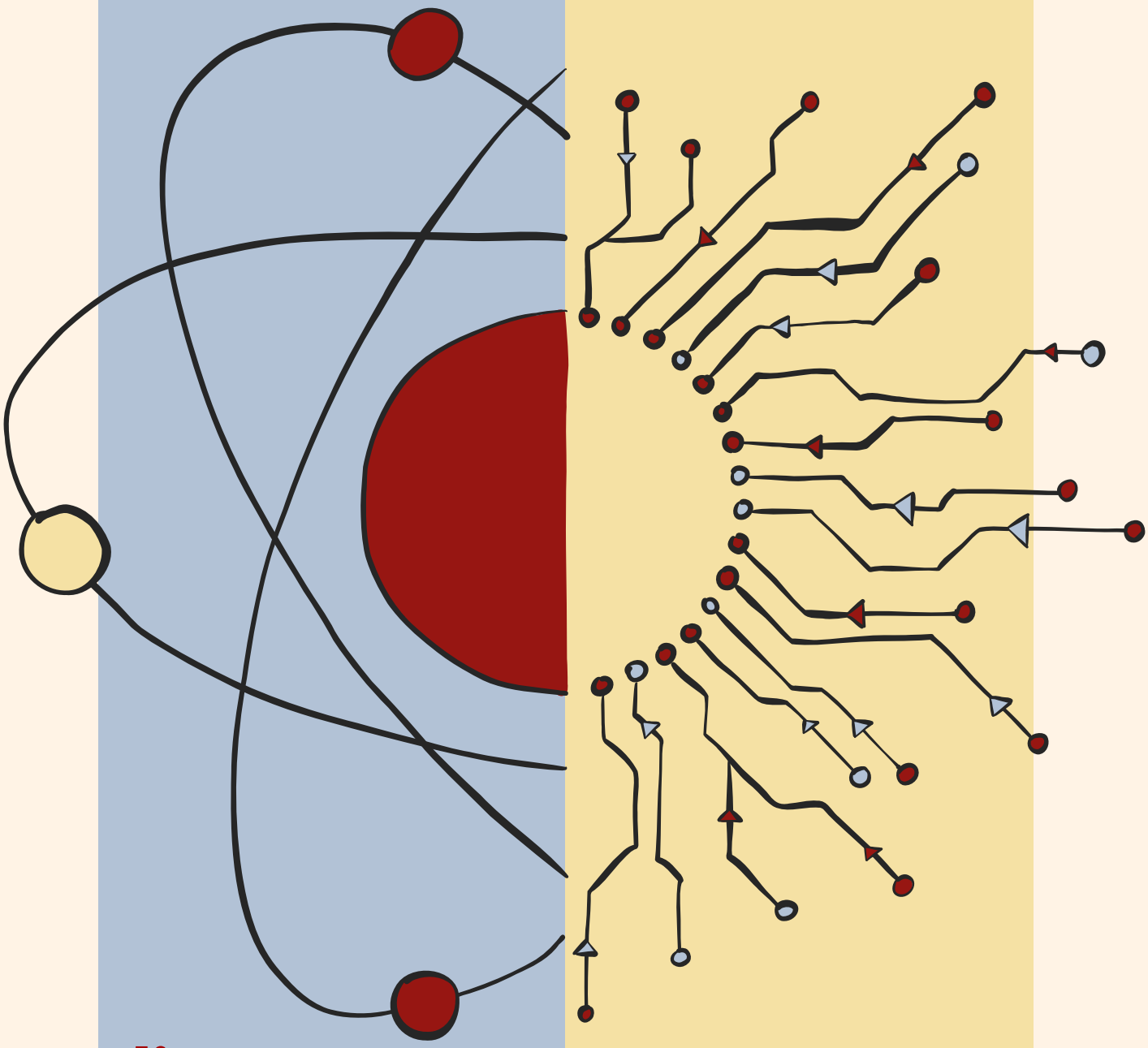
Quantum versus classical resources in computational complexity



Jordi Weggemans

Quantum versus classical resources in computational complexity

Jordi Rudo Weggemans

Quantum versus classical resources in computational complexity

ILLC Dissertation Series DS-2025-09

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

For further information about ILLC-publications, please contact

Institute for Logic, Language and Computation
Universiteit van Amsterdam
Science Park 107
1098 XG Amsterdam
phone: +31-20-525 6051
e-mail: illc@uva.nl
homepage: http://www.illc.uva.nl/

The research for/publication of this doctoral thesis received financial assistance from the Dutch Ministry of Economic Affairs and Climate Policy (EZK), as part of the Quantum Delta NL programme.

Cover design by Galina Pass.
Printed and bound by Ipskamp Printing.

ISBN: 978-94-6473-923-7

Quantum versus classical resources in computational complexity

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. P.P.C.C. Verbeek
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op donderdag 13 november 2025, te 16.00 uur

door Jordi Rudo Weggemans
geboren te Zwolle

Promotiecommissie

Promotor:	prof. dr. H.M. Buhrman	Universiteit van Amsterdam
Copromotor:	dr. F. Speelman	Universiteit van Amsterdam
Overige leden:	prof. dr. R.M. de Wolf	Universiteit van Amsterdam
	prof. dr. C.J.M. Schoutens	Universiteit van Amsterdam
	prof. dr. C. Schaffner	Universiteit van Amsterdam
	prof. dr. S. Gharibian	Paderborn University
	dr. J. Helsen	CWI

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

Publications and preprints

This dissertation is based on the following papers. For the papers where the authors are ordered alphabetically the co-authorship is shared equally. The dissertation's author is the main contributor of the papers where he is the first author.

[CFW22] Chris Cade, Marten Folkertsma, and Jordi Weggemans. Complexity of the guided local Hamiltonian problem: improved parameters and extension to excited states, 2022. arXiv: 2207.10097
Accepted (merged with [GHGM22]) talk at QIP 2022.

[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)*, volume 261, pages 32:1–32:19, 2023. arXiv: 2207.10250
This is a merged work which unifies [CFW22] and [GHGM22].

[WFC24] Jordi Weggemans, Marten Folkertsma, and Chris Cade. Guidable Local Hamiltonian Problems with Implications to Heuristic Ansatz State Preparation and the Quantum PCP Conjecture. In *19th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2024)*, volume 310, pages 10:1–10:24, 2024. arXiv: 2302.11578

[Weg24] Jordi Weggemans. Finding quantum partial assignments by search-to-decision reductions, 2024. arXiv: 2408.03986

[BHW25] Harry Buhrman, Jonas Helsen, and Jordi Weggemans. Quantum PCPs: on Adaptivity, Multiple Provers and Reductions to Local Hamiltonians. *Quantum*, 9:1791, July 2025. arXiv: 2403.04841
Accepted talk at TQC 2024.

[Weg25] Jordi Weggemans. Lower Bounds for Unitary Property Testing with Proofs and Advice. *Quantum*, 9:1717, April 2025. arXiv: 2401.07912

[BGW24] Harry Buhrman, François Le Gall, and Jordi Weggemans. Classical versus quantum queries in quantum PCPs with classical proofs, November 2024. arXiv: 2411.00946

[BBW25] Marcello Benedetti, Harry Buhrman, and Jordi Weggemans. Complement Sampling: Provable, Verifiable and NISQable Quantum Advantage in Sample Complexity, 2025. arXiv: 2502.08721

During the course of this PhD, the author has additionally coauthored the following articles that are not included in this dissertation.

[OOWK+21] L. A. B. Olde Olthof, J. R. Weggemans, G. Kimbell, J. W. A. Robinson, and X. Montiel. Tunable critical field in Rashba superconductor thin films. *Phys. Rev. B*, 103:L020504, Jan 2021. arXiv: 2009.14592

[WUR+22] Jordi Weggemans, Alexander Urech, Alexander Rausch, Robert Spreeuw, Richard Boucherie, Florian Schreck, Kareljan Schoutens, Jiří Minář, and Florian Speelman. Solving correlation clustering with QAOA and a Rydberg qudit system: a full-stack approach. *Quantum*, 6:687, April 2022. arXiv: 2106.11672

[CFNW23] Chris Cade, Marten Folkertsma, Ido Niesen, and Jordi Weggemans. Quantifying Grover speed-ups beyond asymptotic analysis. *Quantum*, 7:1133, October 2023. arXiv: 2203.04975
Accepted talk at QCTIP 2022.

[CFNW24] Chris Cade, Marten Folkertsma, Ido Niesen, and Jordi Weggemans. Quantum algorithms for community detection and their empirical run-times. *Quantum Information & Computation*, 24(5&6):0361–0410, 2024. arXiv: 2203.06208
Accepted talk at QCTIP 2022 (merged submission with [CFNW23]).

[BGLW24] Harry Buhrman, Dmitry Grinko, Philip Verduyn Lunel, and Jordi Weggemans. Permutation tests for quantum state identity, 2024. arXiv: 2405.09626
Accepted talk at TQC 2024, accepted talk at Young Quantum Information Scientists (YQIS 2024).

[BW24] Martijn Brehm and Jordi Weggemans. Assessing fault-tolerant quantum advantage for k-SAT with structure, 2024. arXiv: 2412.13274

Contents

Appendices

Acknowledgments

First, I would like to thank my promotor, Harry Buhrman, and my co-promotor, Florian Speelman, for taking me on as their student, for their guidance, and for the opportunities they provided throughout the years.

I want to extend my gratitude to Kareljan Schoutens, Jonas Helsen, Sevag Gharibian, Ronald de Wolf, and Chris Schaffner for being willing to serve on my committee. I would especially like to thank Ronald de Wolf, who, on numerous occasions over the years, gave me invaluable advice on how to navigate some of the challenges that can arise in academia.

I am also very grateful to Chris Cade and Iso Niessen, who, next to being collaborators, were always open to sharing questions, ideas, advice, and good conversations during the first half of my PhD. I also thank you for the brief part-time stay at Fermioniq, though I'm afraid that by that point I was already too absorbed in the abstractions of complexity theory to find the proper time to worry about the world in practice.

Scientifically, I would like to thank the co-authors I've had the pleasure of working with during my PhD: Marcello Benedetti, Martijn Brehm, Harry Buhrman, Chris Cade, Marten Folkertsma, Sevag Gharibian, Dmitry Grinko, Ryu Hayakawa, Jonas Helsen, François Le Gall, Tomoyuki Morimae, Ido Niesen, and Philip Verduyn Lunel. I also express my gratitude to Alex Grilo and Sevag Gharibian for hosting me for research visits in Paris and Paderborn, respectively.

CWI and QuSoft have been fantastic places to do research over the past few years. Thank you, Adam, Ailsa, Aldo, Álvaro, Ake, Akshay, Amira, Anna, Arjan, Bob, Carla, Chanelle, Chris, Chris, Crownie, Daan, Davi, Dmitry, Dyon, Doutzen, Farokh, Filippo, Florian, Fran, Francesco, Freek, Galina, Garazi, Giada, Gina, Harold, Harry, Ido, Jana, Jelena, Jeroen, John, Jonas, Jop, Joppe, Joran, Joris, Kareljan, Koen, Koen, Léo, Llorenç, Lorenzo, Luca, Ludovico, Lynn, Marten, Manideep, Maris, Maxim, Mehrdad, Mert, Michael, Niels, Nikhil, Peter, Philip, Poojith, Quinten, Randy, René, Ronald, Salvatore, Samuel, Sarah, Sebastian, Seenivasan, Stacey, Subha, Susanne, Tony, Vania, Victor, Yanlin, Yaroslav, and

Yvonne for all the 10 o'clock coffees, foosball games, post- and pre-conference trips, and random-topic lunch conversations.

Regarding this thesis, special thanks go to Galina for designing the cover, and to Lynn and Tom for being my paranymphs.

Finally, I want to thank my parents, family, and friends for all the things that aren't part of a thesis—without which this thesis would not exist. I have often felt as though I were living in two separate worlds at the same time, and your support and friendship in one often provided a welcome escape from the other.

Amsterdam Jordi Weggemans
July 2025

Chapter 1

Quantum versus classical resources in computational complexity

In the early 20th century, Hilbert advocated that mathematicians' primary aim should be to establish mathematics on a solid and provably consistent foundation of axioms, from which all mathematical truths could be deduced [Hil00]. With such a system in place, all mathematical truths could, at least in principle, be obtained using methods of first-order logic.

The above is a slight simplification of what is known as *Hilbert's program*, from which we can identify three distinct objectives. First, *consistency*: the set of axioms should be free of contradictions and provably so. Second, *completeness*: all mathematical truths should be derivable from these axioms. Third, *decidability*: there should be a well-defined procedure that can determine, within a finite time, whether any given mathematical statement can be proven from the set of axioms.

In 1931, it was shown that Hilbert's program could not be fully realised. Gödel's incompleteness theorems [Göd31] intuitively state that (i) in any consistent formal system that can express arithmetic, there are true mathematical statements that cannot be proven within the system, meaning such systems are inherently incomplete; and (ii) a consistent formal system cannot prove its own consistency, shattering the possibility of achieving the first two of Hilbert's objectives.

What about decidability? Of course, the existence of a well-defined procedure to decide the truth of a statement is only possible if the mathematical system is complete to begin with, so (ii) already implies that such a procedure cannot exist. However, at the time, the notion of an algorithm had not yet been precisely defined. This is what Turing accomplished in 1936 by introducing a formal notion of effective computability through the concept of the *Turing machine* [Tur36]. He proved that certain problems, such as the halting problem—determining, given a Turing machine M and an input x, whether M on input x will eventually halt or continue running forever—cannot be solved by any Turing machine, demonstrating the impossibility of a universal decision procedure within an algorithmic

framework.

Yet, Turing was not alone (nor the first) in doing so. Around the same time he introduced Turing machines and the halting problem, Church independently proved the existence of undecidable problems using λ-calculus [Chu36b, Chu36a]. Meanwhile, work by Gödel and Kleene on recursive functions provided yet another formalisation of computation [Göd31, Kle36]. The works of Church and Turing showed that all three models—Turing machines, λ-calculus, and recursive functions—are equivalent in terms of computational power: a function is computable in one model if and only if it is computable in the others. This forms the basis of the *Church–Turing thesis*, which asserts that any function that can be computed by an algorithm can be computed by a Turing machine. As far as is known, no physical or mathematical evidence contradicts the thesis, and it is therefore widely accepted as holding across all models of computation.

Once one accepts this thesis, it is natural to hypothesise further: the *efficient Church–Turing thesis* extends the notion of computational equivalence to computational *efficiency* [vEB91].[1] It posits that any "realistic" model of computation can be efficiently simulated by a (randomised) Turing machine, with at most a polynomial overhead in time and space. To discuss its plausibility, we need to make a leap to the field of physics.

Only a few decades before Hilbert's address, the physics community had been undergoing its own scientific revolutions. In 1894, Michelson expressed the sentiment among the physics community as follows [Mic96]:

> While it is never safe to affirm that the future of Physical Science has no marvels in store even more astonishing than those of the past, it seems probable that most of the grand underlying principles have been firmly established [...] the future truths of physical science are to be looked for in the sixth place of decimals.

Michelson's words still echoing, the next decades would produce the theory of relativity—showing that space and time should really be *space-time*; and quantum theory—telling us that mechanics should really be *quantum mechanics*. It became clear that at the very large and very small scales, the world behaved radically differently from what everyday intuition, based on classical mechanics, would predict.

However, quantum theory comes with its challenges when one attempts to simulate it to make predictions about our world. In classical mechanics, one typically assumes knowledge of a system's individual components, such as initial positions, masses, energies, charges, and initial (angular) momenta, along with any external forces. These parameters are tracked and updated as the system

[1] Also referred to as the strong Church–Turing thesis or the feasibility thesis, though it does not originate from Church or Turing.

evolves according to the laws of classical mechanics. In the Turing machine model, this generally poses no problem when it comes to *storage*: all relevant information can be efficiently stored (up to a certain precision) and updated on a large data storage tape. Quantum mechanics, however, reveals that a system's information is actually "stored" in a wave function, which requires an exponentially growing vector description as the number of components increases. As a result, even for moderately sized quantum systems, tracking the system's wave function may require Turing machines to use computational resources (e.g., time, memory, etc.) that exceed even the scale of the universe!

