p_2 with $p_1 - p_2 = \Omega(1/\,poly(n))$ for the soundness and completeness parameters, respectively. Then there exists a randomized reduction to a* UQCMA *instance $\langle \tilde{U}, \tilde{p_1}, \tilde{p_2}\rangle$, such that if there exists a witness y s.t. $\langle U, p_1, p_2\rangle$ accepts with probability larger than p_1 then there exists a single y^* s.t. $\langle \tilde{U}, \tilde{p_1}, \tilde{p_2}\rangle$ accepts with probability $\geq \tilde{p_1}$, and accepts all other y with probability $\leq \tilde{p_2}$, and if $\langle U, p_1, p_2\rangle$ accepts with probability $\leq p_2$ for all y then $\langle \tilde{U}, \tilde{p_1}, \tilde{p_2}\rangle$ accepts with probability $\leq \tilde{p_2}$ for all y. This randomized reduction succeeds with probability $\Omega(1/|y|)$.*

Another one of these properties is the equivalence of one-sided and two-sided error in the acceptance and rejectance probabilities, which just as in the MA setting holds for QCMA (assuming robustness under the choice of the universal gate-set that is used to construct the verification circuits). Formally, this is established via the following lemma.

Lemma 2.2 (Perfect completeness QCMA [JKNN12])**.** *Let $\mathcal{G} = \{H, X, Toffoli\}$ be a fixed gate set. For any $c, s \in [0,1]$ satisfying $c - s \geq 1/q$ for some polynomial $q(n) : \mathbb{N} \to \mathbb{R}_{>0}$, we have that*

$$\mathsf{QCMA}_{\mathcal{G}}[c, s] \subseteq \mathsf{QCMA}_{\mathcal{G}}[1, s'],$$

where $s' = \frac{1}{2}(1 - (c - s))\left(1 + (1 + c - s)^2\right) < 1$.

Oracle access For a (promise) class $\mathcal{C}$ with complete (promise) problem A, the class $\mathsf{P}^{\mathcal{C}} = \mathsf{P}^{A}$ is the class of all (promise) problems that can be decided by a polynomial-time verifier circuit V with the ability to query an oracle for A. If V makes invalid queries (i.e. $x \in A_{\text{inv}}$), the oracle may respond arbitrarily. However, since V is deterministic, it is required to output the same final answer regardless of how such invalid queries are answered [GY19, Gol06]. Hence, the answer to any query outside of the promise set should not influence the final output bit. For a function f, we define $\mathsf{P}^{\mathcal{C}[f]}$ to be just as $\mathsf{P}^{\mathcal{C}}$ but with the additional restriction that V may ask at most $f(n)$ queries on an input of length n.[15] One defines $\mathsf{NP}^{\mathcal{C}}$ or $\mathsf{NP}^{\mathcal{C}[f]}$ in the same way but replacing the polynomial-time deterministic verifier V by a nondeterministic polynomial-time verifier V', taking an additional input $y \in \{0,1\}^{p(n)}$ for some polynomial $p(n)$.

2.3 Locality reducing perturbative gadgets

Perturbative gadgets are standard techniques from the Hamiltonian complexity toolbox and are used to transform one Hamiltonian into another whilst approximately preserving the (low-energy) spectrum. We will use such gadgets here, and will be particularly interested in those that preserve not only the low-energy spectrum of the original Hamiltonian, but also the structure of the low-energy eigenstates. In [CMP18], the authors introduce the following definition of simulation, and demonstrate via the use of perturbative gadgets that there are families of Hamiltonians which can be 'reduced' to different families of Hamiltonians with simpler/lower locality interactions. Note that these results originally only applied to qubits, but can be extended to qudits [PM21].

Definition 2.6 (Approximate Hamiltonian simulation [CMP18])**.** *We say that an m-qubit Hamiltonian H' is a (Δ, η, ϵ)-simulation of a n-qubit Hamiltonian H if there exists a local encoding $\mathcal{E}(H) = V(H \otimes P + \bar{H} \otimes Q)V^\dagger$ such that*

1. *There exists an encoding $\tilde{\mathcal{E}}(H) = \tilde{V}(H \otimes P + \bar{H} \otimes Q)\tilde{V}^\dagger$ such that $\tilde{\mathcal{E}}(\mathbb{1}) = P_{\leq \Delta(H')}$ and $\|V - V'\| \leq \eta$, where $P_{\leq \Delta(H')}$ is the projector onto the subspace spanned by eigenvectors of H' with eigenvalue below Δ,*
2. *$\left\| H'_{\leq \Delta} - \tilde{\mathcal{E}}(H) \right\| \leq \epsilon$, where $H'_{\leq \Delta} := P_{\leq \Delta(H')} H'$.*

Here, V is a local isometry that can be written as $V = \otimes_i V_i$, where each V_i is an isometry acting on at most 1 qubit, and P and Q are locally orthogonal projectors such that $P + Q = I$, and $\bar{M}$ is the complex conjugate of M. Moreover, we say that the simulation is efficient if m and $\|H'\|$ are at most $\mathcal{O}\left(poly(n, \eta^{-1}, \epsilon^{-1}, \Delta)\right)$, and the description of H' can be computed in $poly(n)$ time given the description of H.

For guided local Hamiltonian problems one is not just interested in the *energy values*, but also in what happens to the actual *eigenstates* throughout such transformations. In [GHLGM22], Appendix B, the authors check for a large range of Hamiltonian transformations to what extent the initial eigenstates are affected. In order to obtain our results, we need the following lemma which summarizes a whole chain of reductions in [GHLGM22].

Lemma 2.3 ('Classical evaluability'-preserving eigenstate encodings)**.** *Suppose H is an arbitrary k-local Hamiltonian on n qubits with a non-degenerate ground state $|g\rangle$ separated from excited states by a gap γ. Then H can be efficiently (Δ, η, ϵ) simulated by a 2-local Hamiltonian H' on $m = poly(n)$ qubits which has a non-degenerate ground state $|g'\rangle$, such that*

$$\left\| \mathcal{E}_{state}(|g\rangle) - |g'\rangle \right\| \leq \eta + \mathcal{O}(\gamma^{-1}\epsilon),$$

where $\mathcal{E}_{state}(\cdot)$ appends only states of a semi-classical form as a tensor product to $|g\rangle$, i.e. preserves the classical evaluability as in Definition 3.2.

[15] This is different from the convention, where usually $\mathcal{O}(f(n))$ is used instead.

Proof. This follows immediately from the proofs of Proposition 2 and Proposition 3 in [GHLGM22], while making the observation that all encodings up to the Spatially sparse 2-local Hamiltonian (with Pauli interactions with no Y-terms) only append states that satisfy the definition of poly-sized subset states (see the proof of Theorem 4.1 in the main text) to the original eigenstate of H. □

3 Guidable local Hamiltonian problems

3.1 Classically evaluatable states

Let us first introduce Gharibian and Le Gall's definition of query and sampling access to quantum states [GLG22], which slightly generalizes the original definition as first proposed by Tang used to dequantize quantum algorithms for recommendation systems [Tan19].

Definition 3.1 (Query and sampling access, from [GLG22])**.** *We say that we have* query and ξ-sampling access *to a vector $u \in \mathbb{C}^N$ if the following two conditions are satisfied:*

(i) we have access to an $\mathcal{O}\left(poly(\log(N)\right)$-time classical algorithm $\mathcal{Q}_u$ that on input $i \in [N]$ outputs the entry u_i.

(ii) we have access to an $\mathcal{O}\left(poly(\log(N)\right)$-time classical algorithm $\mathcal{SQ}_u$ that samples from a probability distribution $p : [N] \to [0, 1]$ such that

$$p(j) \in \left[(1-\xi)\frac{|u_j|^2}{\|u\|^2}, (1+\xi)\frac{|u_j|^2}{\|u\|^2}\right]$$

for all $j \in [N]$.

(iii) we are given a real number m satisfying $|m - \|u\|| \leq \xi\|u\|$.

We simply say that we have sampling access to u (without specifying ξ) if we have 0-sampling access.

In this work we propose a new class of quantum states, conceptually different from those of Definition 3.1, which we will call *classically evaluatable quantum states*. Our main motivations for doing so are the following:

1. It seems rather difficult to find Ansätze that are used in practice for ground state energy estimation that satisfy all conditions of Definition 3.1. As one of the main motivations of this work is to investigate the power of quantum versus classical state preparation when one has access to Quantum Phase Estimation, we wanted to define a class of states that can both be prepared efficiently on a quantum computer and which contains a large class of Ansätze commonly used in practice.

2. Analogous to Dinur's construction, one would expect that determining if a local Hamiltonian has ground state energy exactly zero or some constant away from zero is QMA-hard if the quantum PCP conjecture is true. However, there are arguments from physics[16] on why one might expect this problem to be in NP [PH11]. To study the question of containment in NP it is necessary to be able to work with states within a deterministic setting, and therefore it does not make sense to rely on a form of sampling access which inherently relies on a probabilistic model of computation.

[16] In this setting the LH problem becomes equivalent to determining whether the free energy of the system becomes negative at a finite temperature. One expects then that at such temperatures, the system loses its quantum characteristics on the large scale, making the effects of long-range entanglement become negligible. Hence, this means that the ground state of such a system should have some classical description, which places the problem in NP [Ara11].

3. To add to the previous point, being able to study containment in NP comes with the additional advantage of being able to make statements about whether the problem admits a PCP by the classical PCP theorem. No such theorem is currently known for MA, and we exploit this further in Section 6 where we introduce a new type of 'quantum' PCP.

We will define these quantum states in a slightly more general setting for completeness – by allowing for probabilistic computation of expectation values as well – but this will not be important for the remainder of this work.

Definition 3.2 (ϵ-classically evaluatable and quantumly preparable states)**.** *Let O be some $\mathcal{O}(\log n)$-local observable satisfying $\|O\| \leq 1$, for which we have an efficient classical description (in the sense that we can query all its matrix elements efficiently). We say a state $u \in \mathbb{C}^{2^n}$ is* ***ϵ-classically evaluatable*** *if*

(i) there exists a classical description of u, denoted as $desc(u)$, which requires at most $poly(n)$ bits to write down, and

(ii) there exists a classical probabilistic algorithm $\mathcal{Q}_O$ which, given $desc(u)$, computes an estimate $\hat{z}$ such that $|\hat{z} - \langle u| O |u\rangle| \leq \epsilon$ in time $\mathcal{O}(poly(n, 1/\epsilon))$, with success probability $\geq 2/3$.

Furthermore, we say a state $u \in \mathbb{C}^{2^n}$ is also ***quantumly preparable*** *if*

(iii) there exists a quantum circuit V of at most $poly(n)$ 1- and 2-qubit gates that prepares u as a quantum state, i.e. $|u\rangle$. The description of V can be computed efficiently using some efficient classical algorithm $\mathcal{A}_V$, which only takes $desc(u)$ as an input.

Finally, if $\epsilon = 0$ and the algorithm used in (ii) is deterministic instead of probabilistic, we simply say that u is ***classically evaluatable****.*

Note that it is not required that u is normalized, however by requirement (ii) it is possible to calculate the norm of u. Normalization is of course required for u to be quantumly preparable. Also note that if condition (iii) holds, condition (ii) (for $\epsilon > 0$) is no longer necessary in order to work with the class of states as a suitable Ansatz provided that one has access to a quantum computer, since there exist quantum algorithms to estimate the expectation values of the observables up to arbitrarily precise inverse polynomial precision. However, the current definition allows one to adopt the two-step classical-quantum procedure of *classical* Ansatz generation and *quantum* ground state preparation, as described in Section 1.

To demonstrate the practical relevance of Definition 3.2, we give four examples of Ansätze which all satisfy the required conditions to be (ϵ-)classically evaluatable and quantumly preparable. The first two examples will also be perfectly samplable, as in Definition 3.1, of which the proofs are given in Appendix A.

Example 3.1 (Matrix-product states with bounded bond and physical dimensions)**.** Matrix-product states are quantum states of the following form

$$|u\rangle = \sum_{\{s\}} \mathrm{Tr}[A_1^{(s_1)} A_2^{(s_2)} \dots A_n^{(s_n)}] \, |s_1, \dots, s_{n-1}\rangle \,,$$

where s_i are qudits of 'physical' dimension p, the $A_i^{(s_i)}$ are complex, square matrices of bond dimension D (except at the 'edges' $A_1^{(s_i)}$ and $A_i^{(s_{n-1})}$, where they are vectors of dimension D), and n denotes the total number of qudits. We say that the bond dimension is bounded if it is at most polynomial in n, and that the physical dimension is bounded if it is taken to be some constant independent of n. MPS are also 0-samplable, which is shown in Appendix A.

Conditions check:

(i) The MPS is fully determined by the set of matrices $\{A_i^{s_i}\}$, and can be described explicitly using at most $npD^2 = \mathcal{O}(poly(n))$ complex numbers.

(ii) $\langle u| O |u\rangle$ can be computed exactly in time $\mathcal{O}(npD^3)$, which is efficient as long as none of the parameters grows exponentially in n. Since O is $\mathcal{O}(\log(n))$-local, it can be represented by a matrix product operator acting on at most $\mathcal{O}(\log(n))$ qubits and with bond dimension therefore at most $p^{\mathcal{O}(\log(n))} = \mathcal{O}(poly(n))$ for constant p [Orú14, Sch11].

(iii) An MPS on n qubits with bond dimensions D can be prepared on a quantum computer up to distance ϵ using at most $\mathcal{O}(nD\log(D)^2\log(n/\epsilon))$ 1- and 2-qubit gates and requiring $\lceil\log(D)\rceil$ additional ancilla qubits. A method for constructing such a circuit can be found in Appendix B and is based on [SSV+05].

Example 3.2 (Stabilizer states)**.** Gottesmann and Knill [Got98] showed that there exists a class of quantum states, containing states that exhibit large entanglement, that can be efficiently simulated on a classical computer. These states are called *stabilizer states* and are those generated by circuits consisting of Clifford gates, $\mathcal{C} = \langle \text{CNOT}, H, S\rangle$ where $S = \sqrt{Z}$ is a phase gate, starting on a computational basis state. Any measurement of local Pauli's on these states can be efficiently classically simulated. Amongst other things, stabilizer states have been used to formulate error correcting codes [Ste03], study entanglement [BDSW96], and in evaluating quantum hardware through randomised benchmarking [KLR+08]. Stabilizer states are also 0-samplable, again shown in Appendix A.

Conditions check:

(i) Any stabilizer state can be described by a linear depth circuit consisting of Clifford gates starting on the $|0^n\rangle$ state [MR18]. A possible description of such a circuit is a list of tuples (q_1, q_2, t, g), where q_1 (resp. q_2) denotes the first (resp. second) qubit that $g \in \mathcal{C}$ acts on at depth t. This description takes at most $\tilde{\mathcal{O}}(n^2)$ bits to write down.

(ii) The Gnottesman-Knill [Got98] theorem shows that stabilizer states allow for strong classical simulation and efficient classical computation of probabilities for Pauli measurements. This in particular allows for the calculation of expectation values of log-local observable.

(iii) The description is given as a quantum circuit, which can be implemented to prepare the quantum state.

We will now give two examples of Ansätze that have been shown to not be ξ-samplable, even up to some large constant values of ξ.

Example 3.3 (Constant depth quantum circuits)**.** Constant depth quantum circuits are circuits that, given some fixed gate set $\mathcal{G}$ with just local operations, are only allowed to apply at most $t = \mathcal{O}(1)$ consecutive layers of operations from $\mathcal{G}$ on some initial quantum state, which we take to be the all-zero state $|0\ldots0\rangle$. An example of constant depth quantum circuits that are used as classical Ansätze would be the simple case of the *product state Ansatz*, where one only considers one-qubit gates applied per site. Product state Ansätze are widely used in classical approximation algorithms to Local Hamiltonian problems, see for example [BH13, GP19]. In [TD04] it was shown that the ability to perform approximate weak sampling from the output of a constant depth quantum circuit up to relative error $0 < \xi < 1/3$ implies that $\mathsf{BQP} \subseteq \mathsf{AM}$, which means that it is unlikely that constant depth quantum circuits are ξ-samplable for any $\xi < 1/3$.

Conditions check:

(i) Let us for simplicity assume that we use a fixed gate set $\mathcal{G}$ of only two-qubit gates. A possible description could be a concatenated string of tuples, where each tuple looks like (q_1, q_2, t, g), where q_1 (resp. q_2) denotes the first (resp. second) qubit the $g \in \mathcal{G}$ acts on at depth t. This description takes at most $poly(n)$ bits to write down.

(ii) $\langle u| O |u\rangle = \langle 0| U^\dagger |O| U |0\rangle$, where $U^\dagger O U$ is a $\log(n)2^t$-local observable (via a light-cone argument), and hence we can compute $\langle 0| U^\dagger |O| U |0\rangle$ in time $\mathcal{O}\left(2^{\mathcal{O}(\log(n)2^t)} \cdot poly(n)\right)$ which is $\mathcal{O}(poly(n))$ if $t = \mathcal{O}(1)$.

(iii) This holds by definition.

By combining Example 3.2 and Example 3.3 we find that any state of the form $UC|0^n\rangle$, with U a constant-depth circuit and C a Clifford circuit, is also classically evaluatable and quantumly preparable. Our final example is of a class of states that are not perfectly classically evaluatable, but are ϵ-classically evaluatable for any $\epsilon = \Omega(1/poly(n))$.

Example 3.4 (Instantaneous quantum polynomial (IQP) circuits)**.** IQP circuits start in $|0^n\rangle$ and apply a polynomial number of local gates that are diagonal in the X-basis, followed by a computational basis measurement [BJS11]. An equivalent definition would be to consider circuits with gates that are diagonal in the Z-basis, but then sandwiched in two layers of Hadamard gates (again followed by a measurement in the computational basis). It is well known that IQP circuits are difficult to sample from: if IQP circuits could be weakly simulated to within multiplicative error $1 \leq c < \sqrt{2}$, then the polynomial hierarchy would collapse to its third level [BJS11]. Hence, they are not ξ-samplable for any $\xi < \sqrt{2} - 1$. However, we will now show that states generated by IQP circuits are ϵ-classically evaluatable for all $\epsilon = \Omega(1/poly(n))$.

Conditions check:

(i) This follows by definition, since all gates are local and there are only a polynomial number of them.

(ii) This is a corollary from Theorem 3 in [BJS11], where it is shown that one can exactly sample basis states on $\mathcal{O}(\log n)$ qubits according to their l_2-norm. Let C be an IQP-circuit of n qubits which produces the state $|u\rangle = C|0^n\rangle$, and let $S \subseteq [n]$ with $|S| = \mathcal{O}(\log n)$ be the qubits on which a $\mathcal{O}(\log n)$-local observable O acts. Following the proof of Theorem 3 in [BJS11], the state right before the last layer of Hadamard is given by

$$|\phi\rangle = \frac{1}{N} \sum_{x \in S, y \in [n]\setminus S} e^{if(x,y)} |x, y\rangle ,$$

where the pair $x, y \in \{0,1\}^n$ denotes the bit string state corresponding to the concatenation (with the correct indexing) of the bit strings $x \in \{0,1\}^{|S|}$ and $y \in \{0,1\}^{n-|S|}$. Here $f(x,y)$ is a phase function which can be computed efficiently, by accumulating the relevant diagonal entries of the successive commuting gates. Since O does not act on the qubits with indices $[n] \setminus S$, and they only get acted upon by Hadamards, further measurements on this register should not influence any POVM that only acts on S by the no-signaling principle. By this observation, the protocol is now very simple: one samples a random bit string $y' \in \{0,1\}^{n-|S|}$ and computes the random variable

$$X_i = \frac{1}{2^{|S|}} \sum_{x, x' \in \{0,1\}^{|S|}} \langle x| e^{-if(x,y')} H^{\otimes |S|} O H^{\otimes |S|} e^{if(x,y')} |x'\rangle ,$$

which can be done exactly since $|S| = \mathcal{O}(\log n)$ and $f(x,y')$ can be computed efficiently. Since $\|O\| \leq 1$, we have that $\mathbb{E}[X_i^2] \leq 1$ and $|\mathbb{E}[X_i]| \leq 1$, and therefore $\mathrm{Var}[X_i] =$

$\mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 \leq 2$. Therefore, taking $s = c/\epsilon^2$ samples of X_i (which are independent random variables) and computing $\hat{z} = \frac{1}{s}\sum_{i\in[s]} X_i$ ensures that

$$|\hat{z} - \langle u| O |u\rangle| \leq \epsilon,$$

with probability $\geq 2/3$, provided that $c \geq 6$. This follows from a simple application of Chebyshev's inequality.

(iii) This follows also by definition.

In general quantum states will *not* be classically evaluatable (as that would imply $\mathsf{QMA} = \mathsf{NP}$ as they could be used as witnesses for the QMA-hard local Hamiltonian problem), and some other notable examples of classes of states which are not expected to be classically evaluatable are Projected Entangled Pair States (PEPS) (since computing expectation values of local observables is #P-hard [SWVC07]) and collections of local reduced density matrices (to check whether they are consistent with a global quantum state is QMA-hard [Liu07, BG22]).

We have seen that constant-depth quantum circuits are not even approximately samplable (under the conjecture that $\mathsf{BQP} \not\subset \mathsf{AM}$ [TD04]). We can formalize this in the following proposition which relates ξ-samplable states to ξ-classically evaluatable states.

Theorem 3.1. *For any $\xi > 0$, any ξ-samplable state is also $\mathcal{O}(\xi)$-classically evaluatable. On the other hand, there exist states that are perfectly classically evaluatable but not ξ'-samplable for all $0 < \xi' < 1/3$, unless $\mathsf{BQP} \subseteq \mathsf{AM}$.*

Proof. Let $u \in \mathbb{C}^N$ be a ξ-samplable state with $N = 2^n$. The first part of the proposition follows by checking the two conditions.

(i) u is described by giving the algorithms $\mathcal{Q}_u$ and $\mathcal{SQ}_u$. Both these algorithms run in $\mathcal{O}(poly(log(N)))$-time, which implies that both have an efficient description of length at most $\mathcal{O}(poly(log(N)))$ (in terms of local classical operations, i.e. logic gates).

(ii) Estimates of log-local observables can be calculated by using Theorem 3 in [GLG22]. Let the polynomial $P(x) = x$ (which has degree 1), the sparse matrix A be the $N \times N$ matrix given by O (the log-local observable) tensored with identities acting on the locations on which O does not act (which means that A is at most $s = 2^{\mathcal{O}(\log n)} = poly(n)$-sparse), and $v = u$. This gives an estimate $\hat{z}$ such that

$$|\hat{z} - \langle u| O |u\rangle| \leq \epsilon$$

in time $\mathcal{O}^*(s^2/\epsilon) = poly(n, 1/\epsilon)$ for any $\epsilon \geq 8\xi$, $\epsilon > 0$.