This "memory problem" was one of the prime motivations for Feynman to propose in 1982 that we should use quantum-mechanical computing devices, more conveniently called *quantum computers*, to solve quantum-mechanical computational problems [Fey82]. In fact, in 1980 Paul Benioff already proposed the idea of a quantum mechanical model of a Turing machine, which could perform computations using quantum mechanical principles [Ben80]. Inspired by this, Deutsch generalised the Turing machine to the *quantum Turing machine* and proved that as far as computability is concerned, the Church–Turing thesis still holds its ground: any function that can be computed by a quantum Turing machine can be computed by a Turing machine (and vice versa) [Deu85]. But what about *efficiency*?

In a spectacular result, Shor showed in 1994 that the computational problem of integer factorisation can be efficiently solved on a quantum computer [Sho94]. This result implies that at least one of the following three statements must be true: (i) the efficient Church–Turing thesis is false; (ii) integer factorisation can be solved efficiently classically; (iii) quantum mechanics is incorrect. Arguably the "least unlikely" of the three, it is now widely accepted that the efficient Church–Turing thesis is likely incorrect and should be extended to refer to quantum Turing machines instead. Hence, while quantum computers do not allow us to compute more functions than classical computers, they can compute certain functions more efficiently [Mon16, HM17, DMB$^+$23].

Since quantum information and computation generalise their classical counterparts, not much seems lost by making all computational resources quantum from a purely theoretical point of view, as long as we are willing to incur an asymptotically small overhead to account for reversibility [Ben89] and extensive error-correction to deal with the fragility of quantum information [Sho96]. So, should we just replace all computational resources with "something quantum"?

Though asymptotically small, more and more evidence shows that in practice these overheads can be astronomical [CKM19, BMN$^+$21]. This means that, for example, when quantum algorithms offer only a small (polynomial) advantage, the required problem sizes may be so large that these speedups become meaningful only for those willing to perform computations over months, years, or even millennia [BMN$^+$21, CFNW23, CFNW24, DGLM24, BW24]. Hence, from a prac-

tical standpoint, quantum computation and information should not be viewed as a universal tool applicable to all problems, but rather as a highly specialized instrument: potentially powerful, yet not suitable for every task.

At the heart of this *practical* consideration lies a fundamentally *theoretical* question: in what contexts do quantum resources truly shine—offering dramatic asymptotic advantages that dwarf even the most pessimistic estimates of the overheads—and for what tasks might we as well leave it to classical resources, despite marginal asymptotic quantum advantages? The central theme of this thesis can therefore be summarised by the following question:

> *What is the interplay between quantum and classical resources in computation, and for which tasks are quantum resources provably (much) better?*

This dissertation will explore this question in the context of the theory of *computational complexity*, which studies the resources required to solve problems within different computational models. It classifies problems based on their inherent difficulty, considering factors such as time, space, and access to additional resources like provers or oracles. In computational complexity theory, one can rigorously compare quantum versus classical algorithms, classical versus quantum proofs, classical versus quantum advice, classical versus quantum query access to oracles, classical versus quantum samples, classical versus quantum reductions, and so on.

In this way, one can show that using quantum resources in computation can lead to provable superpolynomial advantages in communication complexity [BCW98, Raz99, BCWdW01], sample complexity [BJ95, GKZ19], query complexity [Sim97, CCD$^+$03, AA15], and space complexity [GKK$^+$07, KPV24]. In each of these settings, quantum models outperform classical ones by exploiting more powerful access models and/or by leveraging inherently quantum phenomena like entanglement or interference.

Moreover, it is also possible to study the *interplay* between quantum and classical resources by combining classical and quantum resources in a single computational setting—such as quantum verifiers with access to classical proofs or classical proof systems with access to quantum reductions. This way, we can gain insights into what aspects of computation truly benefit from quantum over classical. Computational complexity theory also allows us to study problems with access models that only make sense in the context of quantum computing. And perhaps most importantly, it allows us to study our main question in a precise manner, making definite statements that hold independently of hardware considerations.[2]

[2] And, importantly for quantum, the hardware's current limitations.

Beyond the practical motivation, we will see that this angle occasionally gives insights into the peculiar nature of quantum physics. Sometimes quantum resources cannot do much better than classical ones, hindered by the fact that some simple classical concepts—such as copying, deleting and reading out information—have no perfect quantum analogue. Yet, as is well known[3] (and as we will see in this dissertation), these limitations are sometimes precisely the reason that quantum *can* do something spectacular.

A vast body of research already addresses our central question in many different ways. Yet, even 40 years since David Deutsch first formalised the quantum Turing machine [Deu85], there still seems to be room to add new insights into the power and limitations of, as well as the connections between, quantum and classical resources in computation. Specifically, we will explore our central question in three main directions, each of which forms its own part:

I. Low-energy states and their descriptions.

II. Quantum probabilistically checkable proof systems.

III. Unitary query and sample complexity.

For the rest of this introductory chapter, we will introduce each of these directions, explaining their relevance in a broader context, how they connect to our central question, and what the main contributions of this dissertation are. This will be done at a fairly informal level, with references therein to the chapters where the topics are treated at a formal level.

1.1 Low-energy states and their descriptions

Part I of this dissertation follows Feynman's original intuition that "quantum problems should be solved by quantum computing devices". Specifically, we will focus on computational problems related to low-energy states of quantum-mechanical systems, which play an important role in many applications. A quantum-mechanical system is described via its energy operator, the so-called *Hamiltonian* H, which models the interactions between the different components of the system. For example, in a spin system, the Hamiltonian can model exchange interactions, which arise due to the Pauli exclusion principle and the Coulomb interaction between electrons; in a molecule, this could be the individual kinetic energies and interactions among atomic nuclei and electrons. For most systems, it is assumed that these interactions are *local*: this means that H can

[3]For example, no-cloning is one of the most important assets in quantum cryptography.

be decomposed as a sum of Hermitian operators, i.e.,

$$H = \sum_{i \in [m]} H_i,$$

where each H_i acts only non-trivially on at most k subsystems. Locality in this sense is not to be confused with *geometrical locality*, which describes to what extent individual systems can be put on some kind of hypergraph structure, such that systems far apart do not (or only weakly) interact.

One of the most basic tasks in determining properties of a quantum system is to compute its *ground-state energy*, i.e., the minimum eigenvalue of H, which we will denote by λ_0 in this introductory chapter. Information about the ground state provides insights into reaction rates, molecular geometry, binding energies, band structures, and more—quantities that are, for instance, important in the design of drugs [CRO$^+$19] and materials [IW18].

Computing the ground state energy of a k-local Hamiltonian can be formalized as the following computational problem: you are given as an input classical descriptions of the local terms which together form the Hamiltonian H, and two parameters a, b such that $b - a = \delta$, where δ is the so-called *promise gap*. The task is now to determine whether $\lambda_0 \leq a$ or $\lambda_0 \geq b$, promised that either one is the case. It is known that this problem is QMA-complete when $k = 2$ and $\delta = 1/\mathrm{poly}(n)$ [KKR06], where QMA is the quantum analogue of NP. This remains true for more restricted families of Hamiltonians, for instance nearest-neighbour Hamiltonians on a two-dimensional lattice of qubits [OT08], as well as many other physically realistic models [SV09, WMN10, CMP18, OIWF22]. Moreover, this hardness persists even in the excited state setting, where one is interested in computing energies above the ground state level [JGL10]. Assuming that $\mathsf{QMA} \neq \mathsf{BQP}$, where BQP captures all problems quantum computers can solve using polynomially-bounded space and time, this suggests that estimating energies of low-energy eigenstates of many physical Hamiltonians up to inverse polynomial precision is a hard problem even for quantum computers.

Yet, it is widely believed that Feynman's intuition—computational quantum problems should be solved with quantum computational devices—should still apply to some extent and that, despite these hardness results, quantum computers will have a profound impact on performing computations for quantum systems [Aar09, BBMC20]. The central question in Part I is to investigate in what setting in the context of ground and excited state energy estimation this belief can indeed be mathematically justified:[4]

(Q1) *Is there evidence for the existence of a well-defined, practically motivated setting in ground- or excited-state energy estimation where quantum computing offers a superpolynomial advantage over classical computing?*

[4]The use of the word "setting" here is important, as showing so unconditionally would prove $\mathsf{P} \neq \mathsf{PSPACE}$, resolving one of the major open problems in computational complexity theory.

1.1.1 Ansätze

To set the stage for our considered approach to ground state energy estimation (discussed next), we first consider the notion of *ansätze* from the field of physics. In the context of ground states, an ansatz usually refers to an "educated guess" for a class of states that

(a) has some desirable properties, such as an efficient classical description from which useful information can be extracted;

(b) forms a good approximation to the ground state (or another state of interest).

Some well-known examples of ansätze are tensor network states (e.g., matrix product states, projected entangled pair states), product states, and variational quantum circuits. For some families of Hamiltonians, one can even prove that the ground states come from such a restricted family of states. For example, it is known that the ground state of any 1D gapped local Hamiltonian can be well approximated by matrix-product states [Has07].

However, both criteria (a) and (b) above are not well-defined in a mathematical sense. Moreover, what qualifies as "desirable properties" can vary depending on the application. For instance, in the context of ground state energy estimation, you might at the very least want to be able to (potentially heuristically) compute the energy of the state represented by the ansatz. If you have access to a quantum computer, you may wish to implement the state as an actual quantum state on the computer to further probe its properties.

In Chapter 3, we formalise several distinct types of desirable properties within what we refer to as *access models*. Here "access" refers to the ways in which information can be extracted from a given description of a state in an ansatz. This is motivated by the work of Gharibian and Le Gall, who defined a class of states that capture all ansätze allowing the following access model: (i) sampling basis states according to the Born rule, and (ii) querying the amplitude of an individual basis state [GL22].

We then compare this access model with two alternative ones: one requiring efficient classical (approximate) computation of local observables, and another in which the state can be efficiently prepared as a quantum state. We also study some properties of these different access models and define classes of associated states—those that satisfy the criteria of the access model—which will then "guide" us in solving our ground state energy estimation problem in the next section.

1.1.2 Computing ground state energies with guiding states

In an attempt to bypass worst-case QMA-hardness results, the quantum computing community usually proposes a two-step procedure [AL99, AGDLHG05]:

1. A classical (or quantum) heuristic algorithm is applied to obtain a so-called *guiding state* $|\psi\rangle$, which is hoped to have "good" fidelity with a ground space.

2. The guiding state $|\psi\rangle$ is used in Quantum Phase Estimation (QPE) [Kit95] (or an alternative method [LT20b, LT20a]) to efficiently compute the corresponding ground state energy.

The motivation behind this two-step procedure is that it seems to capture the best of both worlds: for Step 1, classical heuristics have had decades of development, are fast to implement, and can be optimized more easily compared to quantum heuristics (e.g., variational quantum eigensolvers, see [CAB$^+$21] for a survey). Additionally, for many of these guiding states, it is known that it is possible to prepare the state on a quantum computer (as further discussed in Chapter 3). Step 2, on the other hand, seems to capture a unique strength of quantum computers: the ability to resolve an eigenvalue within additive $1/\mathrm{poly}(n)$ precision of a (sparse) Hermitian matrix given just an *approximation* $|\psi\rangle$ to the corresponding eigenvector. In the context of ground-state energy estimation, the approximation via a guiding state does not need to be good in terms of its energy: the above procedure runs in polynomial time as long as the fidelity with the ground space is $1/\mathrm{poly}(n)$.

In [GL22], Gharibian and Le Gall initiated the formal study of the complexity of the second step outlined above. Specifically, they introduced the *Guided k-local Hamiltonian problem (k-GLH)*, which is roughly stated as follows: given a k-local Hamiltonian H, properly normalized so that $\|H\| \leq 1$, an appropriate representation of a guiding state $|\psi\rangle$ with ζ-fidelity with the ground space of H, and real thresholds $b > a$, decide if the ground energy λ_0 satisfies $\lambda_0 \leq a$ or $\geq b$. For the guiding state representation, the definition can be modified to accommodate different types of access models (see Section 1.1.1).

Gharibian and Le Gall then proved that 6-GLH is BQP-hard for *inverse polynomial* precision and up to a maximum allowed fidelity, i.e., $b - a \geq 1/\mathrm{poly}(n)$, and $\zeta = 1/2 - 1/\mathrm{poly}(n)$. However, for any constant k and constant fidelity ζ, k-GLH can be efficiently solved *classically* within *constant* precision, i.e., for $b - a \in \Theta(1)$. Here, constant precision refers to precision that is constant with respect to a renormalised Hamiltonian with operator norm at most 1. Since k-GLH is in BQP for any $b - a = \Omega(1/\mathrm{poly}(n))$ and $\zeta = \Omega(1/\mathrm{poly}(n))$, the BQP-hardness in the inverse-polynomial precision regime provides a *provable* superpolynomial quantum advantage in the ground state energy estimation context (under the assumption that $\mathsf{BQP} \neq \mathsf{BPP}$).

The result in [GL22] partly addresses our question, yet many aspects remain open. For instance, the Hamiltonian for which BQP-hardness is proven in [GL22] is not very physical: it is 6-local and lacks geometrical locality. Thus, it is desirable to extend the result to more physically motivated Hamiltonians if we aim to capture the "practically motivated"-part of the question. Moreover, beyond

ground states, excited states also play a crucial role in many applications. For example, many photophysical and photochemical processes, such as energy transfer [NOAO$^+$18], bond dissociation [ZLTN20], light emission, and non-adiabatic dynamics [NWB$^+$20, WLZ23], revolve around electronically excited states. Can the hardness result be extended to this setting as well? Furthermore, what happens for larger fidelities? Lastly, what is the complexity of the problem if we also consider the complexity of Step 1? In Chapter 4, we address all of these questions, which can be informally summarised by the result:

(R1) There exist 2-local, physically motivated Hamiltonians for which estimating the ground- (or excited-) state energy, given access to a guiding state with fidelity at least $1 - 1/\mathrm{poly}(n)$, is BQP-complete. If the guiding state is only promised to exist, the problem becomes QCMA-complete.

Here, the class QCMA is analogous to QMA, but with classical instead of quantum proofs: the verifier is still a quantum computer, but the proof (in this case, a classical description of the guiding state) must be efficiently describable using a classical bit string.

Under the (widely believed) assumption that BQP $\neq$ BPP, the above result demonstrates the existence of a well-defined, practically motivated setting for ground state energy estimation in which quantum computers are superpolynomially more efficient than the best classical algorithms. We will also show that, from a complexity-theoretic perspective, no power is lost by restricting the guiding states to a subset of all states for which the expectation values of local observables can be computed classically, provided these states can still be prepared on a quantum computer. This implies that the guiding state assumption is significantly weaker than imposing structure on the ground state itself: requiring structure only on the guiding state (the former) would imply QMA = QCMA, whereas requiring it on the ground state (the latter) would imply QMA = NP. Finally, this provides complexity-theoretic justification for using classical heuristics rather than quantum heuristics in Step 1 of the proposed two-step procedure.

1.1.3 Finding ground state descriptions

Ground-state *preparation*, as opposed to ground-state energy *estimation*, is crucial for applications where one seeks to probe properties of the ground state beyond its energy value, such as when studying its behaviour under time evolution [Llo96] or when one wants to compute two-point correlation functions (e.g., to characterise quantum chaos [GHST20]). So far, we have only discussed estimating the ground-state energy and potential ground-state ansätze, without considering the complexity of preparing an arbitrary ground state itself. Trivially, the ability to prepare the ground state enables one to estimate its energy, but what about the converse?