This shows that any ξ-samplable state is at least 8ξ-classically evaluatable. The second part follows directly from [TD04], Theorem 3, which shows that the ability to perform approximate weak sampling from the output of a constant depth quantum circuit up to relative error $0 < \xi < 1/3$ implies that $\mathsf{BQP} \subseteq \mathsf{AM}$, obstructing the ability to satisfy condition (ii) in Definition 3.1. By Example 3.3, we already showed that constant-depth quantum circuits produce classically evaluatable states, completing the proof. □

This gives rise to a (conjectured) hierarchical structure of states as depicted in Figure 2. An interesting observation is a supposedly significant leap in the hierarchy when we allow for a small error ϵ in the definition of ϵ-classically evaluatable states. A straightforward way to explain this is by considering how it affects our ability to determine a global property of a quantum state, like its energy with respect to a Hamiltonian H.

Let H be a sum of m log-local terms, i.e. $H = \sum_{i=0}^{m-1} H_i$, satisfying $\|H\| \leq 1$. If one wants to evaluate the energy of an ϵ-classically evaluatable state with respect to H up to accuracy

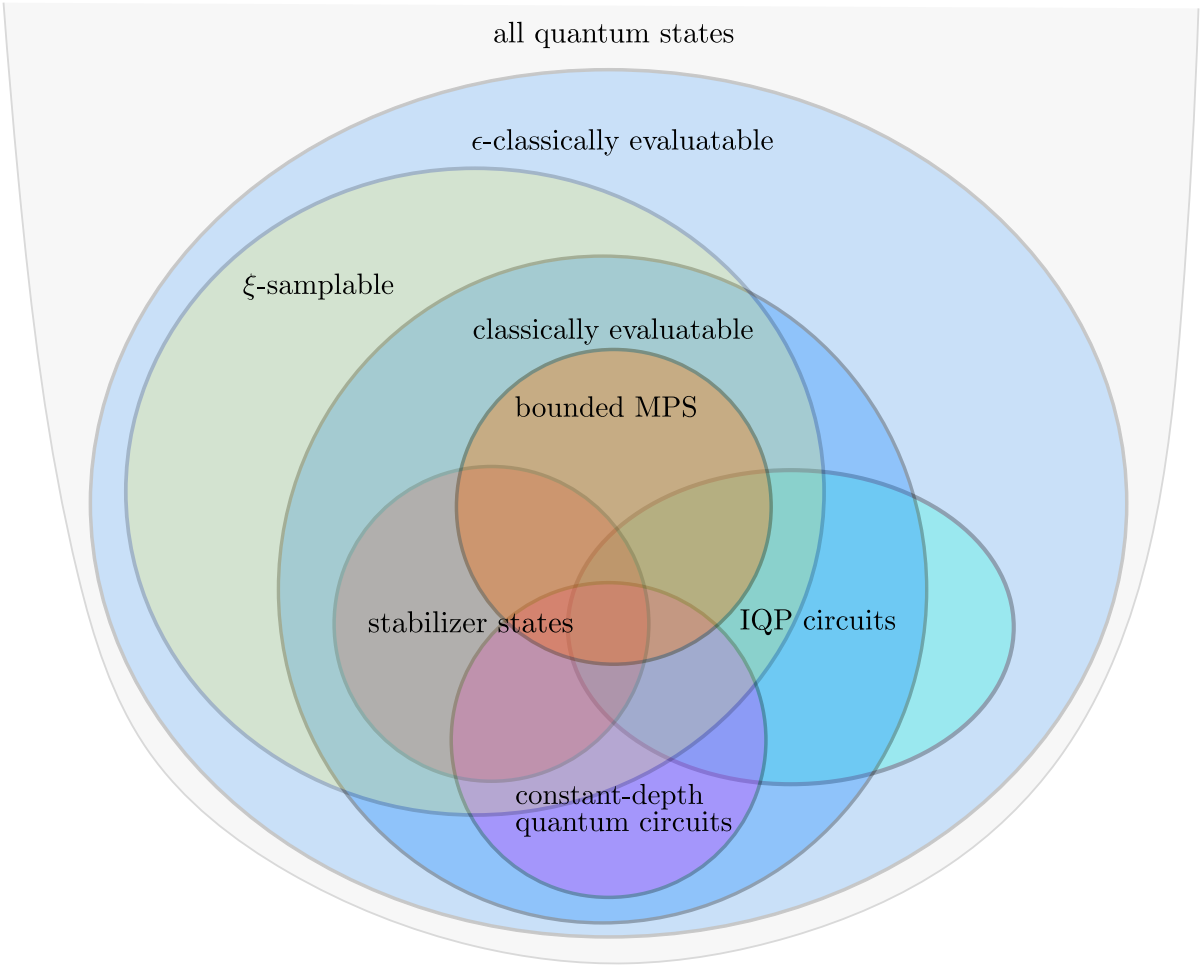

Figure 2: Visualization of the (conjectured) relations between classes of quantum states considered in this work, given a Hilbert space of a fixed dimension. For MPS, we only consider states with polynomially-bounded bond and local dimension. We have that $\xi \leq \epsilon/8 \leq 1/3$, such that by Theorem 3.1 we have that (i) all ξ-samplable states are also ϵ-classically evaluatable and (ii) constant-depth and IQP circuits are not ξ-samplable. One also expects that there are quantum states (which can be prepared by a polynomial time quantum circuit) which are neither classically evaluatable nor samplable, or else QMA (QCMA) would be in NP or MA, respectively.

ϵ', then ϵ has to be less than ϵ'/m since in the worst case the error grows linearly with the number of terms. Instead, ξ-samplable states have a requirement on the accuracy of sampling, which is a property of the global state. [GLG22] shows that this property can be used for energy estimation, where the requirement on ξ only depends on the precision with which one wants to measure the energy. We see this reflected in Theorem 3.1, which shows that if a state has the property of being ξ-samplable this implies that the state is $\mathcal{O}(\xi)$-classically evaluatable, but not the other way around. However, we are not aware of any classes of states which are provably only ξ-samplable for a constant, but small, $\xi > 0$ (all examples that we give in this work are in fact 0-samplable).

For the remainder of our work, we will focus on (0-)classically evaluatable states, which by Definition 3.2 means that $\mathcal{Q}_O$ is deterministic. A notable advantage of this approach, as opposed to 0-samplable states, lies in its compatibility with deterministic algorithms, allowing us to give NP containment results (see Section 5). This is also a prerequisite to make connections to MA as well, see Appendix C.

3.2 Variants of guidable local Hamiltonian problems

Let us define the following class of local Hamiltonian problems, which can be viewed as 'Merlinized' versions of the original guided local Hamiltonian problem. We make a distinction between different types of promises one can make with respect to the existence of guiding states: we either assume that the guiding states are of the form of Definition 3.2 (with or without the promise that the states are also quantumly preparable), or that there exists an efficient quantum circuit that prepares the guiding state.

Definition 3.3 (Guidable Local Hamiltonian Problems)**.** *Guidable Local Hamiltonian Problems are a class of problems defined by having the following input, promise, one of the extra promises and output:*
Input: *A k-local Hamiltonian H with $\|H\| \leq 1$ acting on n qubits, threshold parameters $a, b \in \mathbb{R}$ such that $b - a \geq \delta > 0$ and a fidelity parameter $\zeta \in (0, 1]$.*
Promise: *We have that either $\lambda_0(H) \leq a$ or $\lambda_0(H) \geq b$ holds, where $\lambda_0(H)$ denotes the ground state energy of H.*
Extra promises: *Denote Π_{gs} for the projection on the subspace spanned by the ground state of H. Then for each problem class, we have that either one of the following promises hold:*

1. *There exists a classically evaluatable state $u \in \mathbb{C}^{2^n}$ for which $\|\Pi_{gs} u\|^2 \geq \zeta$. Then the problem is called the **Classically Guidable Local Hamiltonian Problem**, shortened as $\mathsf{CGaLH}(k, \delta, \zeta)$. If u is also quantumly preparable, we call the problem the **Classically Guidable and Quantumly Preparable Local Hamiltonian Problem**, shortened as $\mathsf{CGaLH}^*(k, \delta, \zeta)$.*

2. *There exists a unitary V implemented by a quantum circuit composed of at most $T = poly(n)$ gates from a fixed gate set $\mathcal{G}$ that produces the state $|\phi\rangle = V|0\rangle$ (with high probability), which has $\|\Pi_{gs}|\phi\rangle\|^2 \geq \zeta$. Then the problem is called the **Quantumly Guidable Local Hamiltonian problem**, shortened as $\mathsf{QGaLH}(k, \delta, \zeta)$.*

Output:

- *If $\lambda_0(H) \leq a$, output* YES.
- *If $\lambda_0(H) \geq b$, output* NO.

One can also consider other types of guiding states, for example the samplable states as in Definition 3.1. This guidable local Hamiltonian problem variant was already introduced in Section 5 of [GLG22].

The $\mathsf{QGaLH}(k, \delta, \zeta)$ problem is very similar to the low complexity low energy states problem from [WJB03], but differs in some key ways. In the low complexity low energy states problem one is promised that for all states $\{|\phi\rangle\}$ that can be prepared from $|0 \ldots 0\rangle$ with a polynomially bounded number of gates from a fixed gate set, one has that either there exists at least one such $|\phi\rangle$ such that $\langle\phi| H |\phi\rangle \leq a$ or for all these $|\phi\rangle$ we have $\langle\phi| H |\phi\rangle \geq b$. Instead, in $\mathsf{QGaLH}(k, \delta, \zeta)$ one is promised that there exists a state $|\psi\rangle$ which can be prepared efficiently on a quantum computer that has fidelity ζ with the ground space of H. This promise in the fidelity does not imply that the energy of this $|\phi\rangle$ is necessarily low, as it might have a large fidelity with states in the high-energy spectrum of H. Nevertheless, it does imply that in the YES-case there exists a low complexity low energy state $|\phi\rangle$. One can make use of the state $|\psi\rangle$ that has significant overlap with the ground state and use phase estimation to project $|\psi\rangle$ onto a state $|\phi\rangle$ with energy at least inverse polynomially close to the ground state (which implies $|\phi\rangle$ can be prepared by a quantum circuit). However, in the NO-case this promise on the fidelity implies that every possible state $|\psi\rangle$ has energy $\langle\psi| H |\psi\rangle \geq b - \mathcal{O}(1/\exp(n))$, as even in the NO-case it is still possible to approximate the ground state energy up to polynomial precision. This is different

from the NO-case of the low complexity low energy states problem, where there might exist states with energy lower than a, as long as these states are not preparable by a polynomial-time quantum circuit, making the $\mathsf{QGaLH}(k, \delta, \zeta)$ problem more restrictive than the low complexity low energy states problem. In principle, this could be remedied by relaxing the requirement in $\mathsf{QGaLH}(k, \delta, \zeta)$ from having fidelity with the ground space to having fidelity with the space of states with sufficiently low energy in the YES-case only. All our results that follow would still hold, and this new problem could then be seen as a generalisation of the low complexity low energy states problem.

In the upcoming section we will characterize the complexity of these guidable local Hamiltonian problems in various parameter regimes.

4 QCMA-completeness of guidable local Hamiltonian problems

Before we prove one of the main theorems of this work, we first prove the following claim, which can be done by making some simple observations about the original proof constructions in [ABOBS22] and [JKNN12]. Although not strictly needed to get a result, the use of the claim does allow for an improvement in the parameter range for which the later theorem holds, at the cost of resorting to randomized reductions.

Claim 4.1. $\mathsf{UQCMA}[c, s] \subseteq \mathsf{UQCMA}_1[s']$, *where* $s' \leq 1 - \Omega(|y|^{-2})$, *where* $|y|$ *denotes the witness length in the original* $\mathsf{UQCMA}[c, s]$ *protocol.*

Proof. Let $\langle U, p_1, p_2 \rangle$ be QCMA a promise problem, for which we assume that it only uses gates from the gate set $\mathcal{G} = \{H, X, \text{Toffoli}\}$, which is justified by the robustness of the class under the choice of a universal gate set. The key observation is that the randomized reduction described in [ABOBS22], which maps $\langle U, p_1, p_2 \rangle$ to a UQCMA instance $\langle \tilde{U}, \tilde{p_1}, \tilde{p_2} \rangle$, only does so by modifying the completeness and soundness parameters (uniformly random from a finite set) and by adding a 'filter', again sampled uniformly at random from a pairwise independent hash function family. This reduction succeeds with probability $\Omega(1/|y|)$. The filter appends a solely classical – but made reversible in order to allow for coherent unitary implementation – circuit in front of the original QCMA circuit. Hence, this part of the new circuit can be implemented solely by Toffoli gates, which are in our gate set $\mathcal{G}$: therefore meeting the requirements of Lemma 2.2. In the construction of [ABOBS22], the completeness and soundness parameters get mapped to $((k+1)/|y|$ and $k/|y|$, respectively, where k is randomly sampled from the set $\{1, \ldots, |y| - 2\}$. By Lemma 2.2, we have that s' is then upper bounded by

$$\begin{aligned} s' &= \max_{k \in \{1,\ldots,|y|-2\}} \frac{1}{2}\left(1 - \left(\frac{k+1}{|y|} - \frac{k}{|y|}\right)\right)\left(1 + \left(1 + \frac{k+1}{|y|} - \frac{k}{|y|}\right)^2\right) \\ &= 1 - \Omega(|y|^{-2}) \end{aligned}$$

□

In QCMA the promises of completeness and soundness are always for computational basis state witnesses. Hence, these might no longer hold when *any* quantum state can be considered as witness: for example, in the NO-case there might be highly entangled states which are accepted with probability $\geq 2/3$. When considering a circuit problem, the verifier (Arthur) can easily work around this by simply measuring the witness and then proceeding his verification with the resulting computational basis state. However, there is also another trick, which retains the unitarity of the verification circuit – and which we will denote as the 'CNOT-trick' from now on – to force the witness to be classical, first used in proving QCMA-completeness of the *Low complexity low energy states* problem in [WJB03]. Since the authors do not explain the precise

mechanism behind the workings of this CNOT-trick, we provide a short proof of the lemma below.

Lemma 4.1 (The 'CNOT-trick')**.** *Let $p(n) : \mathbb{N} \to \mathbb{R}_{>0}, q(n) : \mathbb{N} \to \mathbb{R}_{>0}$ be polynomials. Let V be a quantum polynomial-time verifier circuit that acts on an n-qubit input register A, a $p(n)$-qubit witness register B and a $q(n)$-qubit workspace register C, initialized to $|0\rangle^{\otimes q(n)}$. Denote Π_0 for the projection on the first qubit being zero. Let Q be the Marriott-Watrous operator of the circuit, defined as*

$$Q = \left(\langle x| \otimes I_w \otimes \langle 0|^{\otimes q(n)}\right) V^\dagger \Pi_0 V \left(|x\rangle \otimes I_w \otimes |0\rangle^{\otimes q(n)}\right).$$

Consider yet another additional $p(n)$-qubit workspace D initialized to $|0\rangle^{\otimes p(n)}$, on which V does not act. Then by prepending V with $p(n)$ CNOT-operations, each of which is controlled by a single qubit in register B and targeting the corresponding qubit in register D, the corresponding Marriott-Watrous operator becomes diagonal in the computational basis.

Proof. Denote U_{CNOT} for the $2p(n)$ qubit operation that acts on the two registers B and D, and that for each $l \in [p(n)]$ applies a CNOT controlled by qubit l in register B and targets qubit l in register D. Consider the new verifier circuit $\tilde{V} = VU_{\text{CNOT}}$ that acts on the registers A, B, C and D, with the corresponding Mariott-Watrous operator $\tilde{Q}$. Let $|i\rangle$ and $|j\rangle$ for $i, j \in [2^{p(n)}]$ be arbitrary computational basis states. Then we have

$$\begin{aligned}
\langle i|\,\tilde{Q}\,|j\rangle &= \left(\langle x| \otimes \langle i| \otimes \langle 0|^{\otimes q(n)} \otimes \langle 0|^{\otimes p(n)}\right) U_{\text{CNOT}} V^\dagger \Pi_0 V U_{\text{CNOT}} \left(|x\rangle \otimes |j\rangle \otimes |0\rangle^{\otimes q(n)} \otimes |0\rangle^{\otimes p(n)}\right) \\
&= \left(\langle x| \otimes \langle i| \otimes \langle 0|^{\otimes q(n)} \otimes \langle i|\right) V^\dagger \Pi_0 V \left(|x\rangle \otimes |j\rangle \otimes |0\rangle^{\otimes q(n)} \otimes |j\rangle\right) \\
&= \langle i|j\rangle \left(\langle x| \otimes \langle i| \otimes \langle 0|^{\otimes q(n)}\right) V^\dagger \Pi_0 V \left(|x\rangle \otimes |j\rangle \otimes |0\rangle^{\otimes q(n)}\right) \\
&= \delta_{i,j} \langle i|\,Q\,|j\rangle,
\end{aligned}$$

where we used the fact that V and Π_0 themselves do not act on register D. Hence, the operator $\tilde{Q}$ is diagonal in the computational basis, where its entries are taken from the diagonal of Q. □

With this claim and lemma in our toolbox, we proceed to prove the following result, which is one of the main technical contributions of this work.

Theorem 4.1. $\mathsf{CGaLH}(k, \delta, \zeta)$ *is* QCMA*-hard under randomized reductions for $k \geq 2$, $\zeta \in (1/\,poly(n), 1 - 1/\,poly(n))$ and $\delta = 1/\,poly(n)$.*

Proof. Let us first state a 'basic version' reduction, for which we prove completeness and soundness, and finally improve its parameters in terms of the achievable fidelity and locality domains.

The reduction Let $\langle U, p_1, p_2\rangle$ be a QCMA promise problem that only uses gates from $\mathcal{G} = \{H, X, \text{Toffoli}\}$. By using Claim 4 and Lemma 2.2, we let $\langle U, p_1, p_2\rangle$ go through the following chain of reductions using the following inclusions of complexity classes[17]

$$\mathsf{QCMA} \subseteq_r \mathsf{UQCMA} \subseteq \mathsf{UQCMA}_1,$$

where the final UQCMA_1 promise problem is defined as $\langle \tilde{U}, 1, s'\rangle$, where $\tilde{U}$ takes as input x and a witness $\tilde{y} = (y, y')$ (where y' is an additional witness that denotes the probability of acceptance, provided the prover is honest) of size $w := |\tilde{y}| = \mathcal{O}(poly(n))$ (next to the original witness y of the original QCMA promise problem), and consists of at most $T = poly(n)$ gates, and has completeness and soundness parameters 1 and $s' = 1 - \Omega(w^{-2})$, respectively. We will now apply the following modifications to $\langle \tilde{U}, 1, s'\rangle$:

[17]For the reader that does not like randomized reductions: one can also only consider the QCMA to QCMA_1 reduction of this chain. In that case all parts – except for the locality reduction – of the proof below will still hold, and we instead obtain the result for $k \geq 6$. An interesting open question would be if the result for $k \geq 2$ can also be obtained without a randomised reduction.

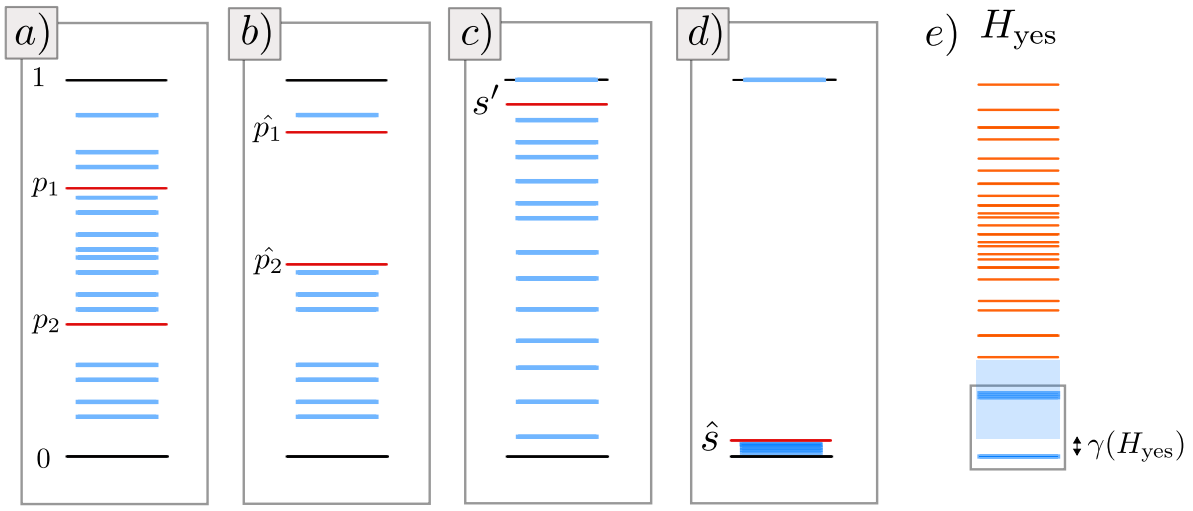

Figure 3: Illustration of the key ideas to construct the desired witness distribution in the YES-case in the first part of the reduction. The blue lines are witnesses, for which their position with respect to the y-axis represents the corresponding acceptance probabilities. The dark red lines represent the completeness and soundness parameters. $a) \to b)$ represents the randomized reduction from a QCMA-problem to a UQCMA one, $b) \to c)$ the reduction from UQCMA one to a UQCMA_1-problem, $c) \to d)$ the "unanimous vote" error reduction and finally $d) \to e)$ the circuit-to-Hamiltonian mapping with the small penalty resulting in the Hamiltonian H_{yes}, whose ground state corresponds exactly the unique witness with acceptance probability 1. The light blue shaded area represents the fact that we do not know the exact energy values corresponding to non-accepting witnesses, except for the fact that they are separated from $\lambda_0(H_{\text{yes}})$ by at least $\gamma(H_{\text{yes}}) = \Omega(1/M^5)$ for our choice of ϵ. Observe that if one was able to show that QMA $\subseteq_r$ UQMA, one could use the same proof construction to show QMA-hardness of inverse-poly-gapped Hamiltonians (in this setting we do not care about the fidelity, so perfect completeness is not needed), for which we only yet know that they are QCMA-hard. This was already shown in [ABOBS22], and rediscovered in this work.

1. First, we force the witness to be classical by adding another register to which we 'copy' all bits of $\tilde{y}$ (through CNOT operations), before running the actual verification protocol – i.e. we use the CNOT trick of Lemma 4.1, which diagonalizes the corresponding Marriot-Watrous operator in the computational basis, and in particular means that the acceptance probability of the verifier can be maximized by using a classical witness.

2. We apply a slightly modified form of *error reduction* to the circuit. Since $s' = 1 - \Omega(w^{-2}) > \frac{1}{2}$, we have to resort to a "unanimous vote" instead of a "majority vote" in our construction to exponentially suppress all witnesses in the NO-case, and all but one witness[18] (the one that achieves perfect completeness) in the YES-case close to zero. This is done by applying the so-called "Marriot and Watrous trick" for error reduction, described in [MW04], which allows one to repeat the verification circuit several times whilst re-using the same witness. By only accepting when *all* repetitions accept, one can quickly verify that the probability of acceptance for the witness that was originally accepted with probability 1 remains 1, and that for all other witnesses the probability of acceptance becomes suppressed exponentially close to zero. More precisely, by repeating k times we have that the probability of acceptance in the NO-case is at most $(1 - \Omega(1/w^2))^k \leq e^{-\Omega(k/w^2)} = e^{-p(n)}$ for p a polynomial such that $k > p(n)\mathcal{O}(w^2)$, implying that we need to repeat the verification circuit only polynomially many times. Note that after these two changes the protocol is still in UQCMA_1, albeit with a new soundness parameter which is now exponentially close to zero.

Let the resulting protocol be denoted by $\langle \hat{U}, 1, \hat{s} \rangle$, where $\hat{U}$ has an input register $\hat{A}$, a witness register $\hat{W}$ and ancilla register $\hat{B}$, uses T gates and where $\hat{s} = 2^{-\mathcal{O}(k)}$, which is exponentially close to zero. We denote $\tilde{y}^*$ for the (unique) witness with acceptance probability 1 in the YES-case. We will also write $P(y) := \Pr[\hat{U} \text{ accepts } (y)]$.

Consider Kitaev's original 5-local clock Hamiltonian [KSV02] with a small penalty on H_{out}, following [DGF22]:

$$H_{yes} = H_{\text{in}} + H_{\text{clock}} + H_{\text{prop}} + \epsilon H_{\text{out}}, \tag{1}$$

with

$$\begin{aligned}
H_{\text{in}} &:= (I - |x\rangle\langle x|)_{\hat{A}} \otimes I_{\hat{W}} \otimes (I - |0\dots0\rangle\langle 0\dots0|)_{\hat{B}} \otimes |0\rangle\langle 0|_C, \\
H_{\text{out}} &:= |0\rangle\langle 0|_{\text{out}} \otimes |T\rangle\langle T|_C, \\
H_{\text{clock}} &:= \sum_{j=1}^{T} |0\rangle\langle 0|_{C_j} \otimes |1\rangle\langle 1|_{C_{j+1}}, \\
H_{\text{prop}} &:= \sum_{t=1}^{T} H_t \qquad \text{where} \\
H_t &:= -\frac{1}{2}\hat{U}_t \otimes |t\rangle\langle t-1|_C - \frac{1}{2}\hat{U}_t^\dagger \otimes |t-1\rangle\langle t|_C + \frac{1}{2} I \otimes |t\rangle\langle t|_C + \frac{1}{2} I \otimes |t-1\rangle\langle t-1|_C,
\end{aligned} \tag{2}$$

where x denotes the input, C denotes the 'clock' register consisting of $T = poly(n)$ qubits. The class of history states parametrized by all possible witnesses $\tilde{y}$ is then given by

$$|\eta(\tilde{y})\rangle = \frac{1}{\sqrt{T+1}} \sum_{t=0}^{T} \hat{U}_t \dots \hat{U}_1 |x\rangle_{\hat{A}} |\tilde{y}\rangle_{\hat{W}} |0\dots0\rangle_{\hat{B}} |t\rangle_C . \tag{3}$$

[18] Note that this is not the case if we do not go through unique QCMA as per the footnote above. However, since we do use QCMA with perfect completeness, the history states corresponding to witnesses with acceptance probability one will all be in the ground space of the resulting Hamiltonian. Hence, all steps of the proof will still go through, except for the locality reduction which relies on the ground state being non-degenerate.

Let q be the total number of qubits that H_{yes} operates on. We will also consider another Hamiltonian, H_{no}, given by

$$H_{no} = \sum_{i=0}^{q-1} |1\rangle\langle 1|_i + bI, \tag{4}$$

where $b > 0$ is a tunable constant. Note that H_{no} has a *non-degenerate* ground state with energy b given by the all zeros state, and the spectrum after that increases in steps of 1 (and so it in particular has a *spectral gap* of 1). Note that $\|H_{no}\| = q + b = \mathcal{O}(poly(n))$. The full Hamiltonian we consider will now be

$$H = H_{yes} \otimes |0\rangle \langle 0|_D + H_{no} \otimes |1\rangle \langle 1|_D . \tag{5}$$

As a guiding state will use the following polynomially-sized subset state

$$|u_{\text{yes}}\rangle = \frac{1}{\sqrt{T+1}} \sum_{t=0}^{T} |x\rangle_{\hat{A}} |\tilde{y}\rangle_{\hat{W}} |0 \dots 0\rangle_{\hat{B}} |t\rangle_C |0\rangle_D ,$$

which satisfies $(\langle 0| \langle \eta(y^*)|) |u_{\text{yes}}\rangle = 1/(T+1) = \mathcal{O}(1/poly(N))$. We will now show that for the choice of $b := \Omega(1/T^6)$ and $\epsilon := \mathcal{O}(1/T^4)$, our reduction achieves the desired result.

Verifying classical evaluatability and quantum preparability Condition (i) follows directly from the definition of polynomially-sized subset states. For condition (ii) we have that $\langle u| O |u\rangle = \frac{1}{|S|} \sum_{i,j \in S} \langle i| O |j\rangle$, for which $\langle i| O |j\rangle$ can be computed efficiently since O, restricted to the Hilbert space it acts on, is described by a $2^{\mathcal{O}(\log n)} \times 2^{\mathcal{O}(\log n)} = poly(n) \times poly(n)$ matrix. Since we only have to compute $\langle i| O |j\rangle$ a total of $\mathcal{O}\left(|S|^2\right) = poly(n)$ times, this can be done efficiently. Finally, for condition (iii), we have that such states can be trivially prepared using $\mathcal{O}(poly(n))$ quantum gates by using a series of controlled rotations on each qubit at a time. For instance, a very simple application of the algorithm from Grover-Rudolph [GR02] would suffice.

Completeness and soundness Let us first analyse the YES-case. Here we have that the history state $|\eta(\tilde{y}^*)\rangle$ precisely gives $\langle \eta(\tilde{y}^*)| H_{yes} |\eta(\tilde{y}^*)\rangle = 0$. For a guiding state witness one can use Since $H_{no} \succeq b > 0$, we must have that $|\eta(\tilde{y}^*)\rangle |0\rangle$ is the ground state of H. However, we are also interested in the spectral gap, since we need it to satisfy our b-parameter as well as for the perturbative gadgets that reduce the locality. We now use a result from [DGF22],[19] Appendix B, which uses the Schrieffer-Wolff transformation to bound the shift caused in the energy levels due to H_{out}, allowing us to determine the relevant spectral gaps of H_{yes}. Let $H_0 = H_{\text{in}} + H_{\text{clock}} + H_{\text{prop}}$ and $H_1 = \epsilon H_{\text{out}}$, such that $\|H_1\| = \epsilon$. The ground space of H_0, denoted as S_0, is then given by

$$S_0 = \text{span}\{|\eta(\tilde{y})\rangle |\tilde{y} \in \{0,1\}^w\}.$$

We know that any state that has no support on S_0 must have energy at least $\Omega(1/T^3)$ [ADK+08], and therefore H_0 has a spectral gap of at least $\gamma(H_0) = \Omega(1/T^3)$ [ADK+08]. From [DGF22], we have that in the ground space of the original Hamiltonian H_0, H_{yes} has eigenvalues

$$\epsilon \frac{1 - P(\tilde{y})}{T+1} \pm \mathcal{O}\left(T^3 \epsilon^2\right), \tag{6}$$

[19] Similar bounds to the ones in [DGF22] can be found in other works (e.g. [KKR06,ADK+08,CLN17]), although we do not use these here.

provided that $\epsilon < \gamma(H_0)/16$. Let us now analyse the spectral gap of H_{yes} itself. If indeed $\epsilon < \gamma(H_0)/16$, we only have to consider states in the ground space of H_0, since all other states will have energies larger than $\Omega(1/T^3)$. We have that

$$\begin{aligned}\gamma(H_{yes}) &\geq \min_{\tilde{y}\in\{0,1\}^{w+1}\setminus\{\tilde{y}^*\}} \left[\epsilon\frac{1-P(\tilde{y})}{T+1} - \epsilon\frac{1-P(\tilde{y}^*)}{T+1} \pm \mathcal{O}\left(T^3\epsilon^2\right)\right] \\ &= \frac{\epsilon}{T+1}\left(1-2^{-k}\right) \pm \mathcal{O}\left(T^3\epsilon^2\right) \\ &= \Omega(1/T^5)\end{aligned}$$

if we set $\epsilon := \mathcal{O}(1/T^4)$. Note that this is in accordance with the condition that $\epsilon < \gamma(H_0)/16 = \Omega(1/T^3)$

Now for the NO-case. We have that all witnesses $\tilde{y}$ get accepted by $\hat{U}$ with at most an exponentially small probability, and hence have that $H_{yes} \succeq \Omega(1/T^5)$. By our choice b we have therefore ensured that the ground state in the NO-case must be the semi-classical polynomially sized subset state $|u_{\text{no}}\rangle := |0^q\rangle|1\rangle$, which has energy $b = \Omega(1/T^6)$. Hence, the promise gap between YES and NO cases is $\delta = \Omega(1/T^6) - 0 = \Omega(1/T^6)$.

Increasing the fidelity range Note that in the NO-case we already have that the ground state is a semi-classical poly-sized subset state. However, in the YES-case, the ground state is a history state with only inverse polynomial fidelity with the semi-classical poly-sized subset state $|u_{\text{yes}}\rangle$. To work around this, we apply the same trick as in [CFW22]:[20] by pre-idling the circuit with a polynomial number of identities, of which we denote the total number by N, the fidelity with a semi-classical poly sized subset state can be made inverse polynomially close to 1. Let $M = T + N = poly(n)$. For the new pre-idled circuit we have to replace in all our results throughout our construction T by M, and then everything goes through as before.

Reducing the locality Finally, we show how to reduce the locality of the constructed Hamiltonian. Assume that we have already increased the fidelity as above, so that the number of gates in the circuit is now M. Then the spectral gap of H, denoted as $\gamma(H)$, can be lower bounded as

$$\begin{aligned}\gamma(H) &\geq \min\left[\gamma(H|_{x\in\Pi_{yes}}), \gamma(H|_{x\in\Pi_{no}})\right] \\ &= \min\left[\delta, (\Omega(1/M^5) - \Omega(1/M^6))\right] \\ &= \Omega(1/M^6).\end{aligned}$$

Since the ground state is unique and inverse-polynomially gapped (in both the YES- and NO-case), we can apply Lemma 2.3 to obtain a 2-local Hamiltonian H' which (Δ, η, ϵ)-simulates H, where we can take $\Delta = 1/\,poly(n) \geq \gamma(H)$ sufficiently large, $\eta, \epsilon = 1/\,poly(n) \leq \delta$ sufficiently small to ensure that the ground energy remains below some a' in a YES instance and above some b' in a NO instance, such that $b' - a' = \delta = \Omega(1/\,poly(n))$ and so that $\|\mathcal{E}_{\text{state}}(|g\rangle) - |g'\rangle\| \leq \eta + \mathcal{O}(\gamma^{-1}\epsilon)$, where $|g\rangle$ is $|u_{\text{yes}}\rangle$ in a YES instance or $|u_{\text{no}}\rangle$ in a NO instance, $|g'\rangle$ is the ground state of H', and $\mathcal{E}_{\text{state}}(|g\rangle)$ is as in Lemma 2.3. That is, H' approximates H in the low energy spectrum (below Δ) in a way that the eigenvalues are perturbed by at most some small inverse-polynomial, and where the ground state can be approximated by the old ground state, plus some semi-classical state added as a tensor product. Finally, note that we can obtain $\|H\| \leq 1$ by simply scaling down by some polynomial, as required by the problem definition. Note that this will not change any of the statements as all relevant parameters and coefficients are (inverse) polynomials in n, albeit of very large degree. □

[20] This trick is not original to [CFW22] and is well known, see e.g. [CLN17].

Since polynomially-sized subset states are also samplable (see [GLG22]), our proof would also go through if one considers a variant of the guidable local Hamiltonian problem which considers samplable states as in Definition 3.1 instead. We have the following corollary.

Corollary 4.1. $\mathsf{CGaLH}^*(k,\delta,\zeta)$ *is* QCMA*-complete, where the hardness is under randomized reductions, for* $k \geq 2$, $\zeta \in (1/\,poly(n), 1 - 1/\,poly(n))$ *and* $\delta = 1/\,poly(n)$.

Proof. Hardness follows from Theorem 4.1. Containment follows trivially from the fact that the YES-and NO-cases can be distinguished by using $\mathrm{desc}(u)$ as a witness, and a verifier circuit that prepares the quantum state $|u\rangle$ (which can be done efficiently possible because of the extra condition on u) followed by quantum phase estimation to an accuracy strictly smaller than the promise gap δ, see Theorem 2 in [CFW22]. □

Now that we have established QCMA-completeness for CGaLH^*, we get QCMA-completeness for QGaLH *for free* for the same range of parameter settings, as the latter is a generalization of the former (containing CGaLH^* as a special case), and containment holds by the same argument as used in the proof of Corollary 4.1. However, with just a little bit of more work we can see that QCMA-hardness for QGaLH actually persists for a larger range of parameter settings. For this, we will use the following lemma by [LT20].

Lemma 4.2 (Ground state preparation with a-priori ground energy bound [LT20].)**.** *Suppose we have a Hamiltonian* $H = \sum_k \lambda_k(H) |\phi_q\rangle\langle\phi_q|$, *where* $\lambda_k(H) \leq \lambda_{k+1}(H)$, *given through its* $(\alpha, m, 0)$*-block-encoding* U_H. *Also suppose we have an initial state* $|\psi\rangle$ *prepared by some circuit* U_{prep} *with the promise that* $|\langle\phi_0|\psi\rangle|^2 \geq \Gamma$, *and that we have the following bounds on the ground energy and spectral gap:* $\lambda_0(H) \leq \mu - \Delta/2 < \mu + \Delta/2 \leq \lambda_1(H)$, *for some* $\mu, \Delta \in \mathbb{R}$. *Then the ground state can be prepared to fidelity* $1-\varepsilon$ *with probability* $1-\nu$ *with the following costs:*

1. *Query complexity:* $\mathcal{O}\left(\frac{\alpha}{\Gamma\Delta}\left(\log\left(\frac{\alpha}{\Delta}\right)\log\left(\frac{1}{\Gamma}\right)\log\left(\frac{\log(\alpha/\Delta)}{\nu}\right)+\log\left(\frac{1}{\varepsilon}\right)\right)\right)$ *queries to* U_H *and* $\mathcal{O}\left(\frac{1}{\Gamma}\log\left(\frac{\alpha}{\Delta}\right)\log\left(\frac{\alpha/\Delta}{\nu}\right)\right)$ *queries to* U_{prep},
2. *Number of qubits:* $\mathcal{O}\left(n+m+\log\left(\frac{1}{\Gamma}\right)\right)$,
3. *Other one- and two-qubit gates:* $\mathcal{O}\left(\frac{m\alpha}{\Gamma\Delta}\left(\log\left(\frac{\alpha}{\Delta}\right)\log\left(\frac{1}{\Gamma}\right)\log\left(\frac{\log(\alpha/\Delta)}{\nu}\right)+\log\left(\frac{1}{\varepsilon}\right)\right)\right)$

Theorem 4.2. $\mathsf{QGaLH}(k,\delta)$ *is* QCMA*-complete for* $k \geq 2$, $\delta = \Theta(1/\,poly(n))$ *and* $\zeta \in (1/\,poly(n), 1 - 1/\exp(n))$, *and even when* $\zeta = 1$ *in the case where* $k \geq 6$.

Proof. This follows immediately from the proof of Theorem 4.1, where the used history states themselves can be prepared by a quantum circuit of at most $poly(n)$ gates. Note that before we evoke locality reductions through the perturbative gadgets the witnesses can be made *exactly* by a quantum circuit[21]. In the locality reduction, we can only ensure that we remain inverse polynomially close to the original ground state. However, due to Lemma 4.2, this fidelity is enough to guarantee the existence of a quantum circuit, still polynomial in n, that produces a new quantum state which is inverse exponentially close to the actual ground state. Let U_{prep} be the quantum circuit that creates $|u\rangle = U_{\mathrm{prep}}|0\rangle$. Let us assume the worst case setting in our construction of Theorem 4.1, where we have that $\Gamma = \zeta = \Omega(1/\,poly(n))$, $\alpha = \mathcal{O}(1)$, $\Delta = \gamma(H) = \Omega(1/T^5)$, and $\mu = \lambda_0(H_{no}) = \Omega(1/T^6)$. Let $m = \mathcal{O}(1)$. Since the inverse fidelity ε appears only logarithmically in Lemma 4.2, we can prepare a state that is exponentially-close in fidelity with (exponentially) high probability in

$$\mathcal{O}\left(poly(n)T^5\left(\log\left(poly(n)T^5\right)\log\left(poly(n)\right)\log\left(\frac{\log(T^5)}{\nu}\right)+\log\left(\exp(n)\right)\right)\right) = \tilde{\mathcal{O}}\left(poly(n)\right)$$

[21] Provided that the circuit will have all gates from $\mathcal{G}$ in its gate set.

queries to U_H (the block encoding of H) and single qubit gates, as well as

$$\mathcal{O}\left(poly(n)\log\left(T^5\right)\log\left(\frac{T^5}{\nu}\right)\right) = \tilde{\mathcal{O}}\left(poly(n)\right)$$

queries to U_{prep}. □

There are a few more observations one can make about our proof of Theorem 4.1. First, before the perturbative gadgets are applied, the ground state energy is exactly zero in the YES-case and inverse polynomially large in the NO-case. This suggests that it might be possible to modify the construction such that it actually shows a result for a quantum satisfiability version for QCMA (which implies hardness for the more general local Hamiltonian version). In Appendix D we show that this is indeed possible. Second, since the spectral gap is bounded by some inverse polynomial, we also have as a direct corollary that 6-local *frustration-free* inverse polynomially gapped Hamiltonians are QCMA-hard, extending the result in [ABOBS22].

5 Classical containment via spectral amplification

To complement our quantum hardness results with classical containment results in certain parameter regimes, we will use a technique based on the dequantization of the quantum singular value transformation as described in [GLG22]. Our algorithm differs conceptually from the one proposed in [GLG22] in the following ways:

- We consider a different (and less restrictive) input model: whereas [GLG22] considers access to states of the form of Definition 3.1, we use states that adhere to the requirements as in Definition 3.2.
- For our purposes, we only consider local Hamiltonians (which are Hermitian sparse matrices) and not arbitrary sparse complex matrices. This simplifies the algorithm in the sense that we can view functions on these Hamiltonians as acting on the *spectrum* instead of the *singular values*.
- We also simplify the algorithm by tailoring it exactly to ground state *decision* instead of *estimation* problems, which allows us to use a different function acting on H as compared to [GLG22] to solve the relevant problems.

Let us introduce and prove bounds on the complexity of the *spectral amplification* algorithm in the next subsection. In the subsequent subsection, we will utilize this algorithm to put classical complexity upper bounds on $\mathsf{CGaLH}(k,\delta,\zeta)$ in specific parameter regimes.

5.1 Spectral amplification

Let $H = \sum_{i=0}^{m-1} H_i$ be a Hamiltonian on n qubits which is a sum of k-local terms H_i, which satisfies $\|H\| \leq 1$. Since H is Hermitian, we can write H as

$$H = \sum_{i=0}^{2^n-1} \lambda_i \left|\psi_i\right\rangle \left\langle\psi_i\right|,$$

where $\lambda_i \in [-1,1]$ (by assumption on the operator norm) denotes the i'th eigenvalue of H with corresponding eigenvector $|\psi_i\rangle$. Consider a polynomial $P \in \mathbb{R}[x]$ of degree d, and write

$$P(x) = a_0 + a_1 x + \cdots + a_d x^d.$$

The *polynomial spectral amplification* of H for P is then defined as

$$\begin{aligned} P(H) &= a_0 I + a_1 H + \cdots + a_d H^d \\ &= a_0 I + a_1 \sum_{i=0}^{2^n-1} \lambda_i \left|\psi_i\right\rangle \left\langle\psi_i\right| + \cdots + a_d \sum_{i=0}^{2^n-1} \lambda_i^d \left|\psi_i\right\rangle \left\langle\psi_i\right| \\ &= \sum_{i=0}^{2^n-1} P(\lambda_i) \left|\psi_i\right\rangle \left\langle\psi_i\right| . \end{aligned}$$

Now for $\alpha \in [-1, 1]$, denote

$$\Pi_\alpha = \sum_{\{i:\lambda_i \leq \alpha\}} \left|\psi_i\right\rangle \left\langle\psi_i\right| \tag{7}$$

for the projection on all eigenstates of H which have eigenvalues at most α, which we will call a *low-energy projector* of H. Note that for any $\alpha \geq \lambda_0$, we must have that $\Pi_{\text{gs}}\Pi_\alpha = \Pi_\alpha \Pi_{\text{gs}} = \Pi_{\text{gs}}$. We can utilize such a projector to solve $\mathsf{CGaLH}(k, \delta, \zeta)$, simply by computing $\|\Pi_\alpha \left|u\right\rangle\|$ for $\alpha = a$ given a classically evaluatable state u. To see why this works, note that in the YES-case, for the witness $\text{desc}(u)$ we have that $\|\Pi_a \left|u\right\rangle\| \geq \|\Pi_{\text{gs}} \left|u\right\rangle\| \geq \sqrt{\zeta}$ and in the NO-case we have that $\|\Pi_a \left|v\right\rangle\| = 0$ for states, which means that both cases are separated by $\sqrt{\zeta}$. However, it is unlikely that an efficient description exists of Π_a, and even if it did, it would not be k-local and therefore $\|\Pi_a \left|u\right\rangle\|$ would not even be necessarily efficiently computable.

The idea is now to approximate this low-energy projector Π_α by a polynomial in H. To see this, note that Π_α can be written exactly as

$$\Pi_\alpha = \frac{1}{2}\left(1 - \text{sgn}(H - \alpha I)\right),$$

where $\text{sgn}(x)$ is the sign function, which for our purposes is defined on $\mathbb{R} :\to \mathbb{R}$ as

$$\text{sgn}(x) = \begin{cases} 1 & \text{if } x > 0, \\ -1 & \text{if } x \leq 0. \end{cases}$$

From [HC17] we can then use the polynomial approximation of the sign function, which achieves optimal scaling in its parameters δ', ϵ. and which can subsequently be shifted to obtain the desired approximate low-energy projector $\tilde{\Pi}_a$.