For most classical problems, the other direction of the reduction is straightforward, as the problem of *finding* a solution is often reducible to deciding whether there *exists* a solution. Following Irani, Natarajan, Nirkhe, Rao and Yuen [INN+22], we study this in a particular type of *oracle* setting: we assume black-box access to a machine that solves some type of decision problem complete for a class of interest, which can be done at unit cost.[5] For example, given a constraint satisfaction problem (CSP) on n bits, an NP oracle can decide whether there exists an assignment that satisfies the constraints. If so, one can make adaptive queries to the oracle—asking whether the constraints remain satisfiable under a given partial assignment—thereby learning one bit of information about some satisfying assignment x^* with each query.

As was pointed out by Irani, Natarajan, Nirkhe, Rao and Yuen, "quantum solutions" (i.e., QMA witnesses) seem to be fundamentally different from classical solutions (i.e., NP witnesses) because

(i) the description-size complexity of a quantum state on n qubits is generally exponential in n;

(ii) there does not appear to be a natural way of conditioning a quantum state on a partial assignment.

This intuitive obstruction can be formalised in the relativised world, as [INN+22] shows that relative to a quantum oracle QMA fails to have search-to-decision reductions. This result is in contrast with some related classes where the witnesses are *classical*: for instance, NP, MA, and QCMA all have search-to-decision reductions relative to all oracles [INN+22].

Suppose we believe that (i) is indeed a truly fundamental obstacle. What would then be the second-best thing one could hope for? Going back to the local Hamiltonian problem, we observe that the full quantum state in fact contains more information than is needed; since the Hamiltonian is local, it suffices to have sufficiently good approximations of all k-local *density matrices* of a low-energy state. Constant-locality density matrices do not suffer from point (i) above, as any n-qubit state only has a polynomial number of them and each has a polynomially-sized description (for inverse exponential accuracy). However, it is well-known that it is again QMA-complete to check if all density matrices are consistent with a global quantum state [Liu06, BG22]. Yet, this leaves open the possibility of finding approximate descriptions of local density matrices of a local

[5] A subtle but important clarification is needed here. In this setting, we assume access to a QMA oracle capable of solving any promise problem in QMA, not just the specific decision problem corresponding to the search task of interest. This is equivalent to assuming access to an oracle for a single QMA-complete problem A, since any problem B in QMA is polynomial-time reducible to A. However, if A is only assumed to be contained in QMA, the corresponding oracle might be too weak, as it would not capture the full power of QMA. In fact, in this case, the result would not even hold for NP, since it is known that there exist problems in NP that do not admit such search-to-decision reductions unless EE = NEE [BG94].

Hamiltonian when given access to a QMA oracle. In Chapter 5, we will prove that this indeed is possible for all local Hamiltonians, and even more generally for any problem in QMA.

(R2) For any problem in QMA, there exists a classical polynomial-time algorithm with access to a QMA oracle that outputs approximations of the density matrices of a near-optimal quantum witness, for any desired constant locality and inverse polynomial error.

Our result allows one to store a classical fingerprint of a low-energy state that can be used to compute expectation values of local observables, like two-point correlation functions, indefinitely. Moreover, our result suggests that the lack of the "bottom-up" property—global quantum states generally cannot be reconstructed from their marginals—lies at the core of why search-to-decision does not seem to work for quantum proofs.

1.2 Quantum probabilistically checkable proof systems

In Part II we will look at proof systems. The aforementioned classes NP, QCMA and QMA are all examples of such proof systems, and share the property that they all consider the setting where an all-powerful prover is allowed to send an untrusted proof to a computationally-bounded verifier. In the 80s, researchers tried to scale up the power of such systems by allowing interaction between the prover and verifier (IP), which could even be further strengthened by allowing multiple uninteracting provers (MIP), leading up to the celebrated IP = PSPACE [Sha92] and MIP = NEXP [BFL90] theorems.

Naturally, people asked what would happen if one were to *scale down* the power of proof systems. A probabilistically checkable proof (PCP) system consists of a polynomial-time verifier that uses $r(n)$ random coins and makes at most $q(n)$ queries to a proof provided by the prover. For a proof written in binary, a query corresponds to reading out a single bit at a specified location in the proof. The PCP theorem [ALM+98, AS98] states that all problems in NP can be decided, with a constant probability of error, by only using a logarithmic number of coin flips and only 3(!) queries to the proof [H$\mathring{\text{o}}$1].

The PCP theorem is also important from a practical point of view in approximation theory: it directly implies that it is NP-hard to decide whether an instance of a CSP is either completely satisfiable or no more than a constant fraction of its constraints can be satisfied. This is usually referred to as the *hardness of approximation* formulation of the PCP theorem. Later, it was shown that it is possible to prove the PCP theorem by reducing a CSP *directly* to another CSP with the above property. This transformation, due to Dinur [Din07], is usually

referred to as *gap amplification*, referring to the increase in the difference (the gap) in the fraction of constraints which can be satisfied in both the YES- and NO-instances.

Naturally, quantum complexity theorists have asked whether an analogous theorem exists in a quantum setting, proposing proof-checking and hardness of approximation versions of a quantum PCP conjecture. The hardness of approximation formulation states that the local Hamiltonian problem with constant promise gap, relative to the operator norm of the Hamiltonian, is QMA-hard. This formulation of the quantum PCP conjecture has been the predominant focus of the quantum PCP literature, and progress has been made in giving evidence both in favour and against the conjecture. Amongst the positive are the NLTS theorem and its cousins, excluding a large set of potential NP witnesses [ABN23, CCNN23b, CCNN23a, HATH24, AGK24]. Evidence against (assuming $\mathsf{NP} \neq \mathsf{QMA}$) are results showing the "quantum hardness" of the local Hamiltonian problem vanishes in the constant precision regime when one imposes some extra structure on the problem (e.g., constraining the interaction graph or ground space structure), whilst the same problem remains (quantumly) hard when the promise gap is inverse polynomial [BBT09, BH13b, ABG19, GL22, CFG$^+$23, WFC24].

We summarise our main direction as the following questions:

(Q2) *Can quantum proofs be more efficiently verified? What about the verification of classical proofs by a quantum verifier?*

1.2.1 Local Hamiltonians versus proof-checking

The proof-checking formulation of the quantum PCP conjecture, which states that one can solve any promise problem in QMA by using a quantum verifier which only accesses a constant number of qubits from a quantum proof, has received considerably less attention than the local Hamiltonian formulation. A reason for this is that both conjectures are known to be equivalent under *quantum* reductions. This was already observed in the first work proposing a quantum PCP [AALV09], although it was not formally written down. Perhaps that is why, even after more than two decades since the question of whether a quantum PCP exists was first posed [AN02], many basic questions regarding the proof-checking formulation have not been addressed. For instance, as already raised in [AALV09], does the choice of the distribution over which the proof qubits are selected matter? Is adaptive access to the proof more powerful than non-adaptive access in the constant-query setting, or do they have the same power, as is the case classically? What about having multiple provers, or classical proofs?

To address previously posed questions regarding quantum PCPs, in Chapter 6 we revisit the definition of the complexity class related to quantum PCPs. We will study a general notion of quantum PCP verifiers, which include the ability to

make adaptive queries to multiple untangled quantum proofs. Having this new definition of a quantum PCP at hand, we study its connection to the hardness-of-approximation formulation of the local Hamiltonian problem via a quantum reduction.[6] As it turns out, our definition and quantum reduction to a constant promise gap local Hamiltonian turn out to be powerful tools in proving several properties of quantum PCPs.

(R3) For any problem that is decided using an adaptive quantum PCP with multiple unentangled provers, there exists a quantum reduction to a local Hamiltonian problem with a constant promise gap. This can be used to show that adaptive proof access offers no additional power over non-adaptive access in the constant-query regime, and that if the multiple-unentangled-prover version of QMA (i.e., QMA(2)) admits a constant-query quantum PCP, then QMA(2) = QMA.

In Chapter 6 we give some additional results that follow from our reduction as well, like proving hardness for an average-energy per site formulation of the quantum PCP conjecture and a connection to the class QCMA.

1.2.2 Quantum PCPs with classical proofs

A major open question regarding quantum proof systems is what the "right" quantum generalisation of NP [Gha24] is. In particular, the QCMA versus QMA question asks whether it suffices to use classical proofs or whether quantum proofs are strictly more powerful. An argument for why they might be different, formalised through a quantum oracle separation by Aaronson and Kuperberg [AK07], is that given a proof of size m qubits, there exist a doubly exponential (in m) number of quantum states with pairwise small overlap; since a polynomial-sized classical string can only describe an exponential number of states, this counting argument shows that most states cannot have an efficient classical description. However, this argument already fails for the simple reason that, assuming a fixed universal gate set, for input size n all possible QMA-verification circuits themselves must be specified using poly(n) bits, which means that there are only an exponential number of possible verifiers. Since for every YES-instance a single quantum proof suffices, descriptions of the verifiers themselves already provide an efficient description of any quantum proof you would ever need.[7] Hence, the space of all relevant quantum proofs occupies only a very tiny fraction of the entire possible Hilbert space. A more physical intuition of

[6] As briefly mentioned earlier, while the existence of reductions between the two formulations is widely known in the community, they have never been fully written down except in the works of [Gri18] and [HATH24], both of which consider restricted formulations of quantum PCPs.

[7] Or, to put it in a local Hamiltonian formulation, the description "ground state of Hamiltonian H" also captures all possible relevant quantum proofs.

why QMA could be equal to QCMA comes from the local Hamiltonian problem: since QMA-hardness even holds for 2-local Hamiltonians defined on a lattice, one only has to care about geometrically local 2-local density matrices, which means that one does not have to worry about longer range entanglement [AN02].

To gain further intuition into how quantum verifiers can process quantum proofs as compared to classical proofs, it makes sense to study this question also in a PCP setting. In Chapter 7, we study quantum PCPs with classical proofs, analogous to how QCMA differs from QMA. In defining the corresponding complexity class, one is immediately faced with a choice: one can assume that queries to the proof are *classical*, which captures the locality aspect of PCPs; or the queries are allowed to be *quantum*, which is more faithful to the query aspect usually considered in a quantum setting. In the standard quantum query model, one assumes access to a bit string y of length 2^n via an $(n+1)$-qubit unitary operator U_y, defined by its action on computational basis states as $U_y |i\rangle |a\rangle = |i\rangle |a \oplus y_i\rangle$, with $a \in \{0,1\}$. This means that in a single quantum query, all entries of y can be accessed in superposition—a feature known to lead to superpolynomial advantages over classical queries for some computational tasks [BV93, Sim97, AA15].

In Chapter 7, we will study quantum-classical PCPs in both query models, arriving at some surprising results which shed light on the interplay between quantum verifiers and classical proofs.

(R4) Quantum-query quantum-classical PCPs with an inverse-polynomial promise gap can be simulated by classical-query quantum-classical PCPs with a constant promise gap, making only three classical queries. Moreover, the corresponding class of promise problems is contained in $\mathsf{BQ} \cdot \mathsf{NP}$, the class of all promise problems that admit a quantum reduction to CSPs. For a (poly-)logarithmic number of queries, the complexity classes corresponding to the two access models can be separated relative to an oracle.

As it is unlikely that $\mathsf{QCMA} \subseteq \mathsf{BQ} \cdot \mathsf{NP}$ (which we support with oracular evidence), this result gives strong evidence there exists no constant query quantum-classical PCP for QCMA, even though we can amplify the promise gap from inverse polynomial to constant.[8] Intuitively, our result is in line with the intuition that classical proofs in a quantum setting generally should be viewed as "uncompiled", and generally are only useful to a quantum verifier if they describe a quantum circuit which can be used to prepare a quantum proof.

[8] This is the much sought-after type of amplification needed to prove the quantum PCP theorem when quantum proofs are considered. Interestingly, the fact that we can show it for a quantum-classical PCP provides strong evidence *against* the existence of a quantum-classical PCP (in our setup) for QCMA.

1.3 Unitary query and sample-to-sample complexity

In our final part, we shift our focus to other measures of computational complexity. Up to this point, there was usually a restriction on the resources in terms of time, space, etc., and we looked at what types of problems could be solved in a complexity class which captures these restrictions. Query and sample complexity are two different notions of complexity, where there will still be restrictions on the allowed model of computation, but we will from that point on only care about how efficient our model is in terms of using some "fundamental object" needed to solve the task at hand (e.g., the number of queries to a function or number of needed samples). Our questions in Part III will be as follows:

(Q3) *Are there other complexity measures for which we can show a quantum advantage over classical computation? Moreover, are there complexity measures that are simply inherently quantum?*

1.3.1 Unitary query complexity

An inherently quantum query complexity model is that of *unitary query complexity*, where one assumes black-box (controlled) access to an n-qubit unitary U or its inverse. In *unitary property testing*, U itself is considered to be the input, and one wants to test whether U satisfies some property or is far away from having this property, minimising the number of queries one has to make to U. Whilst standard query complexity is very useful in obtaining insights into the differences in computational power between different classes of computation, classical or quantum, unitary query complexity gives a potentially useful way to compare inherently quantum classes. These problems, first studied by Wang [Wan11], got considerably more attention recently [SY23, CNY03, WZ23].

Query complexities can vastly differ among different computational models. For example, the search problem, which is to decide whether a string of length N is either the all-zeros string or has at least one entry with a "1", is known to have classical query complexity of $\Theta(N)$ and quantum query complexity of $\Theta(\sqrt{N})$ [Gro96, BBBV97]. However, with the aid of a proof by an untrusted prover, the query complexity of the search problem becomes 1 in both cases, as the prover can provide the location of the entry to be checked by the verifier. A similar result holds for a unitary property testing analogue of search as introduced by Aaronson and Kuperberg [AK07], where one has to decide whether a given black-box unitary U applies either the identity operation $\mathbb{I}$ or the reflection $\mathbb{I} - 2\,|\psi\rangle\langle\psi|$ for some unknown N-dimensional quantum state $|\psi\rangle$. This problem has in general a quantum query complexity of $\Theta(\sqrt{N})$ but can again be solved by just a single

query if a *quantum state* is provided by an untrusted prover as an extra input.[9] For many other unitary property testing problems, it is unclear whether quantum (or classical) proofs and/or trusted advice states might help in solving these tasks.

In Chapter 8, we will study the connection between unitary query complexity and channel discrimination, which is a well-studied problem in quantum information theory.

(R5) Problems in unitary query complexity can be reduced to unitary channel discrimination, giving information-theoretic query lower bounds which hold even in the presence of quantum proofs and advice.

In particular, our technique is useful for investigating the difficulty of quantum tasks involving unitary queries that demand high precision. It is also extremely *simple* to apply—for example, we prove an optimal lower bound for the quantum phase estimation problem in just 7 lines, and our result even holds in a stronger setting than previous proofs [Bes05, WZ23]. We will also discuss the implications for the multiple unentangled prover variant of QMA (i.e., $\mathsf{QMA}(2)$), whose computational power has long been a central open question in quantum complexity theory.

1.3.2 Sample-to-sample problems

Another complexity measure, frequently encountered in a learning context, is that of *sample complexity*. At a basic level, sample complexity typically differs from query complexity in that the problem solver only controls how the given objects are processed, not which objects are provided. While sample complexity can be studied in a fully quantum setting—for example, when learning an unknown quantum state given (a tensor product of) copies of the state [AA24]—our focus will be on the model of quantum samples which follows the one used in quantum learning theory [BJ95, AdW17]. Here, one is given sample access to elements from a finite set S according to a probability distribution $D : S \subseteq \{0,1\}^n \to [0,1]$, and quantum samples correspond to having access to copies of the state

$$|S_D\rangle = \sum_{x \in S} \sqrt{D(x)}\,|x\rangle\,.$$

Measuring $|S_D\rangle$ in the computational basis is then equivalent to classically sampling from S according to D.