Lemma 5.1 (Polynomial approximation to the sign function, from [HC17])**.** *For all $\delta' > 0, \epsilon' \in (0, 1/2)$ there exists an efficiently computable odd polynomial $P \in \mathbb{R}[x]$ of degree $d = \mathcal{O}\left(\frac{\log(1/\epsilon')}{\delta'}\right)$, such that*

- *for all $x \in [-2, 2] : |P(x)| \leq 1$, and*
- *for all $x \in [-2, 2] \setminus (-\delta', \delta') : |P(x) - sgn(x)| \leq \epsilon'$.*

Since Lemma 5.1 holds on the entire interval $[-2, 2]$, choosing any $\alpha \in [-1, 1]$ and scaling the $\text{sgn}(x)$ function with the factor $1/2$ will ensure that the error, as in the lemma, will be $\leq \epsilon/2$. Let $q_\alpha(x) : \mathbb{R} \to [0, 1]$ defined as $q_\alpha(x) = \frac{1}{2}(1 - \text{sgn}(x - \alpha))$ be this function , with polynomial approximation $Q_\alpha \in \mathbb{R}[x]$ of degree d. Note that Q_α can be written as a function of P as $Q_\alpha(x) = \frac{1}{2}(1 - P(x - \alpha)$. We will write $\tilde{\Pi}_\alpha = Q_\alpha(H)$ for the corresponding polynomial approximation of the approximate low-energy ground state "projector". Note that $\tilde{\Pi}_\alpha$ is Hermitian (since H is Hermitian), but that $\tilde{\Pi}_\alpha$ is no longer necessarily a projector and therefore $\tilde{\Pi}_\alpha^2 \neq \tilde{\Pi}_\alpha$. If we now replace Π_α in $\|\Pi_\alpha \left|u\right\rangle\|$ by $\tilde{\Pi}_\alpha$, we get $\left\|\tilde{\Pi}_\alpha \left|u\right\rangle\right\| = \sqrt{\langle u| \tilde{\Pi}_\alpha^\dagger \tilde{\Pi}_\alpha \left|u\right\rangle} = \sqrt{\langle u| \tilde{\Pi}_\alpha^2 \left|u\right\rangle} = \sqrt{\langle u| \left(Q_\alpha(H)\right)^2 \left|u\right\rangle}$, which means that we have to evaluate up to degree $2d$ powers of H. The next lemma will give an upper bound on the number of expectation values that have to be computed when evaluating a polynomial of H of degree d.

Lemma 5.2. *Given access to a classically evaluatable state u, a Hamiltonian $H = \sum_{i=0}^{m-1} H_i$, where each H_i acts on at most k qubits non-trivially, and a polynomial $P[x]$ of degree d, there exists a classical algorithm that computes $\langle u| P(H) |u\rangle$ in $\mathcal{O}(m^d)$ computations of $\langle u| O_i |u\rangle$, where the observables $\{O_i\}$ are at most kd-local.*

Proof. We have that

$$\begin{aligned}\langle u| P(H) |u\rangle &= \langle u| \left(a_0 I + a_1 H + \cdots + a_d H^d\right) |u\rangle \\ &= a_0 + a_1 \langle u| H |u\rangle + \cdots + a_d \langle u| H^d |u\rangle .\end{aligned}$$

Let $l \in [d+1]$ be the different powers for which we have to compute $\langle u| H^l |u\rangle$. By the multinomial theorem, we have that H^l can be written as

$$H^l = \left(\sum_{i=0}^{m-1} H_i\right)^l = \sum_{k_0+k_1+\cdots+k_{m-1}=l; k_0,k_1,\ldots,k_{m-1}\geq 0} \binom{l}{k_0, k_1, \ldots, k_{m-1}} \prod_{t=0}^{m-1} H_t^{k_t}, \tag{8}$$

where

$$\binom{l}{k_0, k_1, \ldots, k_{m-1}} = \frac{l!}{k_0! k_1! \ldots k_{m-1}!},$$

and we have that each term is at most kl local. The total number of expectation values that have to be computed in order to compute $\langle u| P(H) |u\rangle$ is upper bounded by (under the fact that $m \geq 1$, $l > 0$)

$$\begin{aligned}\sum_{l=0}^{d} \binom{m+l-1}{l} &\leq 1 + \sum_{l=1}^{d} \left(\frac{e(m+l-1)}{l}\right)^l \\ &\leq 1 + \sum_{l=1}^{d} \left(\frac{e(m+l)}{l}\right)^l \\ &\leq 1 + d\left(\frac{e(m+d)}{d}\right)^d \\ &\leq 1 + d\left(e + \frac{em}{d}\right)^d \\ &= \mathcal{O}\left(m^d\right),\end{aligned}$$

where we used an upper bound on the binomial coefficient of $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$ and the fact that $\left(\frac{e(m+l)}{l}\right)^l$ is monotonically increasing with respect to l, since

$$\frac{d}{dl}\left(\frac{e(m+l)}{l}\right)^l = \frac{e^l \left(\frac{l+m}{l}\right)^l \left((l+m)\log\left(\frac{l+m}{l}\right) + l\right)}{l+m}$$

is positive for all $m \geq 1$, $l > 0$. Hence, a classical algorithm to compute $\langle u| P(H) |u\rangle$ would be to evaluate the sum

$$\sum_{i=0}^{d} \sum_{k_0+k_1+\cdots+k_{m-1}=i; k_0,k_1,\ldots,k_{m-1}\geq 0} \binom{i}{k_0, k_1, \ldots, k_{m-1}} \prod_{t=0}^{m-1} a_i \langle u| H_t^{k_t}, |u\rangle ,$$

which requires at most $\mathcal{O}(m^d)$ computations of at most kd-local observables. □

All that remains to show is that for constant promise gap δ, using a good enough approximation $\tilde{\Pi}_\alpha$ with a suitable choice of α, will ensure that we can still distinguish both cases in the $\mathsf{CGaLH}(k, \delta, \zeta)$ problem in a polynomial (resp. quasi-polynomial) number of computations of observables when $\zeta = \Omega(1)$ (resp. $\zeta = \Omega(1/\,poly(n))$).

Theorem 5.1. *Let $H = \sum_{i=0}^{m-1} H_i$ be some Hamiltonian, and desc(u) be a description of a classically evaluatable state $u \in \mathbb{C}^{2^n}$. Let $a, b \in [-1, 1]$ such that $b - a \geq \delta$, where $\delta > 0$ is some constant. Consider the following two cases of H, with the promise that either one holds:*

(i) H has an eigenvalue $\leq a$, and $\|\Pi_{gs} |u\rangle\|^2 \geq \zeta$ holds, or

(ii) all eigenvalues of H are $\geq b$.

Then there exists a classical algorithm that is able to distinguish between cases (i) and (ii) using

$$\mathcal{O}\left(m^{c\left(\log\left(1/\sqrt{\zeta}\right)\right)/\delta)}\right)$$

computations of local expectation values, for some constant $c > 0$.

Proof. Let $\tilde{\Pi}_\alpha := Q_\alpha(H)$, where Q is a polynomial of degree d, be the approximate low-energy projector that approximates $\Pi_\alpha = \frac{1}{2}(1 + \mathrm{sgn}(H - (\alpha I)))$. We set $\alpha := \frac{a+b}{2}$, $\delta' := \delta/2$ and $\epsilon' = \epsilon$. Note that we can write $\tilde{\Pi}_\alpha$ as

$$\tilde{\Pi}_\alpha = \sum_{i=0}^{2^n-1} Q(\lambda_i) |\psi_i\rangle \langle\psi_i| ,$$

where we have that

$$\begin{cases} 1 - \epsilon/2 \leq Q(\lambda_i) \leq 1 & \text{if } \lambda_i \leq a, \\ 0 \leq Q(\lambda_i) \leq \epsilon/2 & \text{if } \lambda_i \geq b, \\ 0 \leq Q(\lambda_i) \leq 1 & \text{else,} \end{cases}$$

by Lemma 5.1. The algorithm will consist of calculating $\left\|\tilde{\Pi}_\alpha |u\rangle\right\|$, for sufficiently small ϵ. Let us analyse both cases.

(i) H has an eigenvalue $\leq a$, and $\|\Pi_{\text{gs}} |u\rangle\|^2 \geq \zeta$ holds:

$$\begin{aligned}
\left\|\tilde{\Pi}_\alpha |u\rangle\right\| &\geq \left\|\tilde{\Pi}_\alpha \Pi_{\text{gs}} |u\rangle\right\| \\
&= \left\|\Pi_\alpha \Pi_{\text{gs}} |u\rangle - (\Pi_\alpha - \tilde{\Pi}_\alpha)\Pi_{\text{gs}} |u\rangle\right\| \\
&= \left\|\Pi_{\text{gs}} |u\rangle - \left(\sum_{i:\lambda_i \leq \alpha} |\psi_i\rangle \langle\psi_i| - \sum_{i=0}^{2^n-1} Q(\lambda_i) |\psi_i\rangle \langle\psi_i|\right) \Pi_{\text{gs}} |u\rangle\right\| \\
&= \left\|\Pi_{\text{gs}} |u\rangle - \left(\sum_{i:\lambda_i \leq \alpha} (1 - Q(\lambda_i)) |\psi_i\rangle \langle\psi_i| - \sum_{i:\lambda_i > \alpha} Q(\lambda_i) |\psi_i\rangle \langle\psi_i|\right) \Pi_{\text{gs}} |u\rangle\right\| \\
&\geq \left\|\Pi_{\text{gs}} |u\rangle - \left(\sum_{i:\lambda_i \leq \alpha} \frac{\epsilon}{2} |\psi_i\rangle \langle\psi_i|\right) \Pi_{\text{gs}} |u\rangle\right\| \\
&= \left\|\Pi_{\text{gs}} |u\rangle - \frac{\epsilon}{2}\left(\sum_{i:\lambda_i \leq \alpha} |\psi_i\rangle \langle\psi_i|\right) \Pi_{\text{gs}} |u\rangle\right\| \\
&= \left\|\Pi_{\text{gs}} |u\rangle - \frac{\epsilon}{2}\Pi_\alpha \Pi_{\text{gs}} |u\rangle\right\| \\
&= \left\|\Pi_{\text{gs}} |u\rangle - \frac{\epsilon}{2}\Pi_{\text{gs}} |u\rangle\right\| \\
&= (1 - \frac{\epsilon}{2})\|\Pi_{\text{gs}} |u\rangle\| \\
&\geq (1 - \frac{\epsilon}{2})\sqrt{\zeta}.
\end{aligned}$$

(ii) all eigenvalues of H are $\geq b$:
We must have that

$$\left\|\tilde{\Pi}_\alpha |u\rangle\right\| \leq \epsilon/2,$$

since $\lambda_i \geq b$ for all $i \in \{0, \ldots, 2^n - 1\}$.

Therefore, to be able to distinguish both cases, we must have that $(1 - \epsilon/2)\sqrt{\zeta} - \epsilon/2 > 0$. We set $\epsilon := \sqrt{\zeta}$, such that the gap between both cases becomes equal to $\frac{1}{2}(\sqrt{\zeta} - \zeta)$. We therefore propose the following algorithm:

- $\zeta = 1$: compute $E = \langle u| H |u\rangle$ classically and check if $E \leq a$ or $E \geq b$. This takes m computations of local expectation values.
- $\zeta < 1$: compute $\left\|\tilde{\Pi}_a |u\rangle\right\|$ using a polynomial of degree $2d$ where $d = \mathcal{O}(\log(1/\epsilon'))/\delta'$, for $\epsilon' := \sqrt{\zeta}$ and $\delta' = \delta/2$. By Lemma 5.2, we have that this can be done in at most

  $$\mathcal{O}\left(m^{c\left(\log\left(1/\sqrt{\zeta}\right)\right)/\delta)}\right)$$

  computations of expectation values of local observables, for some constant c. The difference between both cases will be $1/2(\sqrt{\zeta} - \zeta) = \Omega(1/\,poly(n))$ when $\zeta = \Omega(1/\,poly(n)$.

□

Remark 5.1. It should be straightforward to adopt the same derivation as above to a more general setting by considering arbitrarily sparse matrices, a promise with respect to fidelity with the low-energy subspace (i.e. all states with energy $\leq \lambda_0 + \gamma$ for some small γ), as well as $\epsilon > 0$ for ϵ-classically evaluatable states (see Definition 3.2). However, this would likely put constraints on γ and ϵ, where ϵ in principle has to scale inversely proportional to the number of local terms in the Hamiltonian.

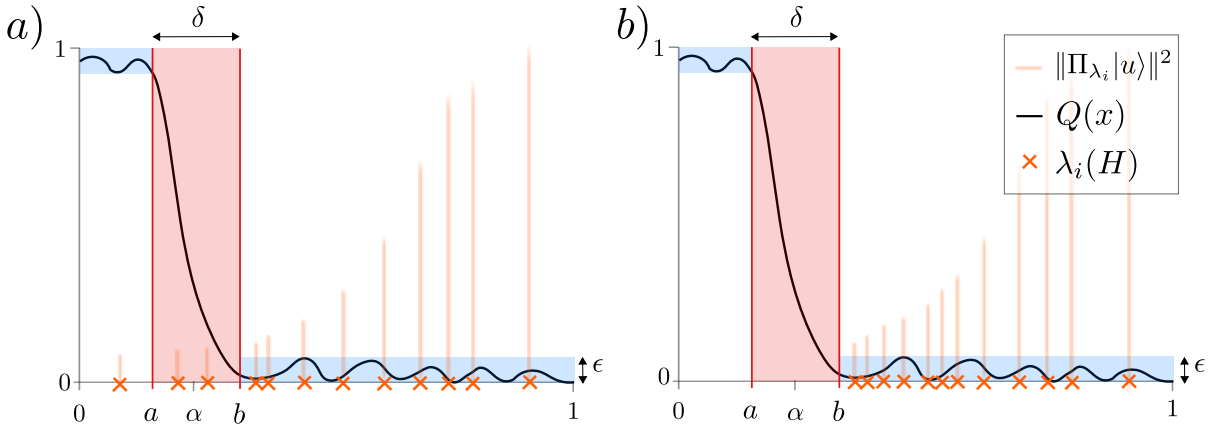

Figure 4: Illustration of the approximate low-energy projector Π_α in both the YES-case (Figure a)) and NO-case (Figure b)), with $\alpha = \frac{a+b}{2}$. The orange crosses correspond to the energy values, and the attached shaded lines indicate the fidelity of the guiding state with the space spanned by all eigenstates $|\psi_l\rangle$ of H that have energy at most λ_i. The polynomial approximation of the shifted sign function is displayed as $Q_\alpha(x)$, and the small error approximation regimes are indicated with the blue-shaded areas. In the red regime we do not have tight bounds on the error, except that the function values are in $[0, 1]$. For small enough ϵ, in the YES-case the contribution of the ground state to the value of $\left\|\tilde{\Pi}_\alpha |u\rangle\right\|^2$ should be larger than that computed in the NO-case due to contributions of higher energy values, as a result from an inexact implementation of the low-energy projector.

5.2 Classical hardness and containment

All results in this section also hold when 'CGaLH' is replaced by 'CGaLH^*', as the containment trivially follows since CGaLH generalises CGaLH^* and the hardness construction uses a diagonal (i.e. classical) Hamiltonian, of which the ground states are basis states and can thus be prepared on a quantum computer. To be able to make completeness statements when we consider NP, let us start by first proving a (straightforward) hardness result.

Lemma 5.3. $\mathsf{CGaLH}(k, \delta, \zeta)$ *is* NP*-hard for* $k \geq 2$*,* $\delta \leq \mathcal{O}(1)$ *and* $\zeta \leq 1$*, where* k, δ, ζ *can also be functions of* n*.*

Proof. We will prove this by a reduction from gapped 3-SAT. Let γ-3-SAT be a promise decision problem where we are given a formula $\phi(x) = \frac{1}{m}\sum_{i=0}^{m-1} C_i$ with $C_i = x_{i_1} \vee x_{i_2} \vee x_{i_3}$, with the promise that either $\phi(x) = 1$ (output YES) or $\phi(x) \leq \gamma$ (output NO), where $\gamma \in (0, 1)$. From (one of) the (equivalent) PCP theorem(s) we know that there exists a constant $\gamma \in (0, 1)$ for which deciding on the correct output (we are allowed to output anything if the promise doesn't hold) is NP-hard [H$\mathring{\text{o}}$1]. Next, we apply the gadget from [AL21], which maps $\phi(x)$ to a 2-SAT instance with formula $\phi'(x) = \frac{1}{10m}\sum_{i=0}^{m-1}\sum_{j\in[10]} C'_{i,j}$. Here we have that $\phi'(x)$ has the property that for every clause C_i there are 10 corresponding clauses $C'_{i,j}$, $j \in [10]$ such that, if a given assignment x satisfies a clause C_i of $\phi(x)$, then exactly 7 clauses of $C'_{i,j}$ can be satisfied, and for those C_i that are not satisfied by x, at most 6 clauses of $C'_{i,j}$ are satisfied. Note that if $\phi(x) = 1$, we then must have that $\phi'(x) = 7/10$, and that if $\phi(x) \leq \gamma$, we have that $\phi'(x) \leq 7/10\gamma + 6/10(1-\gamma) = (\gamma + 6)/10$. Hence, it is still NP-hard to distinguish between those cases, and the promise gap between the YES- and the NO-case is $\frac{1}{10}(1-\gamma) =: \gamma'$, which is some constant. Let us now map $\phi'(x)$ into a 2-local diagonal Hamiltonian H' such that $\langle x| H' |x\rangle = \phi(x)$. This can be done without any circuit-to-Hamiltonian mapping, see for example [Had21]. By our choice of $\phi'(x)$, we have already ensured that the Hamiltonian is (sub)-normalized. To turn the problem into a minimization problem, one can simply invert the

spectrum by letting $H = I - H'$ (note that $H' \succeq 0$). The eigenvectors of H are basis vectors – and thus themselves classically evaluatable states for which $\zeta = 1 = \mathcal{O}(1)$ – and its eigenvalues are precisely the function evaluations of $1 - \phi(x)$. Hence, setting $a := 3/10$ and $b := (4-\gamma)/10$ gives us $\delta = \gamma' = \mathcal{O}(1)$. □

Theorem 5.1 now gives us a very easy way to establish the following upper bounds on $\mathsf{CGaLH}(k, \delta, \zeta)$ when the required precision δ is only constant. Combined with Lemma 5.3, we obtain the following result, reminiscent of Theorem 5 in [GLG22].

Theorem 5.2. $\mathsf{CGaLH}(k, \delta, \zeta)$ *is* NP*-complete for* $k = \mathcal{O}(\log(n))$*, and constants* $\delta \in (0, 1]$ *and* $\zeta \in (0, 1]$*. Furthermore, when* $\zeta = \Omega(1/\,poly(n))$ *we have that* $\mathsf{CGaLH}(k, \delta, \zeta)$ *is in* NqP*.*

Proof. NP-Hardness follows from Lemma 5.3. The containment statements follow from Theorem 5.1, in which the proposed algorithm for $m = \mathcal{O}(n^k)$, $\delta \in (0, 1]$ constant runs in polynomial time when ζ is constant and in quasi-polynomial time when $\zeta = \Omega(1/\,poly(n))$ (using the fact that $\mathcal{O}(n^{\log(n)}) = 2^{\mathcal{O}(\log^c(n))}$ for some constant $c > 0$). □

Moreover, by a little more careful inspection one can show that the problem's hardness depends on how ζ and δ relate to one another, as shown in the following theorem.

Theorem 5.3. *Let* $f(n) : \mathbb{N} \to \mathbb{R}_{>0}$*,* $g(n) : \mathbb{N} \to \mathbb{R}_{>0}$ *be some arbitrary functions with the property that there exists some* n_0*, such that for all* $n < n_0$ *we that* $1/g(n) - 1/f(n) = \Omega(1/\exp(n))$*. Then we have that* $\mathsf{CGaLH}(k, \delta, \zeta)$ *is* NP*-complete for* $k \geq 2$*,* $\delta = 1/g(n)$ *and* $\zeta \geq 1 - 1/f(n)$*.*

Proof. Hardness: NP-hardness follows again trivially from Theorem 5.3, by taking $g(n) = \mathcal{O}(1)$ and letting $f(n)$ be some arbitrarily large function such that $f(n) \gg g(n)$, which gives a loose lower bound on ζ which can be as large as exactly 1.
Containment: To prove this we split the regime into the $n \geq n_0$ case and the $n < n_0$ case, giving separate algorithms for both cases.

$n < n_0$ **case:** In this setting we have that $n \in [1, n_0]$, which is a bounded domain. Therefore, we can run the algorithm considered in the proof of Theorem 5.1 with the worst case parameter settings on this interval, which is for $\delta^* = \max_{n \in [n_0+1]} g(n)$ and $\zeta^* = \min_{n \in [n_0+1]} f(n)$ [22]. Since both δ^* and ζ^* are then constants, this results in an algorithm that runs in a time polynomial in n and, as long as $1 \leq n \leq n_0$, be able to distinguish between the YES and NO-cases.

$n \geq n_0$ **case:** The verifier expect to be given a $\mathrm{desc}(u)$ such that $|\langle u|\phi_0\rangle|^2 \geq \zeta = 1 - 1/f(n)$ and checks if $\langle u| H |u\rangle \leq a + 1/f(n)$. Since H is a sum of local terms, by Lemma 5.2 the verifier can efficiently calculate $\langle u| H |u\rangle$ (since this is a polynomial in H of degree 1) when given $\mathrm{desc}(u)$. Let us now check completeness and soundness of this simple protocol. By the definition of the problem, we must have that $\|H\| \leq 1$. We write $|u\rangle = \alpha_1 |\phi_0\rangle + \alpha_2 \left|\phi_0^\perp\right\rangle$, where $|\alpha_1|^2 + |\alpha_2|^2 = 1$. Note that $|\alpha_2|^2 = 1 - \|\Pi_{\mathrm{gs}} |u\rangle\|^2 \geq 1 - \zeta \geq \frac{1}{f(n)}$. Therefore, we must have that in the YES-case

$$\begin{aligned}
\langle u| H |u\rangle &= |\alpha_1|^2 \langle\phi_0| H |\phi_0\rangle + |\alpha_2|^2 \left\langle \phi_0^\perp \right| H \left| \phi_0^\perp \right\rangle \\
&\leq a + |\alpha_2|^2 \left\langle \phi_0^\perp \right| H \left| \phi_0^\perp \right\rangle \\
&\leq a + \|H\|/f(n) \\
&\leq a + 1/f(n).
\end{aligned}$$

[22] Since n is a discrete variable on a bounded domain of constant size, one can always solve these optimization problems in constant time by brute-force search.