In Chapter 9, we will study sample complexity in what we call a *sample-to-sample* setting: here, you are given samples S according to a distribution D, and the goal is to output a *single* (classical) sample from a set S' (potentially according to a distribution D'). This differs from most problems that use sample

[9] This relies on having controlled access to U. Applying $|0\rangle\langle 0| \otimes \mathbb{I} + |1\rangle\langle 1| \otimes U$ to $|+\rangle\,|\psi\rangle$ leaves the first register in $|+\rangle$ when $U = \mathbb{I}$ and transforms it to $|-\rangle$ when $U = \mathbb{I} - 2\,|\psi\rangle\langle\psi|$.

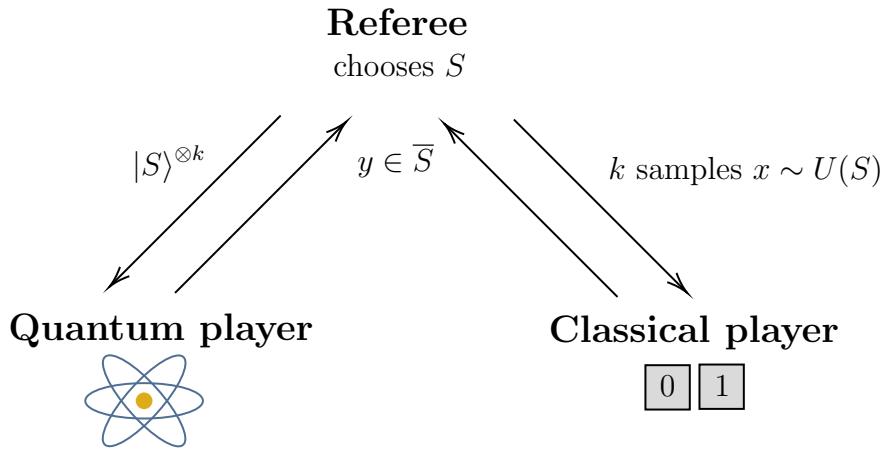


Figure 1.1: Comparing a single round of complement sampling with k input samples for quantum and classical players, given a problem instance S with cardinality K. A quantum player gets access to k copies of the state $|S\rangle$, and a classical player gets access to k elements sampled according to the uniform distribution over S (denoted as $U(S)$). Note that giving the classical player access to measurements of $|S\rangle$ in the computational basis would yield the same input model.

complexity as a measure, in the sense that we are not trying to learn S' to make it reproducible; rather, our focus is that, given some samples from S, we output a *single* correct answer from S' (with high probability). A trivial example is the case where S and S' are constructed by sampling strings of length n, placing y in S and y with its last bit flipped in S' until all strings have been assigned, with D and D' both being uniform distributions. Clearly, a single sample from S suffices to produce a sample from S' (even according to D'). However, the task becomes significantly harder when, instead, we arbitrarily pick all but one of the bit strings of length n, and let S' be the remaining bit string.[10]

From a purely information-theoretic perspective, it is not immediately clear that quantum samples should provide any advantage over classical samples for such sample-to-sample tasks. Since the required output is a classical n-bit string, it is natural to compare the amount of classical information received per sample. Holevo's theorem guarantees that the maximum information content per sample is the same for quantum and classical samples: both yield at most n bits of classical information [Hol73].[11]

Yet, in Chapter 9, we will show that a particular sample-to-sample task cap-

[10]Note that this is just a reformulation of a search problem in a sample-to-sample setting.

[11]The same argument is known to give misleading intuition in the context of quantum communication complexity, where the required output is typically just a single bit. Nonetheless, we include it here to justify our comparison between classical and quantum samples in the sample-to-sample setting, as the models would clearly be incomparable if it were *not* true.

tures, in the best possible way, how quantum computers can leverage quantum resources in ways for which there is no classical counterpart.

(R6) There exists a sample-to-sample problem which has sample complexity

- exactly 1 for quantum samples;
- $\Theta(2^n)$ for classical samples, even when a small failure probability is allowed.

The task, which we call *complement sampling* (see Fig. 1.1), has the property that for some $S \subset \{0,1\}^n$ we always take the set $S' = \bar{S} = \{0,1\}^n \setminus S$ (the complement).

From a theoretical perspective, the above result exhibits a remarkable feature: it shows that in a sample-to-sample setting, quantum computation can achieve the largest possible separation from classical computation, namely a gap of constant versus a linear (in $N = 2^n$) number of samples. This contrasts with, for example, the case of black-box query complexity, as Aaronson and Ambainis showed that no partial Boolean function has constant quantum query complexity while requiring a linear number of randomized queries [AA15].[12]

We will also show that under the assumption of the existence of one-way functions, complement sampling gives provable, verifiable and (possibly) NISQable[13] quantum advantage in a sample complexity setting. Moreover, we argue that complement sampling is another key example of how a *limitation* of quantum information—namely, that of "destroying" a quantum sample upon measurement—is at the heart of why such a large separation in sample complexity is possible.

[12] Aaronson and Ambainis proved that the separation can be at most constant versus $\tilde{\Omega}(\sqrt{N})$, which is achieved by the Forrelation problem.

[13] Implementable on near-term noisy intermediate-scale quantum devices.

Chapter 2

Background on quantum computational complexity

In this chapter, we introduce some basic notation and give an introduction to the basic prerequisites of quantum computational complexity theory. This is by no means exhaustive, nor intended to be; for much more detailed background information, we recommend [AB09, Wat08] for complexity theory, [Wat18] for quantum information theory, and [NC10, dW19] for quantum computing in general.

2.1 Notation

The symbols $\mathbb{C}$, $\mathbb{R}$, $\mathbb{Z}$, $\mathbb{Z}_+$, and $\mathbb{N}$ represent the sets of complex, real, integer, positive integer, and natural numbers, respectively. For a positive integer N, we use the notation $[N]$ to denote the set $\{1, \ldots, N\}$. If X is a finite set, $|X|$ refers to its size, and $\mathrm{conv}(X)$ represents the convex hull of the elements in X.

For a positive integer d, we denote by S_d the symmetric group on d elements; $\mathbb{U}(d)$ the unitary group of degree d; and by $\mathbb{SU}(d)$ the special unitary group, where all matrices have determinant 1.

When the base is not explicitly stated, log always refers to the base-2 logarithm. The natural logarithm is denoted as ln. We use $\binom{n}{k}$ for the binomial coefficient, and write $\binom{[n]}{k}$ for the set of all k-element subsets of $[n]$ (so unordered and without repetitions). We denote $[n]^k$ for the set of all k-tuples of elements from $[n]$ (ordered and allowing repetitions). Given a q-tuple $(i_1, \ldots, i_q)$ of q unique elements, $\mathrm{Perm}(i_1, \ldots, i_q) := \{\pi(i_1, \ldots, i_q) : \pi \in \mathrm{S}_q\}$ denotes the set of all $q!$ permutations of the tuple.

For bit strings $x, y \in \{0,1\}^n$, we use $x \oplus y$ to denote their bitwise XOR operation. For a set S, we write S^* to denote the set of all arbitrary-length tuples of elements from S. Specifically, $\{0,1\}^*$ represents the set of all bit strings.

When we write $\mathcal{H}$, possibly with subscripts or superscripts, it always refers

to a Hilbert space unless stated otherwise, and we will generally omit specifying this explicitly when clear from context. We often write $\mathcal{H} := \mathbb{C}^d$ for some positive integer $d \in \mathbb{N}$, implying that the Hilbert space is equipped with the standard inner product over $\mathbb{C}^d$. Specifically, for $u, v \in \mathbb{C}^d$, the inner product is $u \cdot v = \sum_{i \in [d]} u_i^* v_i$, where u_i^* is the complex conjugate of u_i. Unless otherwise specified, Hilbert spaces used in formal statements in this thesis will be finite-dimensional, and we write $\dim(\mathcal{H})$ for the dimension of $\mathcal{H}$.

We use bra-ket (Dirac) notation to denote quantum states. A ket $|v\rangle$ denotes a vector v in a Hilbert space $\mathcal{H}$. Formally, a bra is a linear functional on $\mathcal{H}$, defined by taking the inner product. Specifically, a vector $\langle w| : \mathcal{H} \to \mathbb{C}$, when evaluated at a vector $|v\rangle \in \mathcal{H}$, equals the inner product of $|w\rangle$ and $|v\rangle$, written as $\langle w|v\rangle$, which is called a *braket*. When $\mathcal{H} = \mathbb{C}^d$, a column vector $\psi = (\psi_1, \ldots, \psi_d)^\intercal$ corresponds to $|\psi\rangle$ and $\langle\psi|$ denotes the conjugate-transpose row vector, i.e., $\langle\psi| = (\psi_1^*, \ldots, \psi_d^*)$.

For Hilbert spaces $\mathcal{H}$ and $\mathcal{H}'$, we write $\mathrm{L}(\mathcal{H}, \mathcal{H}')$ to denote the set of all linear maps from $\mathcal{H}$ to $\mathcal{H}'$. When $\mathcal{H}' = \mathcal{H}$, we simply write $\mathrm{L}(\mathcal{H})$. We denote the sets of Hermitian, positive semidefinite, unitary, and density operator on $\mathcal{H}$ as $\mathrm{Herm}(\mathcal{H})$, $\mathrm{PSD}(\mathcal{H})$, $\mathrm{U}(\mathcal{H})$ and $\mathrm{D}(\mathcal{H})$, respectively. For a subspace $S \subseteq \mathcal{H}$, we write Π_S for the projector onto S. For any linear map $A \in \mathrm{L}(\mathcal{H}, \mathcal{H}')$, we write $\mathrm{rank}(A)$ for the rank of A, and denote the full set of eigenvalues of A as $\mathrm{eig}(A)$. If A is also Hermitian, i.e., $A \in \mathrm{Herm}(\mathcal{H})$, we write $\lambda_i(A)$ for its ith eigenvalue, with the eigenvalues ordered non-decreasingly. If two eigenvalues are equal, their relative order is arbitrary. Again for $A \in \mathrm{Herm}(\mathcal{H})$, we sometimes write $\dim(A)$ instead of $\dim(\mathcal{H})$ when the Hilbert space is only implicitly defined. For $A, B \in \mathrm{Herm}(\mathcal{H})$, we write $A \succeq B$ to indicate that $A - B$ is positive semidefinite. When we write $A \succeq c$ for some scalar $c \in \mathbb{R}$, we mean that $A \succeq c\mathbb{I}$ holds, where $\mathbb{I}$ is the identity operator on $\mathcal{H}$. The trace of an operator $A \in \mathrm{Herm}(\mathcal{H})$ is denoted by $\mathrm{tr}[A]$, and is defined as the sum of the diagonal entries of A with respect to any orthonormal basis of $\mathcal{H}$. For a linear operator $A \in \mathrm{L}(\mathcal{H}, \mathcal{H}')$ and a subspace $\mathcal{S} \subseteq \mathcal{H}$, we write $A|_{\mathcal{S}}$ to denote the restriction of A to the domain $\mathcal{S}$.

Let $x \in \mathcal{H}$ be a vector in a Hilbert space. The ℓ_p-norm of x, for $1 \leq p \leq \infty$, is defined as[1]

$$\|x\|_p := \left(\sum_i |x_i|^p \right)^{1/p}.$$

Let $A \in \mathrm{L}(\mathcal{H}, \mathcal{H}')$ be a linear map between Hilbert spaces. The Schatten-p norm of A, for $1 \leq p \leq \infty$, is defined as

$$\|A\|_p := \left(\mathrm{tr}\left(\left(\sqrt{A^\dagger A} \right)^p \right) \right)^{1/p},$$

[1]Where the case $p = \infty$ corresponds to taking the limit $p \to \infty$, which results in $\|x\|_\infty = \max_i |x_i|$. For the Schatten p-norm on operators, $p = \infty$ is defined similarly.

where $A^\dagger$ is the adjoint of A. Special cases include:

$$\|A\|_1 := \operatorname{tr}\left(\sqrt{A^\dagger A}\right), \quad \|A\|_2 := \left(\operatorname{tr}\left(A^\dagger A\right)\right)^{1/2}, \quad \|A\|_\infty := \max_{\|x\|_2=1} \|Ax\|_2,$$

which are the trace, Frobenius, and operator (or spectral) norms of A, respectively. Because the operator norm on linear maps and the Euclidean norm (ℓ_2-norm) on vectors play such a prominent role throughout this thesis, we will generally omit the subscript in the norm notation when dealing with these two norms; it will be clear from context whether the norm is over a vector or operator.

Many statements in this dissertation will hold asymptotically, i.e., when a relevant parameter grows sufficiently large, and it will often be useful to express this using big-$\mathcal{O}$ notation. Let $f, g : \mathbb{N} \to \mathbb{R}$ be two functions. We write $f(n) = \mathcal{O}(g(n))$ if there exist constants $c > 0$ and $n_0 \in \mathbb{N}$ such that $f(n) \leq c\, g(n)$ for all $n \geq n_0$. We say $f(n) = \Omega(g(n))$ if and only if $g(n) = \mathcal{O}(f(n))$, and $f(n) = \Theta(g(n))$ if both $f(n) = \mathcal{O}(g(n))$ and $f(n) = \Omega(g(n))$ hold. Moreover, $f(n) = o(g(n))$ means that for every $\epsilon > 0$, there exists an $n_0 \in \mathbb{N}$ such that $f(n) < \epsilon g(n)$ for all $n \geq n_0$; conversely, $f(n) = \omega(g(n))$ holds if $g(n) = o(f(n))$. Finally, we write $f(n) = \tilde{\Omega}(g(n))$ to mean that $f(n)$ equals $\Omega(g(n))$ up to polylogarithmic factors. This notation extends analogously to the other asymptotic forms introduced above.

We use $\operatorname{poly}(n)$ to denote any function $f : \mathbb{N} \to \mathbb{R}$ such that $f(n) = \mathcal{O}(n^c)$ for some constant $c > 0$. Similarly, $\operatorname{polylog}(n)$ denotes any function of the form $\mathcal{O}((\log n)^c)$ for some constant $c > 0$, and $\exp(n)$ denotes any function of the form $\mathcal{O}(2^{n^c})$ for some constant $c > 0$. We extend these definitions in the natural way to functions with real-valued inputs and multiple arguments.

2.2 Quantum mechanics: the rules of the game

The discovery of modern quantum theory in the 1920s brought about one of the greatest revolutions in our understanding of nature. By unifying the wave and particle interpretations of light and matter, the theory provided solutions to problems such as the photoelectric effect, Compton scattering, and black-body radiation.

In modern society, the word "quantum" is often associated with being "hard and complex[2]". While many of its implications—like particle-wave duality, superposition, and entanglement—are very counter-intuitive to how we experience the world in our daily lives, the "rules of the game" are deceptively simple: the theory is built on just four basic postulates, all grounded in linear algebra.

[2]Though this association is not entirely unjustified, as the state vectors have complex entries.

2.2.1 The four postulates

In order to do all of quantum mechanics, one only needs the following four axioms.

2.2.1. Axiom (Quantum System). *A quantum system is described by a unit vector $|\psi\rangle$ in a complex Hilbert space $\mathcal{H}$.*

We will only be concerned with finite-dimensional Hilbert spaces, which are finite-dimensional complex vector spaces equipped with an inner product.[3]

The second axiom deals with how closed quantum systems evolve over time.

2.2.2. Axiom (Evolution). *The time evolution of a closed quantum system associated with a Hilbert space $\mathcal{H}$ is governed by the Schrödinger equation. For $t \geq 0$, the state vector $|\psi(t)\rangle \in \mathcal{H}$, initially $|\psi(0)\rangle$, evolves according to a* unitary operator $U(t)$*, given in terms of the system's Hamiltonian H (the total energy operator) by*

$$|\psi(t)\rangle = U(t)\,|\psi(0)\rangle = e^{-iHt/\hbar}\,|\psi(0)\rangle\,,$$

where $\hbar$ is the reduced Planck constant and i is the imaginary unit.

Axiom 2.2.2 states that, at least in principle, knowing the Hamiltonian H and the initial state $|\psi(0)\rangle$ provides all the information needed to compute the state $|\psi(t)\rangle$ at any future time.

Observing a quantum system, formally called a *measurement*, breaks the quantum system's status as "closed", as it necessarily involves an interaction with an external environment (the observer). The next axiom explains how a measurement operation interacts with the quantum system being observed.

2.2.3. Axiom (Measurement). *Let $\mathcal{H}$ be the Hilbert space of a physical system. Let Ω be a set of measurement outcomes, where to each outcome $\omega \in \Omega$ we associate a subspace $S_\omega \subseteq \mathcal{H}$, forming an orthogonal decomposition of $\mathcal{H}$, i.e., $\mathcal{H} = \bigoplus_{\omega\in\Omega} S_\omega$. Suppose the system is in state $|\psi\rangle \in \mathcal{H}$, and we perform a measurement corresponding to this decomposition. Then the probability of observing outcome $\omega \in \Omega$ is given by Born's rule:*

$$\Pr[\omega] = \|\Pi_{S_\omega}\,|\psi\rangle\,\|^2,$$

where Π_{S_ω} denotes the orthogonal projector onto S_ω. Conditioned on observing outcome ω, the post-measurement quantum state is

$$|\psi'\rangle = \frac{\Pi_{S_\omega}\,|\psi\rangle}{\sqrt{\Pr[\omega]}}.$$

[3] In the finite-dimensional setting, every inner product space is complete with respect to the norm it induces, so the completeness condition of a Hilbert space is automatically satisfied.

In the above axiomatic description, we formulated measurements as so-called projection-valued measures (PVMs). It is possible to define measurements more generally in terms of POVMs (Positive Operator-Valued Measures), which allow measurement operators that do not necessarily correspond to orthogonal projectors. A POVM is described by a set of m positive semi-definite operators $\{E_i\}_{i\in[m]}$ acting on the Hilbert space $\mathcal{H}$ with the properties

- (Positivity:) $E_i \geq 0$ for all $i \in [m]$;
- (Completeness:) $\sum_{i\in[m]} E_i = \mathbb{I}$.

Although POVMs describe a broader class of measurements on a given system, Naimark's dilation theorem shows that any POVM can be implemented as a projective measurement on a larger Hilbert space, such that the outcome statistics on the original system are preserved [Nai40, Pau03]. Hence, the axiomatic description does not need to account for these more general measurement types.

Finally, the fourth and last axiom deals with *system composition*.

2.2.4. Axiom (System Composition). *Let n be a positive integer, and consider a quantum system composed of n subsystems with associated Hilbert spaces $\mathcal{H}_1, \ldots, \mathcal{H}_n$. Then the overall Hilbert space is given by the tensor product*

$$\mathcal{H} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n.$$

Axiom 2.2.4 brings us back to Feynman's argument from Chapter 1 about why quantum systems are difficult to simulate using their state vector description: if a quantum system consists of n subsystems, each associated with a two-dimensional Hilbert space $\mathcal{H}_0 := \mathbb{C}^2$, then the total system lives in the Hilbert space $\mathcal{H} = \mathcal{H}_0^{\otimes n}$, which has dimension 2^n. Consequently, fully describing an arbitrary state $|\psi\rangle \in \mathcal{H}$ requires specifying 2^n complex numbers.

2.2.2 The density matrix formalism

More generally, one can also capture *uncertainty* about a quantum system by using the *density matrix* (or *mixed state*) formalism, which generalizes the *pure states* used in our axioms in Section 2.2.1.[4]

[4]Just as in the case of POVMs, axioms based on the density matrix formalism do not introduce any new physics: any experiment involving mixed states can be viewed as a procedure in which a random choice is made at the beginning (e.g., by flipping coins), after which fixed quantum operations are applied deterministically to the resulting pure state. An alternative argument is via the existence of purifications, which will be explained later.

2.2.1. Definition (Quantum state). A quantum state is a semidefinite operator with trace one. We denote by

$$\mathrm{D}(\mathcal{H}) = \{\rho \in \mathrm{PSD}(\mathcal{H}) \mid \mathrm{tr}[\rho] = 1\}$$

the set of all possible quantum states on a Hilbert space $\mathcal{H}$.

The evolution of a mixed state ρ under a unitary U is given by $U\rho U^\dagger$, and Born's rule for a projective measurement Π_{S_ω} is defined in terms of the trace, i.e., $\mathrm{tr}[\Pi_{S_\omega}\rho]$. More generally, the expectation value of an observable A is given by $\langle A\rangle = \mathrm{tr}[\rho A]$. System composition again follows the tensor product formulation. Hence, it is straightforward to define all the axioms in Section 2.2.1 in terms of the density matrix formalism.

By the spectral theorem, any mixed state can be represented as a statistical ensemble of pure states $|\psi_i\rangle$ with corresponding probabilities $p(i)$. The density matrix ρ for a mixed state is then given by

$$\rho = \sum_i p(i)\,|\psi_i\rangle\langle\psi_i|\,.$$

Given a description of the density matrix ρ, it is easy to check if ρ is a pure state: for a pure state, $\rho^2 = \rho$ (i.e., ρ is a rank-one projector) and $\mathrm{tr}[\rho^2] = 1$, whereas for a mixed state, $\mathrm{tr}[\rho^2] < 1$.

For a quantum system of dimension d, the normalized identity matrix $\mathbb{I}_d/d$ represents the so-called *maximally mixed state*, which is the density matrix with maximum entropy: in this state, all possible outcomes are equally probable, and there is complete uncertainty about the system's state.[5]

Partial trace. An important operation in the density matrix formalism is the *partial trace*. The partial trace is used to obtain the reduced density matrix of a subsystem from the density matrix of a larger, composite quantum system. This operation effectively "traces out" the degrees of freedom associated with the part of the system that is not of interest, leaving us with a description of the remaining subsystem.

2.2.2. Definition (Partial Trace). Consider a composite quantum system with associated Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ in a state $\rho_{AB} \in \mathrm{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$. The partial trace over subsystem B, denoted by $\mathrm{tr}_B : \mathrm{D}(\mathcal{H}_A \otimes \mathcal{H}_B) \to \mathrm{D}(\mathcal{H}_A)$, is defined as

$$\mathrm{tr}_B[\rho_{AB}] = \sum_b \left(\mathbb{I}_A \otimes \langle b|\right)\rho_{AB}\left(\mathbb{I}_A \otimes |b\rangle\right),$$

where $\{|b\rangle\}$ is any orthonormal basis for $\mathcal{H}_B$. We often write $\rho_A = \mathrm{tr}_B[\rho_{AB}]$.

[5] Note what happens to the maximally mixed state under any unitary transformation.

Purification. For every mixed state $\rho \in \mathrm{D}(\mathcal{H})$, there exists a *purification* $|\psi\rangle \in \mathcal{H} \otimes \mathcal{H}_{\mathrm{aux}}$, where $\mathcal{H}_{\mathrm{aux}}$ is called the auxiliary space, provided that $\dim(\mathcal{H}_{\mathrm{aux}}) \geq \dim(\mathcal{H})$.[6] This is easily seen by considering the spectral decomposition $\rho = \sum_i p(i) \, |\phi_i\rangle\langle\phi_i|$ and defining

$$|\psi\rangle = \sum_i \sqrt{p(i)} \, |\phi_i\rangle \otimes |i\rangle_{\mathrm{aux}} \, ,$$

where $\{|i\rangle\}$ is an orthonormal basis of $\mathcal{H}_{\mathrm{aux}}$. Since both the auxiliary space and the basis can be chosen arbitrarily, the purification of a mixed state is not unique. However, all purifications are related by isometries on the auxiliary space.

Given two mixed states, it is also possible to relate the different purifications to their fidelity.

2.2.3. Theorem (Uhlmann's Theorem [Uhl76]). *Let $\rho, \sigma \in \mathrm{D}(\mathcal{H}_A)$ be density operators on a Hilbert space $\mathcal{H}_A$, and let $\mathcal{H}_B$ be a Hilbert space such that both states admit purifications on $\mathcal{H}_A \otimes \mathcal{H}_B$. Then*

$$F(\rho, \sigma) = \max_{|\psi_\rho\rangle, |\psi_\sigma\rangle} |\langle\psi_\rho|\psi_\sigma\rangle|^2 \, ,$$

where the maximum is over all purifications $|\psi_\rho\rangle, |\psi_\sigma\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, i.e.,

$$\rho = \mathrm{tr}_{\mathcal{H}_B} \left(|\psi_\rho\rangle\langle\psi_\rho| \right) , \quad \sigma = \mathrm{tr}_{\mathcal{H}_B} \left(|\psi_\sigma\rangle\langle\psi_\sigma| \right) .$$

In other words, the fidelity between two mixed states is the maximum fidelity between all possible purifications of the states.

2.2.3 Some peculiarities of quantum mechanics

Entanglement and non-separability. Let AB be a composite quantum system consisting of subsystems A and B. In general, not every state of the joint system can be written in the product form $|\psi_a\rangle_A \otimes |\psi_b\rangle_B$. For example, consider the state

$$\frac{1}{\sqrt{2}} \left(|\psi_0\rangle_A \, |\psi_0\rangle_B + |\psi_1\rangle_A \, |\psi_1\rangle_B \right) , \tag{2.1}$$

where $|\psi_0\rangle$ and $|\psi_1\rangle$ are two orthogonal states from some Hilbert space $\mathcal{H}$. In particular, a measurement of system A in the basis $\{|\psi_0\rangle \, , |\psi_1\rangle \, , \dots\}$ will also "collapse" the B system to the same state according to the measurement outcome. This inherently quantum phenomenon is called *entanglement*.

[6] In fact, the condition $\dim(\mathcal{H}_{\mathrm{aux}}) \geq \mathrm{rank}(\rho)$ is sufficient, but this refinement is not needed for this dissertation.

2.2.4. Definition (Entanglement). Let $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ be some composite quantum system. A state $|\psi\rangle \in \mathcal{H}_{AB}$ is said to be *entangled* if it cannot be written as a product state of the form

$$|\psi\rangle = |\psi_a\rangle \otimes |\psi_b\rangle \,,$$

where $|\psi_a\rangle \in \mathcal{H}_A$ and $|\psi_b\rangle \in \mathcal{H}_B$.

For mixed states, entangled states are defined as the set of all states that are not *separable*, where a separable state is a generalisation of a pure state of the form

$$\rho = \sum_i p(i) \rho_a^i \otimes \rho_b^i$$

where $\sum_i p(i) = 1$. Hence, for separable states we can have correlations between the two parties' individual measurements, but they can be accounted for solely by classical correlations.

No-signalling. Can entanglement be used to send messages across large distances instantaneously, since measuring system A affects the state of system B, even when they are far apart? To Einstein's relief,[7] the answer is no: the *no-signalling* (or no-communication) theorem asserts that it is impossible for one observer, by measuring part of an entangled quantum state, to communicate information to another observer.

2.2.5. Theorem (No-Signalling). *Let ρ_{AB} be a quantum state on the composite Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$. Let $\{E_a^A\}$ be a POVM on A and $\{E_b^B\}$ a POVM on B. Then, for any unitary $U \in \mathrm{U}(\mathcal{H}_B)$, the marginal probability of obtaining outcome a is*

$$\sum_b \mathrm{tr}\left[(E_a^A \otimes E_b^B)\ (\mathbb{I} \otimes U)\, \rho_{AB} \left(\mathbb{I} \otimes U^\dagger\right)\right] = \mathrm{tr}\left[E_a^A \rho_A\right],$$

where $\rho_A = \mathrm{tr}_B[\rho_{AB}]$. That is, local operations on B (including unitaries and measurements) do not affect the statistics observed on A.

For a proof, see [Chi14, Section 3.4]. Since the role of the registers A and B is symmetric (and their labelling arbitrary), local operations on A do not affect the statistics observed on B.

[7] If this were possible, it would contradict general relativity, which prohibits information transfer faster than the speed of light.

No cloning. Unlike classical information, quantum states cannot be cloned by any unitary operation acting on a larger Hilbert space.

2.2.6. Proposition (No cloning). *Let $\mathcal{H}' = \mathcal{H} \otimes \mathcal{H}_{\text{ext}}$ be a composite Hilbert space with $\dim(\mathcal{H}_{\text{ext}}) \geq \dim(\mathcal{H})$. Let $|0\rangle \in \mathcal{H}_{\text{ext}}$ be an arbitrary fixed state. Then, there exists no unitary $U : \mathcal{H}' \to \mathcal{H}'$ that performs the mapping*

$$U|\psi\rangle|0\rangle = |\psi\rangle|\psi\rangle|g(\psi)\rangle,$$

for some state $|g(\psi)\rangle$ that may depend on $|\psi\rangle$, for every $|\psi\rangle \in \mathcal{H}$.

This is easy to show, as unitarity implies the conservation of inner products.[8]

2.3 Quantum information

Given the rules of quantum mechanics, *quantum information* concerns how quantum systems can be used to store, process, and transmit information. Quantum systems come in many forms, such as photon polarisation, electronic states, spin states, and more. To abstract away from the specific physical systems and their interactions, we will introduce the concepts of qubits and channels.

2.3.1 Qubits and channels

The fundamental unit of information in classical information theory, called the *bit*, is a simple two-level system. In vector notation, the state space of any deterministic two-level system can be written as $\{|0\rangle, |1\rangle\}$, where $|0\rangle$ and $|1\rangle$ are the standard *computational* basis vectors, given by

$$|0\rangle := \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle := \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

If you allow for randomness, i.e., probabilistic mixtures of deterministic states, any possible two-level state can be modelled as $p_0|0\rangle + p_1|1\rangle$ with $p_0, p_1 \geq 0$ and $p_0 + p_1 = 1$. In *quantum information*, the fundamental unit of information is the so-called *qubit*, which generalises the probabilistic case by allowing for complex numbers:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle,$$

[8] Suppose, for the sake of contradiction, that such a unitary U exists. Take two distinct states $|\psi_1\rangle$ and $|\psi_2\rangle$ such that $\langle\psi_1|\psi_2\rangle = \alpha$ with $0 < |\alpha| < 1$. Then, $\alpha = (\langle\psi_1|\langle 0|)(|\psi_2\rangle|0\rangle) = (\langle\psi_1|\langle 0|)U^\dagger U(|\psi_2\rangle|0\rangle) = (\langle\psi_1|\psi_2\rangle)^2 \langle g(\psi_1)|g(\psi_2)\rangle = \alpha^2 \langle g(\psi_1)|g(\psi_2)\rangle$. Taking absolute values gives $|\alpha| = |\alpha|^2 \cdot |\langle g(\psi_1)|g(\psi_2)\rangle| \leq |\alpha|^2$. Since $0 < |\alpha| < 1$, this implies $|\alpha| \leq |\alpha|^2 < |\alpha|$, a contradiction. Therefore, no such unitary U can exist.

where $\alpha, \beta \in \mathbb{C}$ and $|\alpha|^2 + |\beta|^2 = 1$. We will usually write any n-qubit state in the basis of n-bit computational basis states. That is, for any n-qubit state $|\phi\rangle$, we can write

$$|\phi\rangle = \sum_{x \in \{0,1\}^n} \alpha_x |x\rangle ,$$

with $\sum_{x \in \{0,1\}^n} |\alpha_x|^2 = 1$. In the density matrix formalism, a qubit is considered to be any state $\rho \in \mathrm{D}(\mathbb{C}^2)$.