In the NO-case, we can simply evoke the variational principle

$$\langle u| H |u\rangle \geq \lambda_0(H) \geq b = a + 1/g(n) \geq a + 1/f(n) + \Omega(1/exp(n)),$$

for all $|u\rangle$. Therefore, both cases are inverse exponentially separated, and can therefore be distinguished from one another. □

6 Quantum-classical probabilistically checkable proofs

In this section we initiate the study of a new complexity class that sits right between the classical and quantum PCPs. First, let us recall some basic definitions and facts about PCPs.

Definition 6.1 (Probabilistically checkable proofs (PCPs))**.** *Let $n \in \mathbb{N}$ be the input size and $q : \mathbb{N} \to \mathbb{N}, r : \mathbb{N} \to \mathbb{N}$. A promise problem $A = (A_{yes}, A_{no})$ has a $(r(n), q(n))$-PCP verifier if there exists a polynomial-time probabilistic algorithm V which takes an input $x \in \{0,1\}^n$, and has random access to a string $\pi \in \{0,1\}^*$ of length at most $q(n)2^{r(n)}$, uses at most $r(n)$ random coins and makes at most $q(n)$ non-adaptive queries to locations of π, such that*

- ***Completeness.*** *If $x \in A_{yes}$, then there is a proof π such that $V^\pi(x)$ accepts with certainty.*
- ***Soundness.*** *If $y \in A_{no}$, then for all proofs π we have that $V^\pi(x)$ accepts with probability at most $\frac{1}{2}$.*

A promise problem $A = (A_{yes}, A_{no})$ belongs to $\mathsf{PCP}[q, r]$ if it has a $(r(n), q(n))$-PCP verifier.

The celebrated PCP theorem states that $\mathsf{NP} = \mathsf{PCP}[\mathcal{O}(1), \mathcal{O}(\log n)]$ [ALM+98, AS98]. It also implies that there exists a constant α such that it is NP-hard to decide for a constraint satisfaction problem with 'promise gap' α. Dinur [Din07] showed that this implication can be obtained directly, by reducing from a constraint satisfaction problem with inverse polynomial promise gap to one with a constant promise gap, whilst retaining NP-hardness. This type of reduction is commonly referred to as *gap amplification*.

A quantum variant to PCP can naturally be defined as follows [AALV09, AAV13].

Definition 6.2 (Quantum Probabilistically Checkable Proofs (QPCP))**.** *Let $n \in \mathbb{N}$ be the input size and $p, q : \mathbb{N} \to \mathbb{N}$, $c, s : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ with $c - s > 0$. A promise problem $A = (A_{yes}, A_{no})$ has a $(p(n), q(n), c, s)$-QPCP-verifier if there exists a quantum algorithm V which acts on an input $|x\rangle$ and a polynomial number of ancilla qubits, and takes an additional input a quantum state $|\xi\rangle \in \mathbb{C}^{2^{p(n)}}$, from which it is allowed to access at most $q(n)$ qubits, followed by a measurement of the first qubit after which it accepts only if the outcome is $|1\rangle$, such that*

> ***Completeness.*** *If $x \in A_{yes}$, then there is a quantum state $|\xi\rangle$ such that the verifier accepts with probability at least c,*
>
> ***Soundness.*** *If $y \in A_{no}$, then for all quantum states $|\xi\rangle$ the verifier accepts with probability at most s.*

A promise problem $A = (A_{yes}, A_{no})$ belongs to $\mathsf{QPCP}[p, q, c, s]$ if it has a $(p(n), q(n), c, s)$-QPCP verifier. If $p(n) \leq poly(n)$, $c = 2/3$, and $s = 1/3$, we simply write $\mathsf{QPCP}[q]$.

And likewise, there is a formulation of the *quantum* PCP conjecture, using the above notion quantum probabilistically checkable proofs (QPCPs).

Conjecture 6.1 (QPCP conjecture - proof verification version)**.** *There exists a constant $q \in \mathbb{N}$ such that*

$$\mathsf{QMA} = \mathsf{QPCP}[q].$$

As in the classical PCP theorem, there exist equivalent formulations of the QPCP conjecture. In particular, one can formulate the PCP theorem in terms of *inapproximability* of constraint satisfaction problems (CSPs), analogously to the classical setting. In the context of the quantum complexity classes (notably QMA), 'quantum' CSPs are generalized by local Hamiltonian problems. The formulation of the quantum PCP conjecture in terms of inapproximability of local Hamiltonians is:

Conjecture 6.2 (QPCP conjecture - gap amplification version)**.** *There exists a (quantum) reduction from the local Hamiltonian problem with promise gap $\Omega(1/\,poly(n))$ to another instance of the local Hamiltonian problem with promise gap $\Omega(1)$.*

It is well known that, at least under *quantum reductions*, both conjectures are in fact equivalent:

Fact 6.1 ([AALV09])**.** *Conjecture 6.1 holds if and only if conjecture 6.2 holds.*

6.1 Quantum-classical PCPs

We now consider the notion of a quantum PCP conjecture that, intuitively, conjectures the existence of polynomial-time quantum verifiers for QCMA problems that need only check a constant number of (the now classical) bits of the proof to satisfy constant completeness and soundness requirements. More formally, this reads as:

Definition 6.3 (Quantum-Classical Probabilistically Checkable Proofs (QCPCP))**.** *Let $n \in \mathbb{N}$ be the input size and $p, q : \mathbb{N} \to \mathbb{N}$, $c, s : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ with $c - s > 0$. . A promise problem $A = (A_{yes}, A_{no})$ has a $(p(n), q(n), c, s)$-QCPCP-verifier if there exists a quantum algorithm V which acts on an input $|x\rangle$ and a polynomial number of ancilla qubits, plus an additional bit string $y \in \{0,1\}^{2^{p(n)}}$ from which it is allowed to read at most $q(n)$ bits (non-adaptively), followed by a measurement of the first qubit, after which it accepts only if the outcome is $|1\rangle$, and satisfies:*

> ***Completeness.*** *If $x \in A_{yes}$, then there is a basis state state $|y\rangle$ such that the verifier accepts with probability at least c,*
>
> ***Soundness.*** *If $y \in A_{no}$, then for all basis states $|y\rangle$ the verifier accepts with probability at most s.*

A promise problem $A = (A_{yes}, A_{no})$ belongs to $\mathsf{QCPCP}[p, q, c, s]$ if it has a $(p(n), q(n), c, s)$-QCPCP verifier. If $p(n) = \mathcal{O}(poly(n))$, $c = 2/3$, and $s = 1/3$, we simply write $\mathsf{QCPCP}[q]$.

We remark that there are likely several ways to characterise a PCP for QCMA, with some being more or less natural than others. With that said, we believe that the above characterisation is well-motivated for the following reasons:

1. It is a natural definition following the structure of a QPCP verifier as in Definition 6.2, now with proofs given as in the standard definition of QCMA (see Definition 2.5).

2. $\mathsf{QCPCP}[\mathcal{O}(1)]$ captures the power of BQP as well as NP (via the PCP theorem), which are both believed to be strictly different complexity classes. Since techniques used to prove the PCP theorem are difficult (or impossible) to translate to the quantum setting [AAV13],

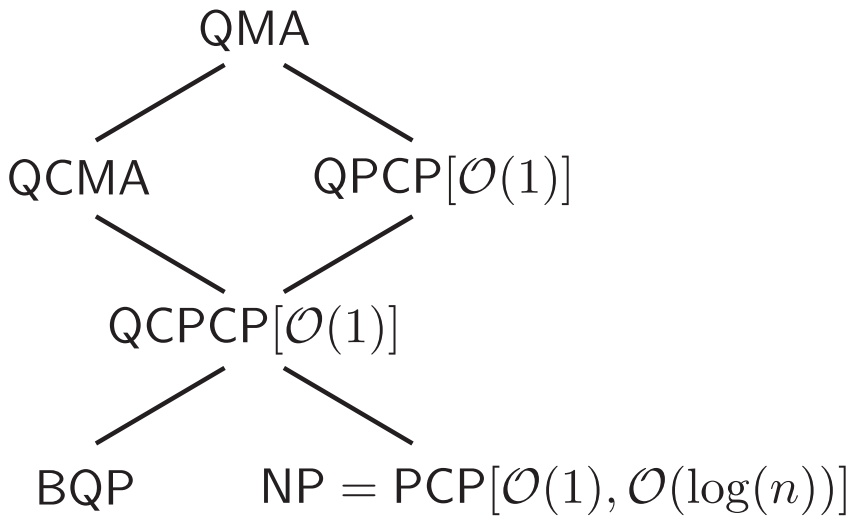

Figure 5: Known inclusions between some complexity classes and our proposed class QCPCP[k], with $k = \mathcal{O}(1)$. A line drawn from complexity class A to another class B, where B is placed above A means that $A \subseteq B$. Note that the complexity of our proposed class QCPCP[$\mathcal{O}(1)$] is non-trivial, as it contains both NP and BQP for which it is believed that both BQP $\not\subset$ NP and BQP $\not\supset$ NP.

studying QCPCP[$\mathcal{O}(1)$] might provide a fruitful direction with which to obtain the first non-trivial lower bound on the complexity of QPCP[$\mathcal{O}(1)$]. Indeed, the currently best known lower bound on the complexity of QPCP[$\mathcal{O}(1)$] is NP via the PCP theorem, so does not yet even capture BQP.

Given this definition for QCPCPs, our 'quantum-classical' PCP conjecture is naturally formulated as follows.

Conjecture 6.3 (quantum-classical PCP conjecture)**.** *There exists a constant $q \in \mathbb{N}$ such that*

$$\mathsf{QCMA} = \mathsf{QCPCP}[q].$$

If true, this conjecture would give a 'QCMA lower bound' on the power of quantum PCP systems, showing that a PCP theorem holds for (quantum) classes above NP, taking a step towards proving the quantum PCP conjecture. If it is false, but the quantum PCP conjecture is true, then this suggests that QPCP systems must take advantage of the quantumness of their proofs to obtain a probabilistically checkable proof system. In particular, since QCMA $\subseteq$ QMA, this would imply the existence of a quantum PCP system for every problem in QCMA, but *not* a quantum-classical one, even though the problem admits a classical proof that can be efficiently verified when we are allowed to look at all of its bits.

Note that we have that QCPCP[$\mathcal{O}(1)$] trivially contains CGaLH(k, ζ, δ) whenever $k = \mathcal{O}(\log n)$ $\zeta = \Omega(1)$ and $\delta = \Omega(1)$, since we have shown in Theorem 5.2 that this problem is in NP and therefore admits a classical PCP system to solve the problem. This would not apriori be clear if we considered guiding states as in Definition 3.1, as we do not know a PCP for the class MA.

6.1.1 Useful facts about QCPCP

We begin by showing two basic properties of QCPCP, which we make use of later on. First, we show the simple fact that QCPCP, just as is the case for PCP, allows for error reduction at the cost of extra queries to the proof.

Proposition 6.1 (Strong error reduction for QCPCP)**.** *The completeness and soundness parameters in* QCPCP *can be made exponentially close in some* $t \in \mathbb{N}$ *to* 1 *and* 0*, respectively, i.e.* $c = 1 - 2^{-\mathcal{O}(t)}$ *and* $s = 2^{-\mathcal{O}(t)}$*, by making* tq *instead of* q *queries to the classical proof.*

Proof. This follows from a standard parallel repetition argument, running the QCPCP protocol t times in parallel and taking a majority vote on the outcomes. Since the proof is classical, it does not matter if the same parts of the proof are queried multiple times by different runs, unlike the case when the proof is quantum. By taking a majority vote on the t outcomes, this yields the desired result by a Chernoff bound. □

Second, we show that when one is interested in $\mathsf{QCPCP}[\mathcal{O}(1)]$, the non-adaptiveness restriction in the definition does not limit the power of the class. The proof of this is similar to how one would prove it for $\mathsf{PCP}[\mathcal{O}(1), \mathcal{O}(\log n)]$, but encounters one difficulty. In the classical setting, the probabilistic PCP verifier can be replaced by a deterministic one that has an auxiliary input for a string which is taken uniformly at random. This way, one can 'fix the randomness' in the verifier, and the queries to the proof form a decision tree on the choices over the values of the proof bits, of which only one path can correspond to the actual proof. However, in the quantum setting there is no such equivalent notion of 'fixing quantumness', and therefore at every step the circuit might output different values for the indices of the proof that are supposed to be queried, even when all preceding steps gave the same indices and yielded identical values for the proof queries as compared to another run. Fortunately, it is easy to fix this by simply running the circuit many times in a simulated manner, using randomly chosen 'fake' proof bits. The probability that one matches the actual proof bits with the randomly chosen ones is small, but still constant when the number of queries to the proof is constant. We formalize this in the following theorem.

Theorem 6.1. *Let* QCPCP_A *be just as* QCPCP *but with the power to make adaptive queries to the proof. We have that*

$$\mathsf{QCPCP}[\mathcal{O}(1)] = \mathsf{QCPCP}_A[\mathcal{O}(1)].$$

Proof. The '$\subseteq$' is trivial since adaptive queries generalise non-adaptive ones, so we only have to show '$\supseteq$'. Let V be the circuit description of an adaptive $\mathsf{QCPCP}[\mathcal{O}(1)]$ protocol which makes q queries to a classical proof y. W.l.o.g., we can view V as a quantum circuit that consists of applying a unitary V_0 to the input and some ancilla qubits all initialized in $|0\rangle$, followed by a measurement to determine which bit of the proof to read, after which a new qubit is initialized in the basis state corresponding to the proof bit value, followed by another unitary V_1 that acts on all qubits, and so on, until at the end a measurement of a single designated output qubit is performed to decide whether to accept or reject (see Figure 6). We can simulate V on proof y non-adaptively in the following way. For a total of $T = C2^q$ times, with $C > 1$ some constant, we pick a $z \in \{0,1\}^q$ uniformly at random, and run circuit V where the answers to the proof queries are taken to be the bits of z. The final output qubit is measured as usual. This yields a total of T tuples $(z, i_1, \ldots, i_q, o)$, where the first entry indicates the value of the sampled $z \in \{0,1\}^q$, the second the q indices of the bits that were supposed to be queried by V, and the third the outcome of the final measurement, which has $o = 1$ if the circuit accepted and $o = 0$ if it rejected. For all T simulations, we can check if each $(z, i_1, \ldots, i_q)$ is consistent with the actual proof y, i.e. whether $z_{i_l} = y_{i_l}$ for all $l \in [q]$, which requires at most $qT = qC2^q = \mathcal{O}(1)$ queries to y. Since the probability that a randomly selected z is consistent with the proof y is $1/2^q$, we have that the probability that at least one of the z values in the tuples is consistent with the proof satisfies

$$\mathbb{P}[\text{At least one } z_{i_1}, \ldots, z_{i_q} \text{ consistent with } y] \geq 1 - \left(1 - \frac{1}{2^q}\right)^T \geq 1 - \frac{1}{1 + \frac{T}{2^q}} = \frac{T}{2^q + T} = \frac{C}{1 + C}.$$

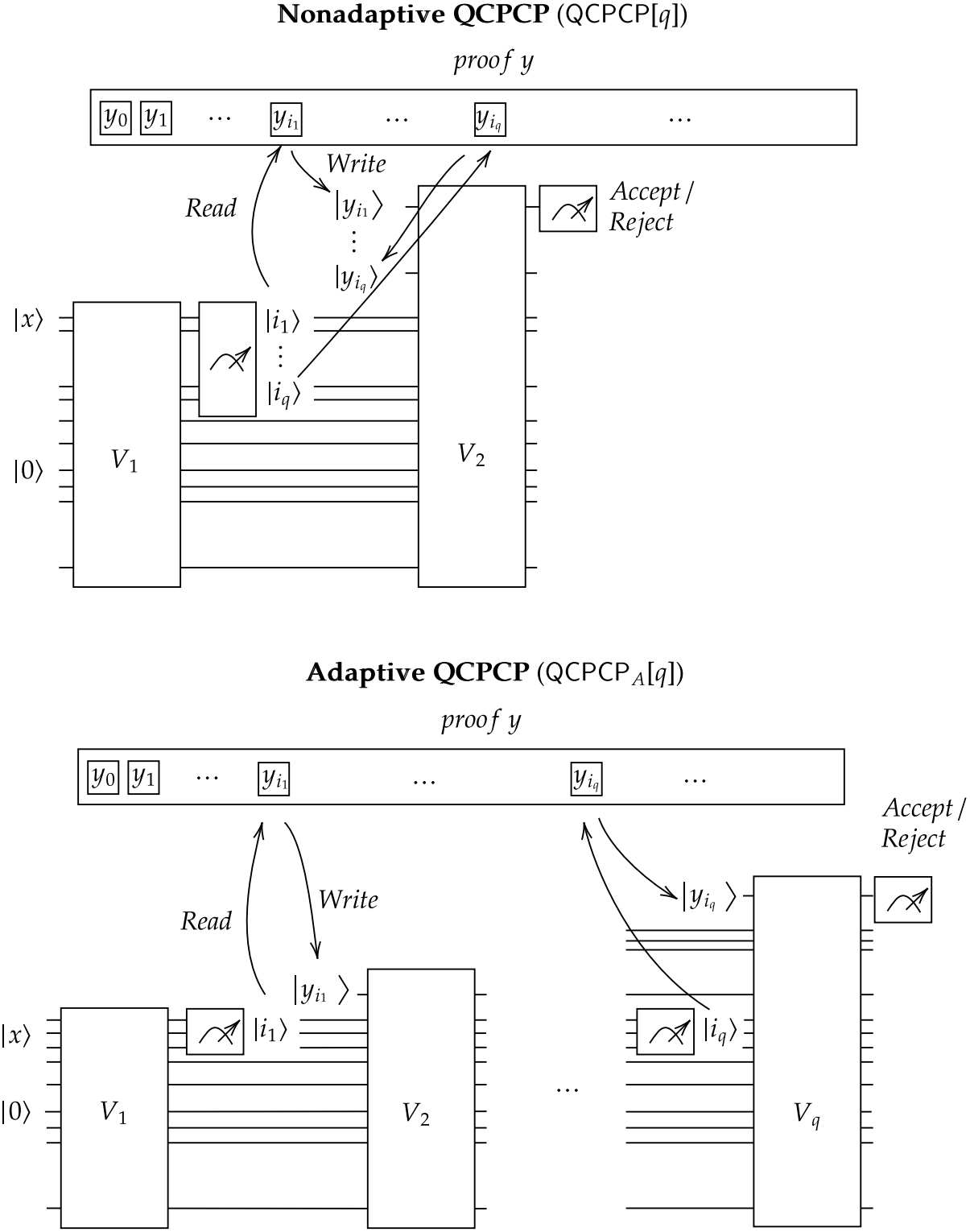

Figure 6: General quantum circuits for nonadaptive QCPCPs (QCPCP[q], top) versus adaptive ones (QCPCP$_A$[q], bottom). When we talk about making a 'query' to the classical proof we mean a single read/write procedure, where a single proof bit is read from the proof and then used to initialise a qubit into the basis state corresponding to the value of the bit. To determine which part of the proof must be read, a measurement of some designated index qubits is performed, which outputs a basis state(s) corresponding to index (indices) of the proof. The entire post-measurement state is allowed to be an input to a subsequent quantum circuit.

Now we loop over all collected tuples, and for the first tuple we encounter that is consistent with the proof we check o: if $o = 1$ we accept, and if $o = 0$ we reject. If all of the tuples are inconsistent with the proof, we reject. Hence, if the QCPCP_A protocol has completeness c and soundness s, this QCPCP protocol has completeness $c' = 2C/3(1 + C)$ and soundness $1/3$. We then have that the YES- and NO-cases are separated by at least

$$\frac{2C}{3(1+C)} - \frac{1}{3} = \Omega(1).$$

if $C > 1$. This can easily be boosted such that the completeness and soundness are $\geq 2/3$ and $\leq 1/3$ at the cost of a constant multiplicative factor in the number of queries, by adopting the strong error reduction of Proposition 6.1. Hence, $\mathsf{QCPCP}[\mathcal{O}(1)] \supseteq \mathsf{QCPCP}_A[\mathcal{O}(1)]$, completing the proof. □

Remark 6.1. The above proof actually holds for all quantum algorithms that have 'non-quantum' query access to a classical string and make only a constant of such queries.

There is an interesting observation one can make when looking more closely at the proof of Theorem 6.1: the resulting non-adaptive protocol employs a quantum circuit only for generating indices/output data needed for queries from the proof. Hence, once the queries are made, all subsequent checks can be performed classically. This means a quantum-classical PCP that is restricted to only using quantum circuits before the proof is accessed is just as powerful as the more general definition as per Definition 6.3. In the next subsection we will see that this idea allows one to put a non-trivial upper bound on the complexity of QCPCPs that make only a constant number of queries to the proof.

6.2 Upper bound on QCPCP with constant proof queries

Here we show that QCPCP with a constant number of proof queries is contained in $\mathsf{BQP}^{\mathsf{NP}[1]}$, i.e. in BQP with only a single query to an NP-oracle. The proof is rather long, but the idea is simple: just as is the case for QPCP, a *quantum* reduction can be used to transform a QCPCP system into a local Hamiltonian problem. However, since the proof is now classical, we can either use the CNOT-trick of Lemma 4.1 to ensure that the resulting Hamiltonian problem is diagonal in the computational basis, or directly learn a diagonal (i.e. classical) Hamiltonian that captures the input/output behaviour of the QCPCP-circuit on basis state inputs.[23] The main technical work required is to derive sufficient parameters in the reduction, thereby ensuring that the reduction succeeds with the desired success probability. We first prove a lemma which upper bounds the norm difference between a certain 'learned' Hamiltonian satisfying certain accuracy constraints, and the actual Hamiltonian.

Lemma 6.1. *Let $H = \sum_{i\in[m]} w_i H_i$ be a k-local Hamiltonian consisting of weights $w_i \in [0, 1]$ such that $\sum_{i\in[m]} w_i = W$, and k-local terms H_i for which $\|H_i\| \leq 1$ for all $i \in [m]$. Let $\Omega_{\geq\gamma} = \{i|w_i \geq \gamma\}$ and $\Omega_{<\gamma} = [m] \setminus \Omega_{\geq\gamma}$, for some parameter $\gamma \in [0, 1]$. Suppose $\tilde{H} = \sum_{i\in\Omega_{\geq\gamma}} \tilde{w}_i \tilde{H}_i$ is another Hamiltonian such that, for all $i \in \Omega_{\geq\gamma}$, we have $|\tilde{w}_i - w_i| \leq \epsilon_0$ and $\left\|H_i - \tilde{H}_i\right\| \leq \epsilon_1$. Then*

$$\left\|H - \tilde{H}\right\| \leq m(\gamma + \epsilon_0) + (W + m\epsilon_0)\epsilon_1$$

[23] We use the language of Hamiltonians here to be consistent with the QPCP setting, but this is not necessary. Since the Hamiltonian is classical, we might as well say that we learn a polynomial $P : \{0, 1\}^{p(n)} \to \mathbb{R}_{\geq 0}$.

Proof.

$$
\begin{aligned}
\left\|H-\tilde{H}\right\| &= \left\|\sum_{i\in\Omega_{\geq\gamma}} \tilde{w}_i\tilde{H}_i - \sum_{i\in[m]} w_i H_i\right\| \\
&\leq \sum_{i\in\Omega_{\geq\gamma}} \left\|\tilde{w}_i\tilde{H}_i - w_i H_i\right\| + \left\|\sum_{i\in\Omega_{<\gamma}} w_i H_i\right\| \\
&\leq m\gamma + \sum_{i\in\Omega_{\geq\gamma}} \left\|\tilde{w}_i\tilde{H}_i - w_i H_i\right\| \\
&\leq m\gamma + \sum_{i\in\Omega_{\geq\gamma}} \left\|\tilde{w}_i\left(\tilde{H}_i - H_i\right) + H_i\left(\tilde{w}_i - w_i\right)\right\| \\
&\leq m\gamma + \sum_{i\in\Omega_{\geq\gamma}} \left\|\tilde{w}_i\left(\tilde{H}_i - H_i\right)\right\| + \left\|H_i\left(\tilde{w}_i - w_i\right)\right\| \\
&\leq m\gamma + \sum_{i\in\Omega_{\geq\gamma}} \tilde{w}_i\left\|\tilde{H}_i - H_i\right\| + \|H_i\||\tilde{w}_i - w_i| \\
&\leq m\gamma + \sum_{i\in\Omega_{\geq\gamma}} \left(w_i+\epsilon_0\right)\epsilon_1 + \epsilon_0 \\
&\leq m(\gamma+\epsilon_0) + (W+m\epsilon_0)\epsilon_1,
\end{aligned}
$$

where in going from line 1 to line 2 we used the triangle inequality and the definition of $\Omega_{\geq\gamma}$, from line 3 to 4 again the definition of $\Omega_{\geq\gamma}$, from line 4 to line 5 the triangle inequality, from line 5 to 6 the absolute homogeneity of the norm, from 6 to 7 and 7 to 8 the properties on norm and absolute value differences as stated in the lemma. □

We will also use the fact that the learning of the Hamiltonian parameters in the reduction can be viewed as the well-known 'Double dixie cup' problem [New60]. In this problem, which is a generalization of the coupon collector problem, a collector wants to obtain m copies of each element from a set of n elements, via a procedure where every item is sampled with equal probability. In our setting, we have that the probability over the items is non-uniform. However, it is straightforward to obtain an upper bound on the non-uniform double dixie cup problem in terms of the expectation value in the uniform setting, as illustrated by the following lemma:

Lemma 6.2 (Upper bound on the non-uniform double dixie cup problem)**.** *Given samples from the set $S_n = [n]$, according to a distribution $\mathcal{P}$, where $\min_{i\in[n]}\mathcal{P}(i) \geq \gamma$, let $T_m^{\mathcal{P}}(S_n)$ be the random variable indicating the first time that all elements in S_n have been sampled at least m times according to distribution $\mathcal{P}$. We write $T_m(S_n)$ when the distribution over S_n is uniform. We have that*

$$\mathbb{E}[T_m^{\mathcal{P}}(S_n)] \leq \mathbb{E}[T_m(S_{\lceil 1/\gamma\rceil})],$$

where $S_{\lceil 1/\gamma\rceil} = [\lceil 1/\gamma\rceil]$ and

$$\mathbb{E}[T_m(S_{\lceil 1/\gamma\rceil})] = \lceil 1/\gamma\rceil \ln\lceil 1/\gamma\rceil + (m-1)\lceil 1/\gamma\rceil \ln\ln\lceil 1/\gamma\rceil + \mathcal{O}\left(\lceil 1/\gamma\rceil\right).$$

Proof. W.l.o.g., let the first n items of $S_{\lceil 1/\gamma\rceil}$ correspond to the items in S_n. We have for all $i \in [n]$ that $\mathcal{P}(i) \geq \gamma \geq \mathbb{P}[\text{sample } i \text{ from } S_{\lceil 1/\gamma\rceil}]$. Hence, if we have seen all elements of $S_{\lceil 1/\gamma\rceil}$ at least m times, we also have seen all elements of S_n at least m times, from which it follows that

$$\mathbb{E}[T_m^{\mathcal{P}}(n)] \leq \mathbb{E}[T_m(S_{\lceil 1/\gamma\rceil})],$$

where

$$\mathbb{E}[T_m(S_{\lceil 1/\gamma \rceil})] = \lceil 1/\gamma \rceil \ln \lceil 1/\gamma \rceil + (m-1)\lceil 1/\gamma \rceil \ln \ln \lceil 1/\gamma \rceil + \mathcal{O}\left(\lceil 1/\gamma \rceil\right)$$

is by the result of [New60]. □

We are now in a position to show the quantum reduction, which proceeds by learning the probability distribution over which indices of the proof string are queried by the verifier circuit, as well as the probability that the verification circuit V of the QCPCP-verifier accepts when those proof bits take on particular values.

Lemma 6.3. *Let $q \in \mathbb{N}$ be some constant and x an input with $|x| = n$. Consider a $\mathsf{QCPCP}[q]$ protocol with verification circuit V_x (which is V but with the input x hardcoded into the circuit), and proof $y \in \{0,1\}^{p(n)}$, and let*

$$\mathcal{P}_x(i_1, \ldots, i_q) = \mathbb{P}[V_x \text{ queries the proof at indices } (i_1, \ldots, i_q)]$$

and

$$\lambda_{x,(i_1,\ldots,i_q)}(z) = \mathbb{P}[V_x \text{ accepts given proof bits } i_1, \ldots, i_q \text{ are queried and are given by } y_{i_1} = z_1, \ldots, y_{i_q} = z_q].$$

Let $\Omega = [(i_1, \ldots, i_q) : i_j \in [p(n)], \forall j \in [q]]$, $\Omega_{\geq\gamma} = \{(i_1, \ldots, i_q) \in \Omega | \mathcal{P}_x(i_1, \ldots, i_q) \geq \gamma\}$ and $\Omega_{<\gamma} = \Omega \setminus \Omega_{\geq\gamma}$, for some parameter $\gamma \in [0,1]$. Then there exists a quantum algorithm that, for all $(i_1, \ldots, i_q) \in \Omega_{\geq\gamma}$ and all $z \in \{0,1\}^q$, provides estimates $\tilde{\mathcal{P}}_x(i_1, \ldots, i_q)$ and $\tilde{\lambda}_{x,(i_1,\ldots,i_q)}(z)$ such that

$$\left|\tilde{\mathcal{P}}_x(i_1, \ldots, i_q) - \mathcal{P}_x(i_1, \ldots, i_q)\right| \leq \epsilon_0,$$

and

$$\left|\tilde{\lambda}_{x,(i_1,\ldots,i_q)}(z) - \lambda_{x,(i_1,\ldots,i_q)}(z)\right| \leq \epsilon_1,$$

with probability $1-\delta$, and runs in time $\leq poly(n, 1/\gamma, \epsilon_0, \epsilon_1, 1/\delta)$.

Proof. Let $t \in [T]$ for some $T \in \mathbb{N}$ to be specified later, and let $O^{t,z} = ((i_1^{t,z}, \ldots, i_k^{t,z}), o^{t,z})$ be the tuple resulting from the t'th simulation of V_x, in which the proof y was supposed to be queried at indices $i_1, \ldots, i_q$ and in which those bits were assigned the values $y_{i_1} = z_1, \ldots, y_{i_q} = z_q$. Now consider the following quantum algorithm:

- For $z \in \{0,1\}^q$:
  - Simulate V for a total of T times to obtain samples $\{O^{t,z}\}_{t\in[T]}$.
  - For all observed $(i_1^{t,z}, \ldots, i_q^{t,z})$, set

$$\tilde{\lambda}_{x,(i_1,\ldots,i_q)}(z) := \frac{\#\text{ times } o^{t,z} = 1 | i_1, \ldots, i_q \text{ observed}}{\#\text{ times } i_1, \ldots, i_q \text{ observed}}.$$

- For all observed $(i_1^{t,z}, \ldots, i_q^{t,z})$, set

$$\tilde{\mathcal{P}}_x(i_1, \ldots, i_q) = \sum_{z\in\{0,1\}^q} \frac{\#\text{ times } (i_1^{t,z}, \ldots, i_q^{t,z}) \text{ observed}}{2^q T}.$$

The intuition is that in each tuple $\{O^{t,z}\}_{t\in[T]}$, the $(i_1^{t,z},\ldots,i_q^{t,z})$ give us information to estimate the $\mathcal{P}_x(i_1,\ldots,i_q)$'s and the combination $(i_1^{t,z},\ldots,i_q^{t,z}),o^{t,z}$ allows us to estimate the $\lambda_{x,(i_1,\ldots,i_q)}(z)$'s. Let us now show that there exists a T not too large such that the criteria of the theorem are satisfied. Since the $\mathcal{P}_x(i_1,\ldots,i_q)$ form a discrete distribution over the set Ω, we know by a standard result in learning theory (see for example [Can20]) that a total of

$$\Theta\left(\frac{|\Omega|+\log(1/\delta_0)}{\epsilon_0^2}\right)$$

samples of $O^{t,z}$ (the 'z'-value is in fact irrelevant here) suffices to get, with probability at least $1-\delta_0$, estimates $\tilde{\mathcal{P}}_x(i_1,\ldots,i_q)$ which satisfy

$$\left|\tilde{\mathcal{P}}_x(i_1,\ldots,i_q)-\mathcal{P}_x(i_1,\ldots,i_q)\right|\leq\epsilon_0.$$

To learn estimates $\tilde{\lambda}_{x,(i_1,\ldots,i_q)}(z)$ for a single index configuration $(i_1,\ldots,i_q)$ and proof configuration z, Hoeffding's inequality tells us that we only need

$$S_\lambda:=\mathcal{O}\left(\frac{\log\left(1/\delta_1\right)}{\epsilon_1^2}\right)$$

samples of $O^{t,z}$ to have that $\left|\tilde{\lambda}_{x,(i_1,\ldots,i_q)}(z)-\lambda_{x,(i_1,\ldots,i_q)}(z)\right|\leq\epsilon_1$, with probability $1-\delta_1$. By Lemma 6.2, we have that the expected number of samples needed such that this condition is met for all $i_1,\ldots,i_q\in\Omega_{\geq\gamma}$ is upper bounded by

$$\left\lceil\frac{1}{\gamma}\right\rceil\ln\left\lceil\frac{1}{\gamma}\right\rceil+(S_\lambda-1)\left\lceil\frac{1}{\gamma}\right\rceil\ln\ln\left\lceil\frac{1}{\gamma}\right\rceil+\mathcal{O}\left(\left\lceil\frac{1}{\gamma}\right\rceil\right),$$

which by Markov's inequality means that

$$\frac{1}{\delta_\lambda}\left(\left\lceil\frac{1}{\gamma}\right\rceil\ln\left\lceil\frac{1}{\gamma}\right\rceil+(S_\lambda-1)\left\lceil\frac{1}{\gamma}\right\rceil\ln\ln\left\lceil\frac{1}{\gamma}\right\rceil+\mathcal{O}\left(\left\lceil\frac{1}{\gamma}\right\rceil\right)\right)$$

samples of $O^{t,z}$ suffice to turn this into an algorithm that achieves success probability $\geq 1-\delta_\lambda$. To ensure that our entire algorithm succeeds with probability $1-\delta$, we require that

$$(1-\delta_\lambda)^{2^q}(1-\delta_0)(1-\delta_1)^{2^q\left\lceil\frac{1}{\gamma}\right\rceil}\geq 1-\delta,$$

which can be achieved by setting $\delta_\lambda=\delta/(2^{q+2})$, $\delta_0=\delta/4$ and $\delta_1=\delta/(\lceil 1/\gamma\rceil 2^{q+2})$. Then the total number of samples T that we must take satisfies

$$\begin{aligned}T&\leq\max\{\Theta\left(\frac{\left\lceil\frac{1}{\gamma}\right\rceil+\log(1/\delta_0)}{\epsilon_0^2}\right),\frac{2^q}{\delta_\lambda}\left(\left\lceil\frac{1}{\gamma}\right\rceil\ln\left\lceil\frac{1}{\gamma}\right\rceil+(T_p-1)\left\lceil\frac{1}{\gamma}\right\rceil\ln\ln\left\lceil\frac{1}{\gamma}\right\rceil+\mathcal{O}\left(\left\lceil\frac{1}{\gamma}\right\rceil\right)\right)\}\\&\leq\max\{\Theta\left(\frac{\left\lceil\frac{1}{\gamma}\right\rceil+\log\left(\frac{1}{\delta}\right)}{\epsilon_0^2}\right),\frac{2^{2(q+1)}}{\delta}\left(\left\lceil\frac{1}{\gamma}\right\rceil\ln\left\lceil\frac{1}{\gamma}\right\rceil+\mathcal{O}\left(\frac{q\log\left(\left\lceil\frac{1}{\gamma}\right\rceil/\delta\right)}{\epsilon_1^2}\right)\left\lceil\frac{1}{\gamma}\right\rceil\ln\ln\left\lceil\frac{1}{\gamma}\right\rceil\right)\}\end{aligned}$$

which yields a total runtime of $\mathcal{O}(poly(n,\lceil 1/\gamma\rceil,1/\delta,1/\epsilon_1,1/\epsilon_0))$ when $q=\mathcal{O}(1)$. □

We now have all the ingredients to obtain the main result of this section.

Theorem 6.2. *For all constant $q\in\mathbb{N}$, we have that*

$$\mathsf{QCPCP}[q]\subseteq\mathsf{BQP}^{\mathsf{NP}[1]}.$$

Proof. Consider a QCPCP protocol with verification circuit V_x (which is again V with input x hardwired into the circuit) which makes $q = \mathcal{O}(1)$ non-adapative queries to a proof $y \in \{0,1\}^{p(n)}$. Let $|\psi\rangle = |x\rangle \left|0^{l(n)}\right\rangle |y\rangle$. We have that[24]

$$\begin{aligned}\mathbb{P}[\mathsf{QCPCP} \text{ protocol accepts } y] &= \sum_{(i_1,\ldots,i_q)\in\Omega} \mathcal{P}_x(i_1,\ldots,i_q)\lambda_{x,(i_1,\ldots,i_q)}(y_{i_1},\ldots,y_{i_q}) \\ &= \langle y| \sum_{(i_1,\ldots,i_q)\in\Omega} \mathcal{P}_x(i_1,\ldots,i_q) \sum_{z\in\{0,1\}^q} \lambda_{x,(i_1,\ldots,i_q)}(z)|z\rangle\langle z| \, |y\rangle \\ &= 1 - \langle y| \, H_x \, |y\rangle \,,\end{aligned}$$

where H_x is a Hamiltonian that is diagonal in the computational basis, and takes the form

$$H_x = \sum_{(i_1,\ldots,i_q)\in\Omega} \mathcal{P}_x(i_1,\ldots,i_q) H_{x,(i_1,\ldots,i_q)} \,,$$

with q-local terms given by

$$H_{x,(i_1,\ldots,i_q)} = \sum_{z\in\{0,1\}^q} (1 - \lambda_{x,(i_1,\ldots,i_q)}(z))|z\rangle\langle z|.$$

Here $|z\rangle\langle z|$ should be read as $|z\rangle\langle z|_{i_1,\ldots,i_q}$, as it only acts on the q qubits with indices $i_1,\ldots i_q$. By Lemmas 6.1, 6.2, and 6.3, there exists a polynomial-time quantum algorithm that finds an estimated Hamiltonian

$$\tilde{H}_x = \sum_{(i_1,\ldots,i_q)\in\Omega_{\epsilon_D}} \tilde{\mathcal{P}}_x(i_1,\ldots,i_q)\tilde{H}_{x,(i_1,\ldots,i_q)},$$

where

$$\tilde{H}_{x,(i_1,\ldots,i_q)} = \sum_{z\in\{0,1\}^q} (1 - \tilde{\lambda}_{x,(i_1,\ldots,i_q)})(z)|z\rangle\langle z| \,,$$

such that $\left\|\tilde{H}_x - H_x\right\| \leq \epsilon$ with probability $\geq 1-\delta$. This means that with probability $\geq 1-\delta$, we have:

- $x \in P_{yes} \Rightarrow \exists y \in \{0,1\}^{p(n)} : \langle y| \, \tilde{H}_x \, |y\rangle \leq \frac{1}{3} + \epsilon$
- $x \in P_{no} \Rightarrow \forall y \in \{0,1\}^{p(n)} : \langle y| \, \tilde{H}_x \, |y\rangle \geq \frac{2}{3} - \epsilon \,.$

Note that $\frac{1}{3} + \epsilon$ and $\frac{2}{3} - \epsilon$ are separated by a constant for any constant $\epsilon < 1/6$. Since our reduction created a *diagonal* (and thus classical) Hamiltonian, which is a $\mathsf{CGaLH}(q,\zeta,\delta)$ instance with $\zeta = 1$ and $\delta = \Omega(1)$, the corresponding q-local Hamiltonian problem can be solved in NP.[25] The $\mathsf{BQP}^{\mathsf{NP}[1]}$ protocol would then consist of performing the quantum reduction to obtain the

[24] We write $1 - \langle y| \, H_x \, |y\rangle$ only to follow the convention that the local Hamiltonian problem is usually stated as a minimization problem.

[25] We would like to stress here that the NP containment only holds because all complexity classes considered are defined as classes of promise problems and not just languages. However, the promise on the final diagonal Hamiltonian eigenvalue problem can be removed by using that (i) every diagonal entry of each local term is only learned to a certain number of bits of precision and (ii) the eigenvalues of a sum of diagonal matrices are given by the sums of the eigenvalues (here we have to consider the full matrix, so also with the tensored identities which are left out in the notation). This way, one knows exactly what values the eigenvalues of the final Hamiltonian can take, and one can simply modify the completeness and soundness parameters such that they are both represented exactly by a certain number of bits and have a promise gap which is given by the difference between two successive numbers in the used binary representation. The promise gap will shrink considerably (and no longer be constant), but this is fine as the diagonal local Hamiltonian problem will still be in NP (it is even for an exponentially small promise gap).

$\mathsf{CGaLH}(q, \zeta, \delta)$ instance, followed by a single call to an NP oracle. This protocol succeeds with the same success probability as the reduction (which can be made exponentially close to 1 with polynomial overhead), which ensures completeness $\geq 1 - \delta \geq 2/3$ and soundness $\leq \delta \leq 1/3$ for any $\delta \leq 1/3$.

□

The above theorem yields an interesting implication in a world where the quantum-classical PCP conjecture is true, as illustrated by the following corollary.

Corollary 6.1. *If Conjecture 6.3 is true, then we have that* $\mathsf{NP}^{\mathsf{BQP}} \subseteq \mathsf{BQP}^{\mathsf{NP}}$.

Proof. We only have to show that $\mathsf{NP}^{\mathsf{BQP}} \subseteq \mathsf{QCMA}$, since if Conjecture 6.3 is true it implies that $\mathsf{QCMA} \subseteq \mathsf{QCPCP} \subseteq \mathsf{BQP}^{\mathsf{NP}[1]} \subseteq \mathsf{BQP}^{\mathsf{NP}}$ by Theorem 6.2. Let $q(n) : \mathbb{N} \to \mathbb{N}$ be a polynomial and $M^{\Pi}(x, y)$ be a deterministic polynomial-time verification circuit that uses as an additional input a proof y and make queries to a BQP oracle $\Pi = (\Pi_{\text{yes}}, \Pi_{\text{no}}, \Pi_{\text{inv}})$ at most $q(n)$ times as a black box. We define $z \in \{0, 1, \perp\}^{q(n)}$ as the string that describes the sets that each query input belonged to (here '$\perp$' indicates Π_{inv}). Define $I = \{i : z_i = 0 \text{ or } z_i = 1\}$. Since M^{Π} is deterministic, we have that any string z' that matches z on all indices I must produce the same output, so all of such z's can be considered correct query strings. For each $i \in I$, a QCMA verifier can, conditioned on all previous z_j for $j < i$ being computed such that it matches the first $i - 1$ bits of a correct query string, perform the BQP computation required to compute z_i with success probability exponentially close to 1 (this follows by the fact that BQP allows for probability amplification). If $i \in [q(n)] \setminus I$, any output will match in a correct query string in this case. Since there are a total of $q(n) \leq poly(n)$ such queries, the overall success probability of simulating a query sequence can be made $\geq 2/3$. The QCMA protocol then simply simulates M^{Π} by executing all operations in M directly and replacing every oracle call by a direct BQP-computation, which can be easily made to succeed with success probability $\geq 2/3$, ensuring completeness and soundness. □

An implication of Corollary 6.1 is that it can be used to show that under the assumption $\mathsf{NP} \subseteq \mathsf{BQP}$ and the quantum-classical PCP conjecture being true, we have that $\mathsf{PH} \subseteq \mathsf{BQP}$, where the class PH is the union of all complexity classes in the polynomial hierarchy, i.e. $\mathsf{PH} = \mathsf{P}^{\mathsf{NP}^{\mathsf{NP}^{\cdots}}}$. This follows from

$$\mathsf{NP}^{\mathsf{NP}} \subseteq \mathsf{NP}^{\mathsf{BQP}} \subseteq \mathsf{BQP}^{\mathsf{NP}} \subseteq \mathsf{BQP}^{\mathsf{BQP}} = \mathsf{BQP},$$

where the first and the third '$\subseteq$' are by assumption, the second is by the assumption of Conjecture 6.3 to be true and the last equality follows from the fact that BQP is self-low. We then have that $\mathsf{PH} \subseteq \mathsf{BQP}$ follows by induction, just as is the case for BPP [Zac88].[26] Moreover, this would also imply that under these assumptions $\mathsf{QCMA} \subseteq \mathsf{BQP}$, since

$$\mathsf{QCMA} \subseteq \mathsf{QCPCP}[\mathcal{O}(1)] \subseteq \mathsf{BQP}^{\mathsf{NP}} \subseteq \mathsf{BQP}^{\mathsf{BQP}} \subseteq \mathsf{BQP}.$$

Both of these implications would provide further evidence that it is unlikely that $\mathsf{BQP} \subseteq \mathsf{NP}$.