The most general way to manipulate quantum information is through *quantum channels*.

2.3.1. Definition (Quantum channel). A quantum channel is a completely positive, trace-preserving (CPTP) linear map.

Examples of quantum channels include unitary or isometric transformations ($\mathcal{E}(\rho) = V\rho V^\dagger$), measurements ($\mathcal{E}(\rho) = \sum_j \mathrm{tr}[\rho O_j]|j\rangle\langle j|$), and tracing out a subsystem ($\mathcal{E}(\rho_{AB}) = \mathrm{tr}_B(\rho_{AB})$).

2.3.2 Distance measures

The following distance measures on quantum states and channels will play an important role throughout this dissertation.

States. We first consider the *trace distance*, which is a measure of the distance between two quantum states.

2.3.2. Definition (Trace distance). For two density matrices $\rho, \sigma \in \mathrm{D}(\mathcal{H})$, the trace distance $D(\rho, \sigma)$ is defined as

$$D(\rho, \sigma) = \frac{1}{2}\|\rho - \sigma\|_1 = \frac{1}{2}\ \mathrm{tr}\left[\sqrt{(\rho - \sigma)^\dagger(\rho - \sigma)}\right].$$

2.3.3. Proposition (Properties of the trace distance). *The following properties hold for the trace distance with respect to all density matrices $\rho, \sigma \in \mathrm{D}(\mathcal{H}_A)$:*

(i) $D(\rho, \sigma)$ is a metric on the space of density matrices: it is non-negative, symmetric, satisfies the triangle inequality, and $D(\rho, \sigma) = 0$ if and only if $\rho = \sigma$.

(ii) Unitary invariance: For any unitary $U \in \mathrm{U}(\mathcal{H}_A)$, we have $D(\rho, \sigma) = D(U\rho U^\dagger, U\sigma U^\dagger)$.

(iii) Tensor product property: For any $\gamma \in \mathrm{D}(\mathcal{H}_B)$, $D(\rho \otimes \gamma, \sigma \otimes \gamma) = D(\rho, \sigma)$.

(iv) Variational form: $D(\rho, \sigma) = \sup_{0 \preceq P \preceq \mathbb{I}} \mathrm{tr}[P(\rho - \sigma)]$.

Point (iv) of Proposition 2.3.3 is also known as the operational interpretation of the trace distance, as it shows that the trace distance precisely characterises the optimal success probability of distinguishing two input states ρ and σ given with equal prior probability. This can be seen as follows: consider any POVM $E = \{E_1, E_2\}$ with $P = E_1$ and $E_2 = \mathbb{I} - P$. If we observe outcome "1", we conclude ρ; if we observe outcome "2", we conclude σ. Assuming that ρ and σ are each given with probability $1/2$, the success probability of correctly identifying the input state is

$$p_{\text{guess}} = \frac{1}{2}\Pr[1|\rho] + \frac{1}{2}\Pr[2|\sigma] = \frac{1}{2}\operatorname{tr}[E_1\rho] + \frac{1}{2}\operatorname{tr}[E_2\sigma] = \frac{1}{2}\left(1 + \operatorname{tr}\left[P(\rho - \sigma)\right]\right).$$

Maximising over all such measurements yields

$$p_{\text{guess}}^{\max} = \sup_{0 \preceq P \preceq \mathbb{I}} \frac{1}{2}\left(1 + \operatorname{tr}\left[P(\rho - \sigma)\right]\right) = \frac{1}{2} + \frac{1}{2}D(\rho, \sigma). \tag{2.2}$$

More generally, the trace distance can also be used to characterise the so-called *total variation* distance between two probability distributions induced by POVMs on either ρ or σ. To do this, we first define the total variation distance for probability distributions on finite domains.

2.3.4. Definition (Total variation distance). Let $p, q : E \to [0, 1]$ be probability distributions over a finite set E. The total variation distance between p and q is defined as

$$D_{\text{TV}}(p, q) := \frac{1}{2}\sum_{x \in E} |p(x) - q(x)|.$$

Analogous to the trace distance, total variation distance characterises the maximum achievable bias for a distinguisher between two probability distributions given only a single sample.

Observe that if ρ and σ are diagonal matrices, given by $\rho = \sum_i p(i)\,|i\rangle\langle i|$ and $\sigma = \sum_i q(i)\,|i\rangle\langle i|$ for some distributions p and q on a finite set $E = \{i\}$, then we automatically have

$$D(\rho, \sigma) = \frac{1}{2}\operatorname{tr}\left[\sqrt{(\rho - \sigma)^2}\right] = \frac{1}{2}\sum_i |p(i) - q(i)| = D_{\text{TV}}(p, q),$$

as was to be expected, since diagonal density matrices can be interpreted as classical probability distributions over an orthonormal basis. For arbitrary density matrices, the trace distance between the states equals the total variation distance between the outcome distributions of the POVM that maximises this distance, as shown in the following lemma.

2.3.5. Lemma ([NC10, Theorem 9.1]). *Let $\{E_m\}$ be a POVM. Define probability distributions p and q by $p_m = \mathrm{tr}[E_m\rho]$ and $q_m = \mathrm{tr}[E_m\sigma]$. Then,*

$$\max_{\{E_m\}} D_{\mathrm{TV}}(q_m, p_m) = D(\rho, \sigma).$$

For any fixed POVM applied to two possible input states, Lemma 2.3.5 provides a useful upper bound: the total variation distance between the resulting outcome distributions is bounded above by the trace distance between the states.

Quantum channels. Similar to quantum states, we also need a definition of a distance on the space of quantum channels. For this, we use the *diamond distance*.

2.3.6. Definition (Diamond distance). The diamond distance $\|\mathcal{E} - \mathcal{F}\|_\diamond$ between two quantum channels $\mathcal{E}, \mathcal{F} : \mathrm{D}(\mathcal{H}_1) \mapsto \mathrm{D}(\mathcal{H}_2)$ is defined as

$$\|\mathcal{E} - \mathcal{F}\|_\diamond = \sup_{\rho \in \mathrm{D}(\mathcal{H}_1 \otimes \mathcal{H}_{\mathrm{ext}})} \|(\mathcal{E} \otimes \mathbb{I})(\rho) - (\mathcal{F} \otimes \mathbb{I})(\rho)\|_1,$$

where $\dim(\mathcal{H}_{\mathrm{ext}}) \geq \dim(\mathcal{H}_1)$.

Analogous to how the trace distance characterises the optimal success probability of distinguishing two quantum states given a single copy and no prior bias, the diamond distance captures the optimal success probability of distinguishing two quantum channels when the unknown channel can be applied only once. This can be directly seen from its definition, since it involves a maximisation over both input states and measurements on the output.

The following properties concerning the diamond norm will be useful throughout this dissertation.

2.3.7. Proposition (Properties of the diamond distance). *The following properties hold for the diamond distance with respect to all quantum channels $\mathcal{E}, \mathcal{E}' : \mathrm{D}(\mathcal{H}_1) \mapsto \mathrm{D}(\mathcal{H}_2)$:*

(i) It is a metric on the space of quantum channels: it is non-negative, symmetric, satisfies the triangle inequality, and $\|\mathcal{E} - \mathcal{E}'\|_\diamond = 0$ if and only if $\mathcal{E} = \mathcal{E}'$.

(ii) Unitary invariance: For all unitaries $U \in \mathrm{U}(\mathcal{H}_1)$ and $V \in \mathrm{U}(\mathcal{H}_2)$, we have $\|\mathcal{E} - \mathcal{E}'\|_\diamond = \|\mathcal{F} - \mathcal{F}'\|_\diamond$, where $\mathcal{F}, \mathcal{F}' \in \mathrm{D}(\mathcal{H}_1) \mapsto \mathrm{D}(\mathcal{H}_2)$ are given by

$$\mathcal{F}(\rho) = V\mathcal{E}(U\rho U^\dagger)V^\dagger \quad \text{and} \quad \mathcal{F}'(\rho) = V\mathcal{E}'(U\rho U^\dagger)V^\dagger,$$

for any input $\rho \in \mathrm{D}(\mathcal{H}_1)$.

(iii) Hybrid argument: For all quantum channels $\mathcal{F}, \mathcal{F}' \in \mathrm{D}(\mathcal{H}_2) \mapsto \mathrm{D}(\mathcal{H}_3)$, we have

$$\|\mathcal{F}\mathcal{E} - \mathcal{F}'\mathcal{E}'\|_\diamond \leq \|\mathcal{E} - \mathcal{E}'\|_\diamond + \|\mathcal{F} - \mathcal{F}'\|_\diamond. \tag{2.3}$$

Proofs of all of the above statements can be found in [Wat18].

2.4 Computational complexity theory

Computational complexity theory explores how the computational resources required to solve a problem scale with some measure of the problem size n. These resources can include time, memory (space), energy, bits of communication, queries to an all-powerful oracle, and many more. To meaningfully study and compare how different algorithms or protocols make use of computational resources, we must first fix a model of computation in both the classical and quantum settings. To that end, we introduce the Turing machine and the quantum circuit model.

2.4.1 Turing machines and the quantum circuit model

When reasoning about resource efficiency, it may seem necessary to be extremely careful in choosing a mathematical definition for a computational model. However, it turns out that a single simple model—the *Turing machine*—seems to capture all (classical) realisable computational methods with at most a polynomial loss of efficiency.

Turing machines. The Turing machine, named after Alan Turing who introduced the model in 1936 [Tur36], operates on an infinite tape divided into cells, each labelled by an integer. The machine has a read/write head that moves left or right along the tape (denoted by $D = \{L, R\}$) and can write symbols from a finite alphabet Σ to the current cell. Without loss of generality, we take $\Sigma = \{0, 1\}$ to be binary. The machine's actions are controlled by a finite set of states Q, with a starting state q_0 and a halting state q_f. The machine's behaviour is determined by a transition function δ defined as

$$\delta : Q \times \Sigma \to Q \times \Sigma \times D.$$

When the machine is in state $q \in Q$ and reads a symbol $\sigma \in \Sigma$, the function $\delta(q, \sigma) = (q', \sigma', d)$ instructs the machine to write σ' to the current cell, move the head in direction $d \in D$, and transition to state q'.

The computation begins with an input string $x \in \Sigma^n$ written in the first n cells, with the head positioned at cell 0 and the machine in the initial state q_0. The remaining cells contain a special blank symbol "#", indicating that the cell is empty. The transition function is repeatedly applied until the machine reaches the halting state q_f, at which point the non-empty contents of the tape represent the output.

We will also consider the notion of *oracle Turing machines*. An oracle Turing machine is a Turing machine equipped with an additional oracle tape. It can write a string z, called the query, on this tape and enter a special query state. In one step, it transitions to either the q_{yes} or q_{no} state, depending on whether z is in the oracle set $O \subseteq \{0, 1\}^*$. Regardless of the choice of O, a membership query to

O counts as a single computational step. If M is an oracle machine, $O \subseteq \{0,1\}^*$ an oracle, and $x \in \{0,1\}^*$ an input, we denote the output of M on input x with oracle O by $M^O(x)$.

It is also possible to extend all of the above to a *randomised Turing machine*: one can modify the definition by allowing that, at each step, instead of always following a single transition, the machine may have multiple possible transitions based on random decisions (e.g., coin flips). Likewise, a further generalization is possible by considering *quantum Turing machines* [Deu85], but since their definition is cumbersome (and sparsely used in the literature), it will be more useful to define quantum computation in terms of the *quantum circuit model*, which is, when the circuits are uniformly generated by a polynomial-time classical Turing machine, polynomially equivalent to the quantum Turing machine model [CCY93].

The quantum circuit model. Let $n \in \mathbb{N}$ be some notion of input size. In the quantum circuit model, one generally considers access to an m-qubit quantum system, with $m = \mathrm{poly}(n)$, which is initialised in some fixed initial state (usually $|0^m\rangle$). To this initial state, one is allowed to apply so-called quantum gates from a predefined set of *universal* quantum operations $\mathcal{G}$, and finally, a measurement is performed (typically in the computational basis). There are several notions of gate universality [Aha03], but we will use the following definition throughout this dissertation:

2.4.1. Definition. A finite gate set $\mathcal{G}$ is *universal* if, for every positive integer n, every unitary $U \in \mathbb{U}(2^n)$, and every $\epsilon > 0$, there exists a finite sequence of gates $U_1, \ldots, U_T$, with $U_i \in \mathcal{G}$ for all $i \in [T]$, such that $\|U_1 \ldots U_T - U\| \leq \epsilon$.

If $\mathcal{G}$ is finite, then we can only achieve ϵ approximations, as there are uncountably many unitary transformations but only a countable number of quantum circuits constructed from a finite set of gates.

An example of a universal gate set is "Clifford + T", which consists of the generators of the Clifford group and a T-gate:

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},\ S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix},\ \mathrm{CNOT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix},\ T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}.$$

It is known that all universal gate sets are polylogarithmically equivalent by the Solovay–Kitaev theorem, so little is lost by restricting to a single universal gate set. The version we use is adapted from [Chi17, Chapter 2]:

2.4.2. Theorem (Solovay–Kitaev). *Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be two universal gate sets that are closed under inverses.*[9] *Then any M-gate circuit over $\mathcal{G}_1$ can be approximated*

[9] Recently, it was proven that the condition of being closed under inverses is not necessary [BGT21].

to precision $\epsilon > 0$ by a circuit over $\mathcal{G}_2$ of size $\mathcal{O}(M \cdot \mathrm{polylog}(M/\epsilon))$, and there is a classical algorithm for constructing such a circuit in time $\mathcal{O}(M \cdot \mathrm{polylog}(M/\epsilon))$.

The careful reader may notice that there is one key distinction between how one aspect of "efficiency" is defined in the Turing machine model and that of the quantum circuit model: whilst the former specifies the *time* used by the Turing machine, so far in the quantum circuit model we have only talked about efficiency in terms of the number of *gates*. In this dissertation, we will assume that gates from any "standard" universal gate set (if we do not specify one, pick your favourite and fix it throughout) can be implemented in constant time, and the same holds for measurements in the computational basis. Hence, when we say that a quantum algorithm runs in polynomial time, this implies that its underlying quantum circuit has polynomial gate complexity. Similarly, since error correction can be performed using an additional number of gates that is polylogarithmic in the number of initial gates and the inverse of the desired precision (assuming an error rate per gate below a universal threshold) [Sho96], any quantum algorithm specified by a fixed initial state, the application of a polynomial number of gates, and a polynomial number of (intermediate) measurements in the computational basis is immediately considered polynomial-time.

Often, we require that a circuit description can be efficiently generated from an input description. Moreover, this becomes necessary when the input size varies, as different input sizes require different circuits. To formalise this, it is useful to introduce the notion of P-*uniformity*.

2.4.3. Definition (P-Uniformity). Let $S \subseteq \{0,1\}^*$. A family of quantum circuits $\{Q_x : x \in S\}$ is called *polynomial-time uniform* (abbreviated as P-uniform) if there exists a deterministic Turing machine that, on input x, outputs a description of Q_x in time polynomial in $|x|$.

We often consider families of circuits parametrised by the input length $n := |x|$, i.e., $\{Q_n : n \in \mathbb{N}\}$. In Definition 2.4.3, this implies that $V_x = V_y$ whenever $|x| = |y|$. However, Definition 2.4.3 is particularly useful when we want to hard-code the input x into the circuit, a situation that arises frequently throughout this dissertation.

Now that we have defined our models of computation, let us briefly return to the remark about the Church–Turing thesis and its extended version from Chapter 1. Since it is known that a Turing machine can simulate a quantum Turing machine [Deu85] and the uniform quantum circuit model is polynomially equivalent to quantum Turing machines [CCY93], this implies that, from a computability perspective, quantum Turing machines (or quantum computations in the quantum circuit model) and classical Turing machines have the same power. However, as mentioned in Chapter 1, it is widely believed that the extended Church–Turing thesis is violated by quantum computation, meaning that quantum resources can

be more *efficient* than their classical counterparts. To capture the power of classical and quantum computing when resources are limited, we will study so-called *complexity classes*.