It is known that there exists an oracle relative to which the conclusion of Corollary 6.1 is not true, i.e. there exists an oracle A relative to which $\mathsf{NP}^{\mathsf{BQP}^A} \not\subset \mathsf{BQP}^{\mathsf{NP}^A}$ [AIK21]. Nevertheless, this does not necessarily mean the premise (i.e. the quantum-classical PCP conjecture) is false: one can also easily construct an oracle separation between PCP and NP, and both classes are now known to be equal [For94]. However, this suggests that, if Conjecture 6.3 is true, showing so likely requires non-relativizing techniques, just as was the case for the PCP theorem.

[26] See also `https://blog.computationalcomplexity.org/2005/12/pulling-out-quantumness.html`.

7 Implications to the quantum PCP conjecture

In this final section, we consider some implications from all previous sections to the quantum PCP conjecture. We find that our results give insights into the longstanding open question of whether the reduction from a QPCP verifier to a local Hamiltonian with constant promise gap can be made classical. Furthermore, we give a no-go result for the existence of quantum gap amplification procedures exhibiting certain properties (unless QCMA = NP or QCMA ⊆ NqP), and our results allow us to pose a conjecture which generalizes NLTS (now Theorem [ABN22]) and provides an alternative to NLSS [GLG22].[27]

7.1 'Dequantizing' the QPCP-to-local-Hamiltonian quantum reduction

Recall Fact 6.1, which states that the two types of QPCP conjectures (i.e. the proof verification and local Hamiltonian problem formulations) are known to be equivalent under *quantum* reductions. It has also been a longstanding open question whether the QPCP-to-local-Hamiltonian reduction can be made classical [AAV13]. However, our definition of QCPCP[q] shows that it is unlikely that a reduction having exactly the same properties as the known quantum reduction exists, as illustrated by the following theorem.

Theorem 7.1 (No-go for classical polynomial-time reductions)**.** *Let $q \in \mathbb{N}$ constant. For all constant $\epsilon < 1/6$ there cannot exist a classical polynomial-time reduction from a* QPCP[q] *circuit V_x (which is the* QPCP[q] *verifier circuit V with the input x hardwired into it) to a $\mathcal{O}(q)$-local Hamiltonian H_x such that, given a proof $|\psi\rangle$,*

$$|\mathbb{P}[V_x \text{ accepts } |\psi\rangle] - (1 - \langle\psi| H_x |\psi\rangle)| \leq \epsilon,$$

unless QCPCP[q] ⊆ NP *(which would imply* BQP ⊆ NP*).*

Proof. This follows directly from the fact that a QPCP[q] protocol can simulate a QCPCP[q] protocol by using the exact same verification circuit V_x and asking for basis states as quantum proofs (this can be forced by measuring the q qubits before any quantum operation acts upon them). Since one can compute $\langle y| H_x |y\rangle$ efficiently classically for any basis state y and local Hamiltonian H_x, we have that the behaviour of the QCPCP-system given a proof y can be evaluated up to precision ϵ by computing $\langle y| H_x |y\rangle$, where H_x is the $\mathcal{O}(q)$-local Hamiltonian induced by the QPCP verification circuit V_x. For any constant $\epsilon < (c-s)/2$ completeness and soundness are ensured. □

7.2 Gap amplifications

We now consider the implications of Section 5 to gap amplifications of guidable local Hamiltonian problems.

Theorem 7.2 (No-go's for quantum-classical gap amplification)**.** *There cannot exist*

1. *A polynomial time classical reduction from an instance of* CGaLH(k, ζ, δ) *with $k \geq 2$, some constant $\zeta > 0$, and $\delta = \Theta(1/\,poly(n))$ to some* CGaLH(k', ζ', δ') *with $k' \geq 2$, some constant $\zeta' > 0$, and $\delta' = \Omega(1)$,*

unless QCMA = NP*, and*

2. *A quasi-polynomial time classical reduction from an instance of* CGaLH(k, ζ, δ) *with $k \geq 2$, $\zeta = \Omega(1/\,poly(n))$, and $\delta = \Theta(1/\,poly(n))$ to some* CGaLH(k', ζ', δ') *with $k' \geq 2$, $\zeta' = \Omega(1/\,poly(n))$, and $\delta' = \Omega(1)$,*

[27] See also [CCNN23], which proposes a closely-related conjecture independently of this work.

unless QCMA $\subseteq$ NqP.

Proof. These all follow directly from Theorem 5.2. □

One can also interpret the no-go results for gap amplifications (points 1 and 2 in the above theorem) in a more general setting: if one wants to prove the QPCP conjecture through a gap amplification procedure a la Dinur, the procedure needs to have the property that it doesn't preserve 'classically evaluatable' properties of eigenstates (it cannot even maintain an inverse polynomial fidelity with such states) unless at the same time showing that QCMA = NP (or QCMA $\subseteq$ NqP, which is also very unlikely)! Hence, this result can be viewed as a 'QCMA-analogy' to the result from [AG19], where the authors showed that the existence of quantum gap amplifications that preserve stoqasticity of Hamiltonians would imply that NP = MA. We also point out that it is possible that – even though the complexity of QGaLH and CGaLH* was in the *inverse polynomial* precision regime the same for all $\zeta \leq 1 - 1/\,poly(n)$ – it might very well be that their complexities will differ when considering a *constant* precision, as our containment results of Section 5 crucially use the properties of classically evaluatable states. We note that many of the above results can also be easily adopted to the MA setting. All obtained results and proofs are given in Appendix C.2.

7.3 Classically evaluatable states and QPCP

Finally, we close by formulating a new conjecture which can be viewed as a strengthening of the NLTS theorem, or as an alternative to the NLSS conjecture of [GLG22] in light of our results, and which must hold if the quantum PCP conjecture is true and QMA $\neq$ NP.

Conjecture 7.1 (NLCES (no low-lying classically-evaluatable states) conjecture)**.** *There exists a family of local Hamiltonians $\{H_n\}_{n\in\mathbb{N}}$, where each H_n acts on n qubits, and a constant $\beta > 0$, such that for every classically evaluatable state $|u\rangle \in \mathbb{C}^{2^n}$ as in Definition 3.2, we have that for sufficiently large n that $\langle u| H |u\rangle \geq \lambda_0(H_n) + \beta$.*

Taking into account our results about the containment of the constant-gapped classically guidable local Hamiltonian problem in NP – namely the insight that what really matters is the *fidelity* of a classically evaluatable state with the low-lying energy subspace of the Hamiltonian, and not the energy of the classically evaluatable state itself – we can also define a stronger version of the NLCES conjecture, which must hold if the quantum PCP conjecture holds.

Conjecture 7.2 (Strong-NLCES conjecture)**.** *There exists a family of local Hamiltonians $\{H_n\}_{n\in\mathbb{N}}$, where each H_n acts on n qubits, and a constant $\beta > 0$, such that for sufficiently large n we have that for all classically evaluatable states $|u\rangle \in \mathbb{C}^{2^n}$, as in Definition 3.2, we have that $\|\Pi_{\lambda_0(H_n)+\beta} |u\rangle\|^2 = o(1/\,poly(n))$. Here $\Pi_{\lambda_0(H_n)+\beta}$ is the projector onto the space spanned by eigenvectors of H with energy less than $\lambda_0(H_n) + \beta$.*

Note that the NLCES Conjecture is strictly weaker than the Strong-NLCES conjecture, and that both do not necessarily imply the QPCP conjecture.

Acknowledgements

The authors thank Sevag Gharibian and François Le Gall for useful discussions and comments on an earlier version of the manuscript, with a special thanks to François Le Gall for pointing us towards looking into the implications to the quantum PCP conjecture. We also thank Harry Burhman for his help regarding promise classes, Jonas Helsen for helpful discussions, and Ronald de Wolf for providing feedback on the introduction. We thank the anonymous reviewers for their helpful comments. CC was supported by Fermioniq B.V., and thanks QuSoft and CWI for their accommodation whilst this work was completed. MF and JW were supported by the Dutch

Ministry of Economic Affairs and Climate Policy (EZK), as part of the Quantum Delta NL programme.

References

[AALV09] Dorit Aharonov, Itai Arad, Zeph Landau, and Umesh Vazirani. The detectability lemma and quantum gap amplification. *STOC*, pages 417–426, May 2009. arXiv:0811.3412.

[Aar09] Scott Aaronson. Computational complexity: Why quantum chemistry is hard. *Nature Physics*, 5:707–708, October 2009.

[Aar10] Scott Aaronson. BQP and the polynomial hierarchy. *STOC*, pages 141–150, June 2010. arXiv:0910.4698.

[AAV13] Dorit Aharonov, Itai Arad, and Thomas Vidick. Guest column: the quantum PCP conjecture. *ACM SIGACT news*, 44(2):47–79, June 2013. arXiv:1309.7495.

[ABN22] Anurag Anshu, Nikolas P. Breuckmann, and Chinmay Nirkhe. NLTS hamiltonians from good quantum codes. *arXiv preprint*, June 2022. arXiv:2206.13228.

[ABOBS22] Dorit Aharonov, Michael Ben-Or, Fernando G.S.L. Brandão, and Or Sattath. The pursuit of uniqueness: Extending Valiant-Vazirani theorem to the probabilistic and quantum settings. *Quantum*, 6:668, March 2022. arXiv:0810.4840.

[ADK+08] Dorit Aharonov, Wim van Dam, Julia Kempe, Zeph Landau, Seth Lloyd, and Oded Regev. Adiabatic quantum computation is equivalent to standard quantum computation. *SIAM review*, 50(4):755–787, 2008. arXiv:quant-ph/0405098.

[AG04] Scott Aaronson and Daniel Gottesman. Improved simulation of stabilizer circuits. *Physical Review A*, 70(5):052328, 2004.

[AG19] Dorit Aharonov and Alex B. Grilo. Stoquastic PCP vs. Randomness. *FOCS*, pages 1000–1023, November 2019. arXiv:1901.05270.

[AGL20] Dorit Aharonov, Alex B. Grilo, and Yupan Liu. StoqMA vs. MA: the power of error reduction. *arXiv preprint*, October 2020. arXiv:2010.02835.

[AIK21] Scott Aaronson, DeVon Ingram, and William Kretschmer. The acrobatics of bqp. *arXiv preprint arXiv:2111.10409*, 2021.

[AL18] Tameem Albash and Daniel A. Lidar. Adiabatic quantum computation. *Rev. Mod. Phys.*, 90:015002, January 2018. arXiv:1611.04471 .

[AL21] Carlos Ansótegui and Jordi Levy. Reducing SAT to Max2SAT. *IJCAI*, pages 1367–1373, August 2021.

[ALM+98] 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.

[Ara11] Itai Arad. A note about a partial no-go theorem for quantum PCP. *Quantum Information & Computation*, 11(11-12):1019–1027, November 2011. arXiv:1012.3319.

[AS98] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: A new characterization of NP. *Journal of the ACM*, 45(1):70–122, January 1998.

[ASSZ18] Itai Arad, Miklos Santha, Aarthi Sundaram, and Shengyu Zhang. Linear time algorithm for quantum 2SAT. *Theory of Computing*, 14(1):1–27, March 2018. arXiv:1508.06340.

[BBMC20] Bela Bauer, Sergey Bravyi, Mario Motta, and Garnet K. Chan. Quantum algorithms for quantum chemistry and quantum materials science. *Chemical Reviews*, 120(22):12685–12717, October 2020. arXiv:2001.03685.

[BBT06] Sergey Bravyi, Arvid J. Bessen, and Barbara M. Terhal. Merlin-Arthur games and stoquastic complexity. *arXiv preprint*, November 2006. arXiv:quant-ph/0611021.

[BDLT08] 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.

[BDSW96] Charles H. Bennett, David P. DiVincenzo, John A. Smolin, and William K. Wootters. Mixed-state entanglement and quantum error correction. *Physical Review A*, 54(5):3824–3851, nov 1996.

[Ben80] Paul Benioff. The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines. *Journal of Statistical Physics*, 22(5):563–591, May 1980.

[BG22] 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.

[BGK22] Lennart Bittel, Sevag Gharibian, and Martin Kliesch. Optimizing the depth of variational quantum algorithms is strongly QCMA-hard to approximate. *arXiv preprint*, November 2022. arXiv:2211.12519.

[BH13] Fernando G.S.L. Brandao and Aram W. Harrow. Product-state approximations to quantum ground states. *STOC*, page 871–880, June 2013. arXiv:1310.0017.

[BJS11] Michael J Bremner, Richard Jozsa, and Dan J Shepherd. Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 467(2126):459–472, 2011.

[BLH+21] Vera von Burg, Guang Hao Low, Thomas Häner, Damian S. Steiger, Markus Reiher, Martin Roetteler, and Matthias Troyer. Quantum computing enhanced computational catalysis. *Physical Review Research*, 3(3):033055, July 2021.

[BMT15] Jacob D. Biamonte, Jason Morton, and Jacob W. Turner. Tensor network contractions for #SAT. *Journal of Statistical Physics*, 160:1389–1404, June 2015. arXiv:1405.7375.

[Bra06] Sergey Bravyi. Efficient algorithm for a quantum analogue of 2-SAT. *arXiv preprint*, February 2006. arXive:quant-ph/0602108.

[Bra15] Sergey Bravyi. Monte Carlo simulation of stoquastic Hamiltonians. *Quantum Information & Computation*, 15(13–14):1122–1140, October 2015. arXiv:1402.2295.

[Can20] Clément L Canonne. A short note on learning discrete distributions. *arXiv preprint arXiv:2002.11457*, 2020.

[CCNN23] Nolan J Coble, Matthew Coudron, Jon Nelson, and Seyed Sajjad Nezhadi. Local hamiltonians with no low-energy stabilizer states. *arXiv preprint arXiv:2302.14755*, 2023.

[CFG+23] 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)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2023.

[CFW22] Chris Cade, Marten Folkertsma, and Jordi Weggemans. Complexity of the guided local Hamiltonian problem: Improved parameters and extension to excited states. *arXiv preprint*, July 2022. `arXiv:2207.10097`.

[CGL+20] Nai-Hui Chia, András Gilyén, Tongyang Li, Han-Hsuan Lin, Ewin Tang, and Chunhao Wang. Sampling-based sublinear low-rank matrix arithmetic framework for dequantizing quantum machine learning. *STOC*, page 387–400, June 2020. `arXiv:190.06151`.

[CHM21] Jordan Cotler, Hsin-Yuan Huang, and Jarrod R. McClean. Revisiting dequantization and quantum advantage in learning tasks. *arXiv preprint*, December 2021. `arXiv:2112.00811`.

[CLN17] Libor Caha, Zeph Landau, and Daniel Nagaj. The Feynman-Kitaev computer's clock: bias, gaps, idling and pulse tuning. *arXiv preprint*, 2017. `arXiv:1712.07395`.

[CMP18] Toby S. Cubitt, Ashley Montanaro, and Stephen Piddock. Universal quantum Hamiltonians. *The National Academy of Science*, 115(38):9497–9502, September 2018. `arXiv:1701.05182`.

[Coo71] Stephen A. Cook. The complexity of theorem-proving procedures. *STOC*, pages 151–158, May 1971.

[DGF22] Abhinav Deshpande, Alexey V. Gorshkov, and Bill Fefferman. The importance of the spectral gap in estimating ground-state energies. *PRX Quantum*, 3(4):040327, December 2022. `arXiv:2207.10250`.

[Din07] Irit Dinur. The PCP theorem by gap amplification. *Journal of the ACM*, 54(3):12–es, June 2007.

[Dru11] Andrew Drucker. A PCP characterization of AM. *ICALP*, pages 581–592, July 2011. `arXiv:1002.3664`.

[Fey18] Richard P. Feynman. Simulating physics with computers. In *Feynman and computation*, pages 133–153. CRC Press, 2018.

[For94] Lance Fortnow. The role of relativization in complexity theory. *Bulletin of the EATCS*, 52:229–243, 1994.

[GHLGM22] Sevag Gharibian, Ryu Hayakawa, François Le Gall, and Tomoyuki Morimae. Improved hardness results for the guided local Hamiltonian problem. *arXiv preprint*, July 2022. `arXiv:2207.10250`.

[GLG22] Sevag Gharibian and François Le Gall. Dequantizing the quantum singular value transformation: hardness and applications to quantum chemistry and the quantum PCP conjecture. *STOC*, pages 19–32, June 2022. arXiv:2111.09079.

[GN16] David Gosset and Daniel Nagaj. Quantum 3-SAT is QMA_1-complete. *SIAM Journal of computing*, 45(3):1080–1128, 2016. arXiv:1302.0290.

[Gol06] Oded Goldreich. On promise problems: A survey. In *Theoretical Computer Science: Essays in Memory of Shimon Even*, pages 254–290. Springer, 2006.

[Got98] Daniel Gottesman. The heisenberg representation of quantum computers. *arXiv preprint*, 1998. arXiv:quant-ph/9807006.

[GP19] Sevag Gharibian and Ojas Parekh. Almost optimal classical approximation algorithms for a quantum generalization of Max-Cut. *LIPIcs*, 145:31:1–31:17, October 2019. arXiv:1909.08846.

[GR02] Lov Grover and Terry Rudolph. Creating superpositions that correspond to efficiently integrable probability distributions. *arXiv preprint*, August 2002. arXiv:quant-ph/0208112.

[Gri18] Alex B. Grilo. *Quantum proofs, the local Hamiltonian problem and applications*. PhD thesis, Université Sorbonne Paris Cité, April 2018.

[GSLW19] András Gilyén, Yuan Su, Guang Hao Low, and Nathan Wiebe. Quantum singular value transformation and beyond: Exponential improvements for quantum matrix arithmetics. *STOC*, page 193–204, June 2019. arXiv:1806.01838.

[GY19] Sevag Gharibian and Justin Yirka. The complexity of simulating local measurements on quantum systems. *Quantum*, 3:189, 2019.

[Had21] Stuart Hadfield. On the representation of boolean and real functions as Hamiltonians for quantum computing. *ACM Transactions on Quantum Computing*, 2(4):1–21, December 2021. arXiv:1804.09130.

[HC17] Guang Hao Low and Isaac L. Chuang. Hamiltonian simulation by uniform spectral amplification. *arXiv preprint*, July 2017. arXiv:1707.05391.

[Hå01] Johan Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, July 2001.

[JGS20] Dhawal Jethwani, François Le Gall, and Sanjay K. Singh. Quantum-inspired classical algorithms for singular value transformation. *LIPIcs*, 170:53:1–53:14, August 2020. arXiv:1910.05699.

[JKNN12] Stephen P. Jordan, Hirotada Kobayashi, Daniel Nagaj, and Harumichi Nishimura. Achieving perfect completeness in classical-witness quantum Merlin-Arthur proof systems. *Quantum Information & Computation*, 12(5-6):461–471, May 2012. arXiv:1111.5306.

[KKR06] Julia Kempe, Alexei Kitaev, and Oded Regev. The complexity of the local hamiltonian problem. *SIAM Journal on Computing*, 35(5):1070–1097, 2006. arXiv:quant-ph/0406180.

[KLR+08] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland. Randomized benchmarking of quantum gates. *Physical Review A*, 77(1), jan 2008.

[KSV02] Alexei Y. Kitaev, Alexander Shen, and Mikhail N. Vyalyi. *Classical and quantum computation*. Number 47. American Mathematical Society, 2002.

[Lev73] Leonid A. Levin. Universal sequential search problems. *Problemy peredachi informatsii*, 9(3):115–116, 1973.

[Liu07] Yi-Kai Liu. Consistency of local density matrices is qma-complete, 2007.

[Liu21] Yupan Liu. StoqMA Meets Distribution Testing. In Min-Hsiu Hsieh, editor, *16th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2021)*, volume 197 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 4:1–4:22, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[LLS+22] Hongbin Liu, Guang Hao Low, Damian S. Steiger, Thomas Häner, Markus Reiher, and Matthias Troyer. Prospects of quantum computing for molecular sciences. *Materials Theory*, 6(1):1–17, March 2022. `arXiv:2102.10081`.

[LLZ+22] Seunghoon Lee, Joonho Lee, Huanchen Zhai, Yu Tong, Alexander M. Dalzell, Ashutosh Kumar, Phillip Helms, Johnnie Gray, Zhi-Hao Cui, Wenyuan Liu, Michael Kastoryano, Ryan Babbush, John Preskill, David R. Reichman, Earl T. Campbell, Edward F. Valeev, Lin Lin, and Garnet Kin-Lic Chan. Is there evidence for exponential quantum advantage in quantum chemistry? *arXiv preprint*, August 2022. `arXiv:2208.02199`.

[LT20] Lin Lin and Yu Tong. Near-optimal ground state preparation. *Quantum*, 4:372, December 2020. `arXiv:2002.12508`.

[MR18] Dmitri Maslov and Martin Roetteler. Shorter stabilizer circuits via bruhat decomposition and quantum circuit transformations. *IEEE Transactions on Information Theory*, 64(7):4729–4738, 2018.

[MW04] Chris Marriott and John Watrous. Quantum Arthur-Merlin games. *CCC*, pages 275–285, June 2004. `arXiv:cs/0506068`.

[New60] Donald J Newman. The double dixie cup problem. *The American Mathematical Monthly*, 67(1):58–61, 1960.

[OIWF22] Bryan O'Gorman, Sandy Irani, James Whitfield, and Bill Fefferman. Intractability of electronic structure in a fixed basis. *PRX Quantum*, 3(2):020322, May 2022.

[Orú14] Román Orús. A practical introduction to tensor networks: Matrix product states and projected entangled pair states. *Annals of physics*, 349:117–158, October 2014. `arXiv:1306.2164`.

[PH11] David Poulin and Matthew B. Hastings. Markov entropy decomposition: A variational dual for quantum belief propagation. *Physical review letters*, 106(8):080403, February 2011. `arXiv:1012.2050`.

[PM21] Stephen Piddock and Ashley Montanaro. Universal qudit hamiltonians. *Communications in Mathematical Physics*, 382:721–771, 2021. `arXiv:1802.07130`.

[RT22] Ran Raz and Avishay Tal. Oracle separation of BQP and PH. *Journal of the ACM*, 69(4):1–21, August 2022.

[Sch11] Ulrich Schollwöck. The density-matrix renormalization group: a short introduction. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 369(1946):2643–2661, July 2011.

[SSV+05] Christian Schön, Enrique Solano, Frank Verstraete, J. Ignacio Cirac, and Michael M. Wolf. Sequential generation of entangled multiqubit states. *Physical review letters*, 95(11):110503, September 2005. `arXiv:quant-ph/0501096`.

[Ste03] A. M. Steane. Quantum computing and error correction, 2003.

[SWVC07] Norbert Schuch, Michael M Wolf, Frank Verstraete, and J Ignacio Cirac. Computational complexity of projected entangled pair states. *Physical review letters*, 98(14):140506, 2007.

[Tan19] Ewin Tang. A quantum-inspired classical algorithm for recommendation systems. *STOC*, pages 217–228, June 2019. `arXiv:1807.04271`.

[TCC+22] Jules Tilly, Hongxiang Chen, Shuxiang Cao, Dario Picozzi, Kanav Setia, Ying Li, Edward Grant, Leonard Wossnig, Ivan Rungger, George H. Booth, et al. The variational quantum eigensolver: a review of methods and best practices. *Physics Reports*, 986:1–128, November 2022. `arXiv:2111.05176`.

[TD04] Barbara M. Terhal and David P. DiVincenzo. Adaptive quantum computation, constant depth quantum circuits and Arthur-Merlin games. *Quantum Information & Computation*, 4(2):134–145, March 2004. `arXiv:quant-ph/0205133`.

[TME+18] Norm M. Tubman, Carlos Mejuto-Zaera, Jeffrey M. Epstein, Diptarka Hait, Daniel S. Levine, William Huggins, Zhang Jiang, Jarrod R. McClean, Ryan Babbush, Martin Head-Gordon, and K. Birgitta Whaley. Postponing the orthogonality catastrophe: efficient state preparation for electronic structure simulations on quantum devices. *arXiv preprint*, September 2018. `arXiv:1809.05523`.

[VMC08] Frank Verstraete, Valentin Murg, and J Ignacio Cirac. Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems. *Advances in physics*, 57(2):143–224, 2008.

[WJB03] Pawel Wocjan, Dominik Janzing, and Thomas Beth. Two QCMA-complete problems. *Quantum Information & Computation*, 3(6):635–643, November 2003. `arXiv:quant-ph/0305090`.

[Zac88] Stathis Zachos. Probabilistic quantifiers and games. *Journal of Computer and System Sciences*, 36(3):433–451, 1988.

A Perfect sampling access of MPS and stabilizer states

In this appendix we show that both matrix product states (MPS) and stabilizer states are samplable states, by checking all three conditions of Definition 3.1.

Matrix product states: Let u be a $N = 2^n$-dimensional vector described by an MPS of n particles, bounded bond dimension D and local particle dimension d.

(i) Let $\hat{i}$ be the bit representation of i. The algorithm Q_u can simply be the evaluation of $\mathrm{Tr}[A_1^{(s_1)} A_2^{(s_2)} \dots A_n^{(s_n)}]$ for $s = \hat{i}$, which can be done via a naive matrix multiplication algorithm in time $\mathcal{O}(nD^3)$, and thus clearly runs in time $\mathcal{O}\left(poly(\log(N)\right)$ when $d = \mathcal{O}(poly(n)$, $D = \mathcal{O}(poly(n)$.

(ii) We will use that expectation values that are a tensor product of 1-local observables can be computed efficiently for a MPS im time $\mathcal{O}(nd^2D^3)$ [VMC08]. We assume that m is already known (see item (iii)), and that our MPS is therefore normalized. The algorithm

$\mathcal{SQ}_u$ works as follows: one computes the probability that the first qubit is 1 by computing the expectation value of the 1-local projector $\Pi_1 = |1\rangle\langle 1|_1$. Let $p_1 = \langle u|\Pi_1|u\rangle$ and $p_0 = 1 - p_1$. The algorithm now samples a bit $j_1 \in \{0,1\}$ according to distribution $\{p_0, p_1\}$, and computes the expectation value of the 2-local projector $\Pi_{j_1 1}$ to obtain $p_{j_1,1}$ and $p_{j_1,0}$, from which again a bit is sampled according to the distribution $p_{j_1,0}, p_{j_1,1}$. This procedure is repeated for all $n-2$ remaining sites, which yields a sample j with probability $|u_j|^2$. The total time complexity of this procedure is $\mathcal{O}(n^2 d^2 D^3) = \mathcal{O}(poly(\log N))$, when $d = \mathcal{O}(poly(n))$ and $D = \mathcal{O}(poly(n))$, as desired.

(iii) m can easily be computed by considering the overlap of the MPS with itself, which can be done in time $\mathcal{O}(npD^3)$ as the overlap can be viewed as the expectation value of a 0-local observable.

Stabilizer states: Let $u \in \mathbb{C}^{2^n}$, $N = 2^n$, be a stabilizer state on n qubits.

(i) This follows from the fact that basis states are stabilizer states, and that there exists an algorithm $\mathcal{Q}_u$ that computes inner products between stabilizer states in time $\mathcal{O}(n^3) = \mathcal{O}(poly(\log N))$ [AG04].

(ii) This follows from the fact that stabilizer states can be strongly simulated (i.e. marginals can be computed), which allows for weak simulation as shown in [TD04] at overhead n for the cost of strong simulation. Using the strong simuation algorithm as in [AG04], this gives an algorithm $\mathcal{SQ}_u$ that runs in time $\mathcal{O}(n^3) = \mathcal{O}(poly(\log N))$.

(iii) $m = 1$ by definition.

B MPS to circuit construction

In this section, we show that any MPS on n qubits with bond dimension D can be implemented on a quantum computer up to distance ϵ, with respect to the 2-norm, in $\mathcal{O}(nD\log(D)^2\log(Dn/\epsilon))$ one- and two-qubit gates and a $\mathcal{O}(npoly(D))$-time classical pre-calculation. The result is based on a result from [SSV+05]. For completeness we will first repeat their result. Let $\mathcal{H}_A = \mathbb{C}^D$ and $\mathcal{H}^B = \mathbb{C}^2$ be the Hilbert spaces characterising a D-dimensional ancillary system and a single qubit, respectively. Then every MPS of the form

$$|\psi\rangle = \langle\phi_F| V_n \dots V_1 |\phi_I\rangle$$

with arbitrary maps $V_k : \mathcal{H}_A \mapsto \mathcal{H}_A \otimes \mathcal{H}_B$, and $|\phi_I\rangle, |\phi_F\rangle \in \mathcal{H}_A$ is equivalent to a state

$$|\psi\rangle = \left\langle \tilde{\phi_F} \right| \tilde{V}_n \dots \tilde{V}_1 \left| \tilde{\phi}_I \right\rangle$$

with $\tilde{V}_k : \mathcal{H}_A \mapsto \mathcal{H}_A \otimes \mathcal{H}_B$ isometries and such that the ancillary register decouples in the last step

$$\tilde{V}_n \dots \tilde{V}_1 \left| \tilde{\phi}_I \right\rangle = |\phi_F\rangle \otimes |\psi\rangle .$$

Note that this is the canonical form of the MPS and can be found using $\mathcal{O}(npoly(D))$ classical pre-calculation time. The isometries are of size $2D \times D$ acting on the auxiliary system sequentially and create one qubit each. Every $\tilde{V}_k$ can be embedded into a unitary $U_k : \mathcal{H}_A \otimes \mathcal{H}_B \mapsto \mathcal{H}_A \otimes \mathcal{H}_B$ of size $2D \times 2D$, acting on the auxiliary system and a qubit initialised in $|0\rangle$ such that $U_k \left| \tilde{\phi}_k \right\rangle \otimes |0\rangle = \tilde{V}_k \left| \tilde{\phi}_k \right\rangle$. This gives the quantum circuit

$$U_n \dots U_1 \left| \tilde{\phi}_I \right\rangle |0\rangle^{\otimes n} = \left| \tilde{\phi}_F \right\rangle |\psi\rangle .$$

$\left|\tilde{\phi}_I\right\rangle$ is a state in $\mathcal{H}_A$ which can be generated on $\lceil \log(D) \rceil$ qubits, up to normalisation. By the Solovay-Kitaev theorem, this state can be prepared up to distance ϵ by a circuit of

$$\mathcal{O}(\lceil \log(D) \rceil^2 D \log\big(\lceil \log(D) \rceil^2 D/\epsilon\big)) = \mathcal{O}(D \log(D)^2 \log(1/\epsilon)$$

two and one-qubit gates. The unitaries U_k act on $\lceil \log(D) \rceil + 1$ qubits hence they can be approximated up to error ϵ in

$$\mathcal{O}((\lceil \log(D) \rceil + 1)^2 (D+1) \log\big((\lceil \log(D) \rceil + 1)^2 (D+1)/\epsilon\big)) = \mathcal{O}(D \log(D)^2 \log(D/\epsilon)).$$

Note that because every unitary incurs an error ϵ the entire error can be bounded by $n\epsilon$, setting individual error to $\epsilon' = \frac{\epsilon}{n}$ ensures that the generated state is at most ϵ far from the desired state. This results in a circuit of complexity: $\mathcal{O}(nD \log(D)^2 \log(nD/\epsilon))$ generating the MPS up to normalisation.

C Results for MA

Let us define yet another class of guidable local Hamiltonian problems, which constrains the considered Hamiltonians to be of a specific form.

Definition C.1 (Classically Guidable Local Stoquastic Hamiltonian Problem)**.** *The* ***Classically Guidable Local Stoquastic Hamiltonian Problem****, shortened as* $\mathsf{CGaLSH}(k, \delta, \zeta)$*, has the same input, promise, extra promise and output as* $\mathsf{CGaLH}(k, \delta, \zeta)$ *but with the extra constraint that the considered Hamiltonian is stoqastic. For our purposes this means that all* m *k-local terms* H_i *of the Hamiltonian* $H = \sum_{i=0}^{m-1} H_i$ *have real, non-positive off-diagonal matrix elements in the computational basis.*

By adopting the same proof structure as we used to prove Theorem 4.1, we can obtain a similar result for CGaLSH. For this, we first define a coherent description of MA, denoted as MA_q.

Definition C.2 (Coherent classical verifier [BBT06])**.** *A coherent classical verifier is a tuple* $V = (n, n_w, n_o, n_+, U)$*, where*

$$\begin{aligned} n &= \text{number of input bits} \\ n_w &= \text{number of witness bits} \\ n_0 &= \text{number of ancillas } |0\rangle \\ n_+ &= \text{number of ancillas } |+\rangle \\ U &= \text{quantum circuit on } n + n_w + n_0 + n_+ \text{ qubits with } X, \text{ CNOT and Toffoli gates.} \end{aligned}$$

The acceptance probability of a coherent classical verifier V *on input string* $x \in \{0,1\}^n$ *and witness state* $|\psi\rangle \in \left(\mathbb{C}^2\right)^{\otimes n_w}$ *is defined as*

$$P[V; x, \psi] = \langle \psi_{in}| U^\dagger \Pi_{out} U |\psi_{in}\rangle ,$$

where $|\psi_{in}\rangle = |x\rangle |0\rangle^{\otimes n_0} |+\rangle^{\otimes n_+}$ *is the initial state and* $\Pi_{out} = \langle 0|0\rangle \otimes I_{else}$ *project the first qubit onto the state* $|0\rangle$*.*

Definition C.3 (MA_q)**.** *The class* $\mathsf{MA}_q[c, s]$ *is the set of all languages* $L \subset \{0,1\}^*$ *for which there exists a (uniform family of) coherent classical verifier circuit* V *such that for every* $x \in \{0,1\}^*$ *of length* $n = |x|$*,*

- *if* $x \in L$ *then there exists a poly*(n)*-qubit witness state* $|\psi_x\rangle$ *such that* $V(x, |\psi_x\rangle)$ *accepts with probability* $\geq c (= 2/3)$*,*
- *if* $x \notin L$ *then for every purported poly*(n)*-qubit witness state* $|\psi\rangle$*,* $V(x, |\psi\rangle)$ *accepts with probability* $\leq s (= 1/3)$*.*

Lemma C.1 ([BBT06])**.** $\mathsf{MA} = \mathsf{MA}_q$

C.1 Proof of MA-hardness of the Classically Guidable Local Stoquastic Hamiltonian problem

Theorem C.1. $\mathsf{CGaLSH}(k, \delta, \zeta)$ *is* MA*-hard for* $k \geq 7$*,* $\zeta \in (1/\,poly(n), 1 - 1/\,poly(n))$ *and* $\delta = 1/\,poly(n)$*.*

Proof. This follows by using a nearly identical construction as have used to prove Theorem 4.1, but then starting from a MA_q verification circuit (Definition C.2) instead of a quantum circuit and with some minor changes. Let us go through the relevant steps of the construction and verify that they preserve stoquasticity. We will not concern ourselves with reducing the locality, and hence the parts of the construction needed to achieve this in the QCMA setting are left out.

- We have that $\mathsf{MA}_q[c, s] = \mathsf{MA}_q[1, s]$, since this holds for MA and the MA_q verifier has the same acceptance probability as the MA verifier.
- The CNOT- and Marriot and Watrous-tricks can be applied as the CNOT gate is part of the allowed gate-set in Definition C.2.
- One can again use Kitaev's clock construction with a small penalty ϵ on H_{out}, where the only terms that are off-diagonal in the computational basis come from H_{prop} and H_{out}. Inspection of these terms confirms that the Hamiltonian is stoquastic [BBT06]. Since the Toffoli gate is a 3-local gate, we have that our Hamiltonian becomes 6-local.
- The block-encoding trick also preserves stoquasticity, as both blocks themselves are stoquastic. This does increase the locality of the Hamiltonian by 1, and therefore it becomes 7-local.
- Finally, the pre-idling can be done since the identity is trivially in any gate set.

□

Somewhat surprisingly, a variant of the guidable local Hamiltonian problem for stoqastic Hamiltonians was already considered in a work by Bravyi in 2015 [Bra15].[28] However, the considered promise on the guiding state is different, as can be read from the following definition.

Definition C.4 ([Bra15])**.** *Let* H *be a stoquastic Hamiltonian. We will say that* H *admits a guiding state if and only if there exists a pair of normalized* n*-qubit states* ψ, ϕ *with non-negative amplitudes in the standard basis such that* ψ *is the ground state of* H*, the function* $x \to \langle x|\phi\rangle$ *is computable by a classical circuit of size* $poly(n)$*, and*

$$\langle x|\phi\rangle \geq \frac{\langle x|\psi\rangle}{poly(n)} \quad \text{for all } x \in \{0,1\}^n.$$

Using the Projection Monte Carlo algorithm with a variable number of walkers, Bravyi was able to show MA-containment of a guided Stoquastic Hamiltonian problem that uses guiding states satisfying Definition C.4. Note that Definition C.4 implies that the fidelity satisfies $|\langle\psi|\phi\rangle|^2 = \Omega\,(1/\,poly(n))$, but that the converse is not necessarily true. Therefore, Bravyi's result cannot be directly used to obtain MA-containment of $\mathsf{CGaLSH}(k, \delta, \zeta)$. We leave this as an open problem for future work.

[28] See also [Liu21], which considers StoqMA circuit problems where the witnesses are promised to be of a more restricted form.

C.2 PCP statements for MA

For MA it is not so clear how to define a PCP statement in a proof verification version, as the verifiers used in the original PCP system are already probabilistic. In [AG19], the authors define stoqastic PCP (SPCP) in the following way:

Conjecture C.1 (SPCP conjecture - frustration-free Hamiltonian version [AG19])**.** *There exist constants $\epsilon > 0$, $k', d' > 0$ and an efficient gap amplification procedure that reduces the problem of deciding if a uniform stoquastic d-degree k-local Hamiltonian is frustration-free or at least inverse polynomially frustrated, to the problem of deciding if a uniform stoquastic d'-degree k'-local Hamiltonian is frustration-free or at least ϵ frustrated.*

Theorem C.2 ([AG19])**.** *If conjecture C.1 is true, then* MA = NP*.*

We obtain similar results for our guidable *stoquastic* Local Hamiltonian problem, which differs from [AG19] in the fact that the Hamiltonian does not have to be uniform and the energy decision parameters can be arbitrary real numbers (instead of zero or bounded away from zero, as is the case in deciding on the frustration of the Hamiltonian). However, this comes with the extra constraint that there has to exist a classically describable guiding state, as per Definition 3.2. Formally, we can define another version of a SPCP-conjecture in terms of guidable Hamiltonians in the following way.

Conjecture C.2 (SPCP conjecture - guidable Hamiltonian version)**.** $\mathsf{CGaSLH}(k, \zeta, \delta)$ *with $k \geq 2$, some constant $\zeta > 0$, and $\delta = \Omega(1)$ is* MA*-hard under classical poly-time reductions.*

Just as in the QCMA-setting (see Theorem 7.2), we obtain certain no-go results which must hold if Conjecture C.2 is true.

Theorem C.3. *Conjecture C.2 (SPCP conjecture) is true if and only if* MA = NP*.*

Proof. The '$\Longrightarrow$' implication follows from Theorem 5.2, since $\mathsf{CGaSLH}(k, \zeta, \delta)$ is a special case of $\mathsf{CGaLH}(k, \zeta, \delta)$. '$\Longleftarrow$' follows from the fact that the NP-hard Hamiltonian in Lemma 5.3 is diagonal, which therefore is also stoquastic. □

Furthermore, we can also give no-go results on the existence of stoquastic gap amplification procedures.

Theorem C.4. *There cannot exist*

1. **Stoquastic gap amplification:** *a polynomial-time classical reduction from an instance of $\mathsf{CGaSLH}(k, \zeta, \delta)$ with $k \geq 2$, some constant $\zeta > 0$, and $\delta = \Theta(1/\,poly(n))$ to some $\mathsf{CGaSLH}(k', \zeta', \delta')$ with $k' \geq 2$, some constant $\zeta' > 0$, and $\delta' = \Omega(1)$,*

unless MA = NP*, and*

2. **Stoquastic gap amplification (2):** *a quasi-polynomial time classical reduction from an instance of $\mathsf{CGaSLH}(k, \zeta, \delta)$ with $k \geq 2$, $\zeta = \Omega(1/\,poly(n))$, and $\delta = \Theta(1/\,poly(n))$ to some $\mathsf{CGaSLH}(k', \zeta', \delta')$ with $k' \geq 2$, $\zeta' = \Omega(1/\,poly(n))$, and $\delta' = \Omega(1)$,*

unless MA $\subseteq$ NqP*.*

Proof. This follows again from Theorem 5.2. □

Hence, these results might provide yet another way, next to [AG19], to derandomize the class MA (i.e. show that MA = NP).

D QCMA versions of Quantum SAT

The canonical NP-complete problem is satisfiability, and whilst the general local Hamiltonian problem can be viewed as the 'quantum analogue' of satisfiability, one might argue that an *even closer* analogue would be the slightly different problem of *quantum satisfiability*, introduced in [Bra06]. In quantum satisfiability, shortened as k-QSAT, the Hamiltonian is given by a sum of local projectors and the task is to decide whether there exists a quantum state $|\xi\rangle$ such that all terms project $|\xi\rangle$ to zero or the expectation value of H is greater or equal than some inverse polynomial, for all quantum states. k-QSAT was shown to be QMA_1-complete for $k \geq 3$ [Bra06, GN16], and it can be solved in linear time classically when $k \leq 2$ [ASSZ18].

In the case of QMA it is very much unclear whether $\mathsf{QMA} = \mathsf{QMA}_1$, but for QCMA we know that in fact $\mathsf{QCMA} = \mathsf{QCMA}_1$, as mentioned before. The simple observation that in our proof of Theorem 4.1 the ground state energy is exactly zero in the YES-case and inverse polynomially large in the NO-case, hints that our construction actually shows a result for a quantum satisfiability version for QCMA (which implies hardness for the more general local Hamiltonian version), and indeed with some very minor modifications this can be shown to be the case. First we formally define the 'QCMA versions' of satisfiability.

Definition D.1 (Guidable Quantum Satisfiability)**.** *Guidable Quantum Satisfiability Problems are a class of problems defined as having the following input, promise, one of the extra promises and output:*
Input: *A collection $\{\Pi_i : i = 1, \ldots, m\}$ of k-local projectors acting on n qubits, a precision parameter $\delta > 0$. Let the Hamiltonian $H = \sum_{i=1}^{m} \Pi_i$ be the sum of all these projectors.*
Promise: *Either there exists a state $|w\rangle$ s.t. $\Pi_i |w\rangle = 0$ for all $i = 1, \ldots, m$ (i.e. $\lambda_0(H) = 0$), or otherwise $\sum_{i=1}^{m} |\langle y|\Pi_i|y\rangle| \geq \delta$ for all states $|y\rangle \in \mathbb{C}^{2^n}$ (i.e. $\lambda_0(H) \geq \delta$).*
Extra promises: *Denote Π_{gs} for the projection on the subspace spanned by the ground state of H. Then for each problem class, we have that either one of the following promises hold, each giving a different problem:*

1. $\mathsf{CGa\text{-}QSAT}(k, \delta, \zeta)$ ***Classically Guidable Quantum Satifisfiability Problem*** *:*
   There exists a classically evaluatable state $u \in \mathbb{C}^{2^n}$ for which $\|\Pi_{gs} u\|^2 \geq \zeta$.

2. $\mathsf{QGa\text{-}QSAT}(k, \delta, \zeta)$ ***Quantumly Guidable Quantum Satifisfiability Problem****:*
   There exists a unitary V implemented by a quantum circuit composed of at most $T = poly(n)$ gates from a fixed gate set $\mathcal{G}$ that produces the state $|\phi\rangle = V|0\rangle$, which has $\|\Pi_{gs}|\phi\rangle\|^2 \geq \zeta$.

Output:

- *If $\lambda_0(H) = 0$, output* YES.
- *If $\lambda_0(H) \geq \delta$, output* NO.

Theorem D.1. $\mathsf{CGa\text{-}QSAT}(k, \delta, \zeta)$ *is* QCMA*-complete for $k \geq 6$, $\delta = \Theta(1/poly(n))$ and $\zeta \in (1/poly(n), 1 - 1/poly(n))$.* $\mathsf{QGa\text{-}QSAT}(k, \delta, \zeta)$ *is* QCMA*-complete for $k \geq 6$, $\delta = \Theta(1/poly(n))$ and $\zeta \in (1/poly(n), 1]$.*

Proof. We will follow the same hardness construction as was used to prove Theorem 4.1, making the following three observations: (i) Definition D.1 requires that the Hamiltonian is an unweighted sum of projectors. Note that the Hamiltonian of Eq. (1) is of this form except for the fact that H_{out} is weighted by a factor ϵ, which is needed in order to be able to use Eq. 6. We can easily work around this by instead of multiplying H_{out} by a factor ϵ we add $\lceil 1/\epsilon \rceil$ times each of the H_{in}, H_{clock} and H_{prop} terms, which allows one to follow the same analysis as when

ϵH_{out} was used instead. (ii) H_{no} of Eq. (4) is a sum of projectors, and the block-encoding trick used to construct H of Eq. (5) ensures that all terms remain projectors. (iii) The ground state energy is (before the locality reduction through the perturbative gadgets, which we do not apply here) precisely 0 in the YES-case and inverse polynomially far from zero in the NO-case. This also ensures containment through the use of quantum phase estimation. Also, note that we do not have to scale down the Hamiltonian as $\|H\| \leq 1$ is not required by Definition D.1. □