2.4.2 Languages versus promise problems

Traditionally, complexity theory defined its classes in terms of *languages*. A language is a function $L : \{0,1\}^* \to \{0,1\}$, partitioning the set of all words $x \in \{0,1\}^*$ into those that are part of the language (i.e., $L(x) = 1$, $x \in L$, or $x \in L_{\text{yes}}$) and those that are not (i.e., $L(x) = 0$, $x \notin L$, or $x \in L_{\text{no}}$).

Informally, *syntactic* complexity classes are those for which one can syntactically enforce that the algorithms defining members of the class are valid (for example, that they run in polynomial time, always output 0 or 1, etc.). This means that the complexity classes can be defined in terms of languages, and that each syntactic class has a "standard" complete language (see Section 2.4.3 for those not familiar with this notion), that is

$$\{(M, x) : M \in \mathcal{M} \text{ and } M(x) = 1\}, \tag{2.4}$$

where $\mathcal{M}$ is the class of all machines of the variant that define the class, appropriately standardised (for example, all polynomial-time Turing machines) [Pap03].

Semantic classes are different: these classes usually have in their definition a property that cannot be easily checked, for example, the existence of a randomised Turing machine M which, for each input x, outputs 1 with probability $\geq 2/3$ or $\leq 1/3$. In fact, for semantic classes, the "standard" complete language as in Eq. (2.4) is usually undecidable [Pap03]. As argued in [Pap03], it is possible to define semantic classes by the absence of such a complete language.

To work around this and still consider some notion of complete problems, it will be useful to consider *promise problems* and *promise classes*. A promise problem generalises the concept of a language by partitioning all words $x \in \{0,1\}^*$ into *three* non-intersecting sets $A = (A_{\text{YES}}, A_{\text{NO}}, A_{\text{INV}})$. We are only concerned with how a class behaves on the so-called *valid inputs*, which are all $x \in (A_{\text{YES}} \cup A_{\text{NO}})$.

As is standard in quantum complexity theory, we will define all classes in terms of promise problems, omitting the prefix "Promise" for a class $\mathcal{C}$ when considering its promise version PromiseC [Wat08]. If we refer to the syntactic definition, we will explicitly state this in the text. We will also generally omit A_{INV} when defining a promise class A, as A_{YES} and A_{NO} implicitly define A_{INV}.

The definitions of (almost all) complexity classes can be found in the Complexity Zoo.[10] For completeness, we have included definitions of all classes considered in this dissertation (excluding those mentioned only in footnotes) that are not already introduced in the main text, in Appendix A.

[10] `https://complexityzoo.net/Complexity_Zoo`.

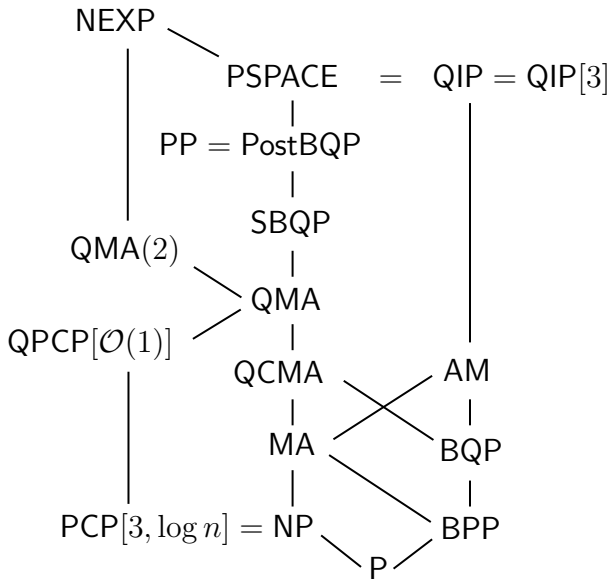


Figure 2.1: Hasse Diagram of (most of) the complexity classes given in Appendix A. Whenever a line from a set A goes upward to a set B, we have $A \subseteq B$.

2.4.3 Reductions, hardness and completeness

We recall three central notions in complexity theory: *reductions*, *hardness*, and *completeness*.

2.4.4. Definition (Polynomial-time reductions). Let $A = (A_{\text{YES}}, A_{\text{NO}})$ and $B = (B_{\text{YES}}, B_{\text{NO}})$ be two promise problems. We say A is polynomial-time Karp reducible to B (shortened to polynomial-time reducible), denoted as $A \leq B$, if there exists a function $f : \{0,1\}^* \to \{0,1\}^*$ such that:

- For every $x \in A_{\text{YES}}$, we have $f(x) \in B_{\text{yes}}$.
- For every $x \in A_{\text{NO}}$, we have $f(x) \in B_{\text{no}}$.
- The function f is computable in time polynomial in $|x|$.

There are also other types of reductions: for example, one can consider functions f that are efficiently computable by randomised computation or even quantum computation, which will be central in Chapter 6 and Chapter 7. One can also consider so-called *Turing reductions*, which informally state that a problem A is Turing-reducible to B if there exists an oracle machine that can compute A by using an oracle for B as a subroutine.

Having defined polynomial-time reductions, we can now give the definitions for hardness and completeness.

2.4.5. Definition (Hardness and completeness). Let $\mathcal{C}$ be a complexity class of (promise) problems. A (promise) problem A is said to be hard for the class $\mathcal{C}$ under polynomial-time reductions if for every (promise) problem $B \in \mathcal{C}$, $B \leq A$. We say that A is complete for $\mathcal{C}$ if it additionally satisfies that $A \in \mathcal{C}$.

Intuitively, $\mathcal{C}$-complete problems can be viewed as the "hardest" problems contained in the class $\mathcal{C}$. However, it is important to stress that this intuition fails if $\mathcal{C} \subset \mathsf{P}$: by equipping $\mathcal{C}$ with the power to perform polynomial-time reductions, it can solve any problem in P, since any problem in P is polynomial-time reducible to the problem of checking whether a single input bit is 0 or 1. However, this problem does not have to be in $\mathcal{C}$, as $\mathcal{C} \subset \mathsf{P}$. This also implies that all languages in P, except for the empty language and the language of all strings, are P-complete under polynomial-time reductions.[11]

2.4.4 Beyond complexity classes: alternative complexity measures

One of the biggest hurdles in complexity theory is that even some of the most plausible separations cannot be proven with our current mathematical techniques. For example, it is widely believed that $\mathsf{P} \neq \mathsf{PSPACE}$, yet we do not know how to prove this.

However, there are also other ways to compare the power of different models of computation, beyond the time and space we have considered so far, such as query complexity, communication complexity, and sample complexity. In particular, query complexity and sample complexity will play important roles in Chapter 8 and Chapter 9, respectively.

Query complexity. Query complexity measures the number of queries to an oracle (i.e., a black box function that provides answers to specific types of questions) needed to solve a problem. In classical query complexity, one of the most studied settings is that of decision tree complexity: here, the goal is to compute a Boolean function $f : \{0,1\}^n \mapsto \{0,1\}$ using queries to the input. Also, f can be a partial, i.e., $f : \{0,1\}^n \mapsto \{0,1,\perp\}$. Here, '$\perp$' denotes that the function value can be 0 or 1 arbitrarily. The simplest query setting is where, for a given input $x \in \{0,1\}^n$, the value $f(x)$ is returned.

In quantum query algorithms for Boolean functions, there are typically two types of query access to f. In the *standard query model*, the oracle acts as an $n+1$-qubit unitary operator U_f defined in terms of its action on basis states by

$$U_f |x\rangle |a\rangle = |x\rangle |a \oplus f(x)\rangle$$

[11] A more meaningful question is which problems are P-complete under reductions weaker than full polynomial-time computation, such as logspace reductions.

with $a \in \{0,1\}$. Another model is the *phase query*, where f is accessed via an n-qubit unitary operator $U_{f,\text{phase}}$ defined by

$$U_{f,\text{phase}} |x\rangle = (-1)^{f(x)} |x\rangle .$$

Since a phase query can be implemented at the cost of a single standard query[12], we use standard query and phase access interchangeably throughout this dissertation.

It is also possible to consider an inherently quantum model of query complexity, where one has (controlled) black-box access to an n-qubit unitary U and its inverse $U^\dagger$, rather than to a Boolean function. In this setting, U itself is typically treated as the input, and the goal is to decide whether it satisfies a given property using as few queries to U as possible.

Sample complexity. Another complexity measure, frequently studied in a learning context, is *sample complexity*. In this dissertation, we consider the sample complexity of solving tasks when given access to a finite set $S \subseteq \{0,1\}^n$, sampled according to a probability distribution $D : S \to [0,1]$. In this setting, we define a quantum sample from D as access to a single copy of the state[13] [BJ95, AdW17]

$$|S_D\rangle = \sum_{x \in S} \sqrt{D(x)} |x\rangle .$$

Observe that if $|S_D\rangle$ is measured in the computational basis, this is equivalent to obtaining classical samples according to the distribution D.

[12] Observe what happens when U_f is applied to $|x\rangle |-\rangle$.

[13] In [BJ95, AdW17], the setup is slightly different, as the individual samples are of the form $|x, c(x)\rangle$ with $c : \{0,1\}^n \to \{0,1\}$ an unknown function. However, this is entirely analogous to our definition, in the sense that we view a quantum sample as a superposition over all possible classical samples with amplitudes given by the square roots of the probabilities.

Part I.

Low-energy states and their descriptions

Chapter 3

Local Hamiltonians, marginals and ansätze

In this chapter, we lay much of the groundwork for Part I and, to some extent, for Part II. We begin by introducing the local Hamiltonian problem, formalising the task of ground state energy estimation, which is central to this part of the thesis. To establish hardness results for the local Hamiltonian problem (and related problems), the canonical approach is the *Feynman-Kitaev clock construction*, which we describe in Section 3.2. We also present a simplified version of Kitaev's original proof that the 5-local Hamiltonian problem is QMA-hard, developing ideas that will be useful in Chapter 5. In Section 3.3, we take a brief detour to discuss another QMA-complete problem: the *consistency of local density matrices*, which appears frequently throughout this thesis. We give a new proof of containment in QMA that extends the range of parameters for which the problem is known to lie in QMA, improving on previous results [Liu06, BG22]. In Sections 3.4 and 3.5, we introduce *access models* and, associated with them, classes of quantum states that define commonly used *ansätze* for ground state energy estimation. These will play a central role in Chapter 4. Finally, in Section 3.6, we establish a connection between these classes and the ground states of certain Hamiltonians.

3.1 The local Hamiltonian problem

A quantum-mechanical system is characterised by its Hamiltonian H, the operator corresponding to its energy. Throughout this dissertation, we restrict ourselves to Hamiltonians acting on finite-dimensional Hilbert spaces, which ensures that their spectra are finite and thus bounded both from below and above.

An important property of a Hamiltonian is the energy difference between its ground state and first excited state, known as the *spectral gap*, which we denote by $\gamma(H)$. A family of Hamiltonians $\{H_n = H(n) : n \in \mathbb{Z}_+\}$, parametrised

by a notion of system size n (e.g., the number of spins), is said to be *gapped* if $\gamma(H_n) = \Omega(1)$, meaning that the spectral gap remains lower bounded by a constant as $n \to \infty$.

When simulating a quantum-mechanical system governed by a Hamiltonian, we can consider the following general setting, as described by Osborne [Osb12].[1] We are given as input a family of Hamiltonians $\{H_x : x \in X \subseteq \{0,1\}^*\}$, a family of observables $\{O_x : x \in X\}$, an initial state ρ_0, a (possibly complex) time $t \in \mathbb{C}$, and an accuracy parameter $\epsilon > 0$. The task is to compute the expectation values $\mathrm{tr}[O_x \rho_x(t)]$ up to additive precision ϵ, where

$$\rho_x(t) = \frac{\left(e^{-itH_x}\right)^\dagger \rho_0 e^{-itH_x}}{\mathrm{tr}\left[\left(e^{-itH_x}\right)^\dagger \rho_0 e^{-itH_x}\right]}.$$

When $t \in \mathbb{R}$, ρ_0 is a simple state that can be efficiently prepared on a quantum computer, and the observables O_x can be efficiently measured, the above task corresponds to what is commonly referred to as *Hamiltonian simulation*, which is known to be universal for quantum computation [Llo96]. When $t = -i\beta/2$ and $\rho_0 = \mathbb{I}/\mathrm{tr}[\mathbb{I}]$, the state $\rho_x(t)$ corresponds to a thermal state at inverse temperature $\beta > 0$. In the limit $\beta \to \infty$, $\rho_x(t)$ approaches the so-called *ground state* ρ_{gs}, i.e., the eigenstate of lowest energy. Similarly, by letting $\rho_0 = (\mathbb{I} - \Pi_{\mathrm{gs}})/\mathrm{tr}[\mathbb{I} - \Pi_{\mathrm{gs}}]$, where Π_{gs} is the projector onto the subspace spanned by all ground states, the same limit yields the first excited state. In the context of energy estimation, the observable of interest is simply the Hamiltonian H_x itself.

All physical systems we consider throughout this thesis will be defined in terms of qubits and have the property of being *local*.

3.1.1. Definition (Local operators). We say that an n-qubit Hermitian operator M is k-local if and only if it can be expressed as

$$M = P_\pi (A \otimes \mathbb{I}) P_\pi^{-1},$$

for an arbitrary k-qubit Hermitian matrix A, where P_π is a permutation matrix defined by

$$P_\pi |x_1 \cdots x_n\rangle = |x_{\pi(1)} \cdots x_{\pi(n)}\rangle.$$

Often, we will only specify A and use it in place of M when the identities are clear from the context, and the specification of the permutation is not relevant for our purposes.

Given a physical system, one of the primary tasks is to compute its ground state energy. For systems composed of qubits, this task can be formalised as the following promise problem:

[1] We slightly modify the definition given in [Osb12], as we do not yet need to formally present the task as a collection of promise problems.

3.1.2. Definition (Local Hamiltonian problem, LH(k, a, b)).

Input: A classical description of a collection of k-local Hermitian operators $\{H_i\}_{i\in[m]}$, with $\|H_i\| \leq 1$ for all $i \in [m]$, which define an n-qubit k-local Hamiltonian $H = \sum_{i=1}^{m} H_i$, where $m = \text{poly}(n)$, and two efficiently computable real numbers a, b such that $b > a$.

Promise: Either $\lambda_0(H) \leq a$ or $\lambda_0(H) \geq b$ holds.

Output: YES when $\lambda_0(H) \leq a$, and NO when $\lambda_0(H) \geq b$.

3.1.3. Remark. Moreover, without loss of generality, each term H_i can be assumed to satisfy $0 \preceq H_i \preceq \mathbb{I}$, since this can be manually enforced by adding an identity matrix and rescaling each term by a factor of $1/2$, which decreases the relative promise gap (which is relative to the total number of terms) by a factor $1/2$.

In Chapter 1, we briefly touched upon the importance of the local Hamiltonian problem in practical applications. In the following two subsections, we substantiate this by providing additional motivation from the perspectives of both physics and theoretical computer science.

3.1.1 Motivation from physics

In many physical systems, the ground state is interesting primarily for two reasons: (i) most quantum phenomena only exhibit themselves at low temperature; and (ii) in some systems the ground state actually forms a very good approximation to any thermal state at low temperature, where "low" can even mean at room temperature.[2] For (i), a key example is that of superconductivity: at standard atmospheric pressure, cuprates currently hold the temperature record—manifesting superconductivity up to 138 K [DCS+95], far below a room temperature of $T = 298$ K. Point (ii) is particularly common for metals, where the Fermi temperature defines the temperature scale at which excited states become important, which is typically when the temperature is of the order of $T \sim 10^4$ K or higher. This means that a metal at room temperature is effectively in its ground state, with only a few excitations governed by Fermi-Dirac statistics.

We can give a particularly simple example of (ii) by considering arguably the simplest possible quantum system: a single two-level particle in a magnetic field, described by the Hamiltonian

$$H = -BZ,$$

[2]Regarding Point (i), it was recently proven that for a local Hamiltonian H, all thermal states with $\beta \geq \beta_0$, where $\beta_0 > 0$ is some constant, are separable [BLMT24]. However, the actual value of β_0 (and the corresponding temperature) may be far lower (higher) than what is typically encountered in practice.

where $B \geq 0$ is the magnetic field strength and Z is the Pauli Z-matrix. The eigenvalues of H are given by $E_\pm = \pm B$, giving a partition function of

$$\mathcal{Z} = \sum_i e^{-\beta E_i} = e^{\beta B} + e^{-\beta B}.$$

Here $\beta = 1/k_B T$ is the inverse temperature with $k_B \approx 8.617 \times 10^{-5}$ eV/K being Boltzmann's constant in electron volts per Kelvin. The next argument is an example of why—unlike theoretical computer scientists—physicists usually do have to worry about the actual numerical values in their expressions. The probability of the system being in the ground state $E_- = -B$ at temperature T is then

$$\frac{e^{\beta B}}{\mathcal{Z}} = \frac{e^{\beta B}}{e^{\beta B} + e^{-\beta B}} = \frac{1}{1 + e^{-2\beta B}}.$$

For large βB, this probability approaches 1, meaning that the system remains in its ground state with high probability at low temperatures. For example, if B is on the order of 1 eV, then at room temperature ($T = 298$ K), we find

$$\frac{1}{1 + e^{-2\beta B}} \approx 1 - 1.5 \times 10^{-34},$$

indicating that the system is overwhelmingly in its ground state. Of course, most physical systems of interest have much more complex spectra, and whether the ground state provides a good approximation to a thermal state depends on the details of the energy level distribution.

Now that we have some motivation for why ground states occupy an important place among all eigenstates, we still need to explain why the corresponding energy value is significant in the first place. Again, we consider the setting where the family of Hamiltonians we are interested in can be viewed as a parametrised Hamiltonian, i.e., $H_x = H(x)$ with $x \in X \subseteq \{0,1\}^*$. A simple example of this is the hydrogen molecular ion H_2^+, whose electronic structure Hamiltonian in the Born–Oppenheimer approximation[3] can be parametrised as $H(x)$, where x is a bit string representing the distance between the two protons [AF11]. Being able to compute the ground state energy for each $x \in X$ (each of which represents some value of r) allows us to identify the most stable configuration. In this way, ground state energy computations provide insight into the geometry of the molecule. For many industrially relevant quantities in quantum chemistry—such as reaction rates, binding energies, and molecular pathways—the simulation problem reduces to computing the ground state energy as a function of the nuclear coordinates of a molecular system [CRO+19].

[3] In its standard form, this Hamiltonian is not defined in terms of qubits as in Definition 3.1.2, as it models the interactions between electrons and atomic nuclei. However, it can be transformed into a related qubit Hamiltonian with the same spectral properties using a fermion-to-qubit mapping such as the Jordan–Wigner transformation [JW93].

3.1.2 Motivation from theoretical computer science

The local Hamiltonian problem can be seen as a quantum generalisation of the canonical NP-complete *Constraint Satisfaction Problem* (CSP). Consider a collection of m clauses $\{C_i\}_{i\in[m]}$, where each clause $C_i : \{0,1\}^k \to \{0,1\}$ assigns a 0/1 value to a setting of its k input bits. Each clause is associated with a subset $S_i \subset [n]$ of k out of n variables, and we say that a global assignment $x \in \{0,1\}^n$ satisfies C_i if the restriction of x to C_i, denoted x_{S_i}, is accepted by C_i. The task of determining whether there exists an x such that all clauses are satisfied, i.e., $C_i(x_{S_i}) = 1$ for all $i \in [m]$, or whether for every x there exists some $i \in [m]$ for which $C_i(x_{S_i}) = 0$, can be framed as a local Hamiltonian problem. To do this, one considers the k-local terms

$$H_i = \sum_{y\in\{0,1\}^k : C_i(y)=0} |y\rangle\langle y| \,,$$

which assign an energy penalty of 1 to each $|y\rangle$ that does not satisfy clause C_i. These act non-trivially only on the k qubits corresponding to the variables involved in C_i. The full n-qubit Hamiltonian is then given by $H = \sum_{i\in[m]} H_i$, where each H_i is implicitly tensored with identities on the other qubits and permuted to act on the qubits with labels S_i (see Definition 3.1.1). The ground state energy of H is 0 if all clauses can be satisfied, and at least 1 otherwise.

More importantly, just as k-CSP problems are considered canonical NP-complete problems, the k-local Hamiltonian problem serves as the canonical complete problem for the quantum analogue of NP, namely QMA [KSV02], which will be the topic of the next section.

3.2 QMA-completeness: tick-tock goes the clock

In this section, we briefly overview the ideas behind the *Feynman-Kitaev circuit-to-Hamiltonian mapping* and some of its variations, which will serve as key techniques in the next two chapters. We aim to be as concise as possible while still introducing the most important ideas and notation, and refer the reader to [KSV02, Gha12] for more comprehensive treatments.

3.2.1 Circuit-to-Hamiltonian mappings

Let $A = (A_{\text{YES}}, A_{\text{NO}})$ be an arbitrary promise problem in QMA, and let $\mathcal{U} = \{U_n : n \in \mathbb{N}\}$ be the corresponding P-uniform family of QMA verifiers. Without loss of generality, we assume that each $U_n \in \mathcal{U}$ consists of only two-qubit gates from some fixed universal gate set $\mathcal{G}$, and has completeness $1 - \alpha(n)$ and soundness $\alpha(n)$ for any choice of $\alpha(n) \geq 1/\exp(n)$. Fix some input size n. With a slight abuse of notation, we omit the subscript and refer to the verification circuit for input size n simply as U.

Let $T = \text{poly}(n)$ denote the number of elementary gates in U, so that $U = V_T \dots V_1$, where each V_t is a two-qubit gate from $\mathcal{G}$ tensored with identities. Following standard convention, we set $V_0 \coloneqq \mathbb{I}$. The Hilbert space on which U acts, denoted by $\mathcal{H}_{\text{comp}}$, can be decomposed as

$$\mathcal{H}_{\text{comp}} = \mathcal{H}_{\text{input}} \otimes \mathcal{H}_{\text{proof}} \otimes \mathcal{H}_{\text{ancilla}},$$

where $\mathcal{H}_{\text{input}} \coloneqq \mathbb{C}^{2^n}$ is the input register containing the input state $|x\rangle$, $\mathcal{H}_{\text{proof}} \coloneqq \mathbb{C}^{2^{p(n)}}$ is the proof register containing a $p(n)$-qubit witness, and $\mathcal{H}_{\text{ancilla}} \coloneqq \mathbb{C}^{2^m}$ is the ancilla register initialized in $|0^m\rangle$, with $m = \text{poly}(n)$.

We first consider the general setting of Kitaev's original circuit-to-Hamiltonian mapping [KSV02], which acts on the Hilbert space

$$\mathcal{H}_{\text{comp}} \otimes \mathcal{H}_{\text{clock}},$$

for some Hilbert space $\mathcal{H}_{\text{clock}}$. We do not specify the dimension of $\mathcal{H}_{\text{clock}}$ at this stage, as it depends on the particular construction being used. We adopt the shorthand "out" for the output qubit with associated subspace $\mathcal{H}_{\text{out}} \subseteq \mathcal{H}_{\text{comp}}$.

The Hamiltonian corresponding to the circuit-to-Hamiltonian mapping is then given by

$$H_{\text{FK}} = H_{\text{in}} + H_{\text{prop}} + H_{\text{out}}, \tag{3.1}$$

with $H_{\text{in}} \in \text{Herm}(\mathcal{H}_{\text{input}} \otimes \mathcal{H}_{\text{ancilla}} \otimes \mathcal{H}_{\text{clock}})$, $H_{\text{prop}} \in \text{Herm}(\mathcal{H}_{\text{comp}} \otimes \mathcal{H}_{\text{clock}})$ and $H_{\text{out}} \in \text{Herm}(\mathcal{H}_{\text{out}} \otimes \mathcal{H}_{\text{clock}})$, given by (omitting some identity matrices)

$$\begin{aligned}
H_{\text{in}} &\coloneqq \left(\underbrace{\sum_{i=1}^{n} (\mathbb{I} - |x\rangle\langle x|_i)}_{\text{acts on } \mathcal{H}_{\text{input}}} + \underbrace{\sum_{j=1}^{m} (\mathbb{I} - |0\rangle\langle 0|_j)}_{\text{acts on } \mathcal{H}_{\text{ancilla}}} \right) \otimes \underbrace{|0\rangle\langle 0|}_{\text{acts on } \mathcal{H}_{\text{clock}}}, \\
H_{\text{prop}} &\coloneqq \sum_{t=1}^{T} H_t \qquad \text{where} \\
H_t &\coloneqq -\frac{1}{2} \underbrace{V_t}_{\text{acts on } \mathcal{H}_{\text{comp}}} \otimes \underbrace{|t\rangle\langle t-1|}_{\text{acts on } \mathcal{H}_{\text{clock}}} - \frac{1}{2} \underbrace{V_t^\dagger}_{\text{acts on } \mathcal{H}_{\text{comp}}} \otimes \underbrace{|t-1\rangle\langle t|}_{\text{acts on } \mathcal{H}_{\text{clock}}} \\
&\quad + \frac{1}{2} \underbrace{\mathbb{I}}_{\text{acts on } \mathcal{H}_{\text{comp}}} \otimes \underbrace{|t\rangle\langle t|}_{\text{acts on } \mathcal{H}_{\text{clock}}} + \frac{1}{2} \underbrace{\mathbb{I}}_{\text{acts on } \mathcal{H}_{\text{comp}}} \otimes \underbrace{|t-1\rangle\langle t-1|}_{\text{acts on } \mathcal{H}_{\text{clock}}}, \\
H_{\text{out}} &\coloneqq \underbrace{|0\rangle\langle 0|}_{\text{acts on } \mathcal{H}_{\text{out}}} \otimes \underbrace{|T\rangle\langle T|}_{\text{acts on } \mathcal{H}_{\text{clock}}}
\end{aligned} \tag{3.2}$$

Note that all local terms are projectors, so the operator norm condition of Definition 3.1.2 is satisfied. Moreover, observe that $H_{\text{FK}} = H_{\text{FK}}(x)$, since $H_{\text{in}} = H_{\text{in}}(x)$

depends on the input x, which means that we associate a different Hamiltonian with each input x. From now on, we omit register subscripts whenever their meaning is clear from context.

Let $H_0 \coloneqq H_{\mathrm{FK}} - H_{\mathrm{out}}$ be the Hamiltonian that is just as H_{FK} but without the H_{out} term. The family of *history states*, parametrized by proofs $\{|\psi\rangle : |\psi\rangle \in \mathcal{H}_{\mathrm{proof}}\}$, is given by

$$|\eta(|\psi\rangle)\rangle = \frac{1}{\sqrt{T+1}} \sum_{t=0}^{T} V_t \dots V_1 |x\rangle |\psi\rangle |0^m\rangle |t\rangle . \tag{3.3}$$

These history states span the ground space of H_0 with ground state energy 0 [KSV02]. It is easily verified that if U accepts $(x, |\psi\rangle)$ with probability p, then we have that the corresponding history state of Eq. (3.3) has energy

$$\langle \eta(|\psi\rangle)| H_{\mathrm{FK}} |\eta(|\psi\rangle)\rangle = \frac{1-p}{T+1}. \tag{3.4}$$

Moreover, by linearity, we have that for any two quantum states $|\psi_1\rangle, |\psi_2\rangle$ and complex numbers $\alpha_1, \alpha_2 \in \mathbb{C}$ satisfying $|\alpha_1|^2 + |\alpha_2|^2 = 1$, it holds that

$$\alpha_1 |\eta(|\psi_1\rangle)\rangle + \alpha_2 |\eta(|\psi_2\rangle)\rangle = |\eta(\alpha_1 |\psi_1\rangle + \alpha_2 |\psi_2\rangle)\rangle ,$$

from which it follows that any linear combination of history states is itself a history state. A useful property of H_0 is that its spectral gap can be lower bounded in terms of T, which is implicitly proven in [KSV02].

3.2.1. Lemma (Adapted from [KSV02, Chapter 15]). *H_0 has a spectral gap of $\Delta \geq \frac{1}{(T+1)^2}$.*

If we use a binary representation of integers to represent the clock space, $|t\rangle = |\mathrm{rep}(t)\rangle$ with $\mathrm{rep}(\cdot) : \mathbb{N} \to \{0,1\}^*$, then we have that $\mathcal{H}_{\mathrm{clock}} \coloneqq \mathbb{C}^{T+1}$ and H_{FK} becomes a $\mathcal{O}(\log n)$ local Hamiltonian (recall that $T = \mathrm{poly}(n)$). We will denote the entire space the Hamiltonian acts upon when using this specific representation of the clock $\mathcal{S} \coloneqq \mathcal{H}_{\mathrm{comp}} \otimes \mathbb{C}^{T+1}$, and will later see that $\mathcal{S}$ can be embedded as a subspace into a larger space to achieve a constant locality for the resulting Hamiltonian.

In [KSV02], QMA-hardness is established by proving the following statement:

3.2.2. Theorem (Adapted from [KSV02, Chapter 15]). *The Hamiltonian H_{FK} as per Eq.* (3.1) *satisfies*

- *If $x \in A_{\mathrm{YES}}$, then its ground state satisfies $\lambda_0(H_{\mathrm{FK}}) \leq \alpha/(T+1)$;*
- *If $x \in A_{\mathrm{NO}}$, then its ground state satisfies $\lambda_0(H_{\mathrm{FK}}) \geq \Omega((1-\sqrt{\alpha})/T^3)$.*

Hence, when $\alpha(n)$ was chosen sufficiently small, the promise gap becomes strictly positive and is lower bounded by $\Omega(1/T^3)$.[4]

Having established some basic properties of this general construction, our goal is now to explore variations that offer improvements—particularly in terms of locality, improving from $k = \mathcal{O}(\log n)$ to $k = \mathcal{O}(1)$. Generally, this is achieved by developing more efficient representations of the "clock component" of the construction. This leads to new Hamiltonians H, but all those we consider include a term H_{out}, allowing us to define $H_0 \coloneqq H - H_{\text{out}}$, just as we did for H_{FK} in Eq. (3.1). As we will see next, a simple modification presented in [Kit95] allows us to make the construction 5-local.

Kitaev's 5-local construction. We use a unary encoding to represent the clock space, which means that $\mathcal{H}_{\text{clock,new}} \coloneqq \mathbb{C}^{2^T}$. Consequently, the resulting Hamiltonian will act upon a larger space than before, and we associate an injective linear map (in fact, an isometry):

$$\varphi : \mathcal{H}_{\text{comp}} \otimes \mathcal{H}_{\text{clock}} \longrightarrow \mathcal{H}_{\text{comp}} \otimes \mathcal{H}_{\text{clock,new}},$$

that maps states from the old construction to the new one. We write $\hat{t}$ to denote the unary representation of the time step t in the computation, i.e.,

$$\left|\hat{t}\right\rangle = |\underbrace{1\ldots1}_{t}\underbrace{0\ldots0}_{T-t}\rangle,$$

so that the history states can be written as

$$|\eta(|\psi\rangle)\rangle = \frac{1}{\sqrt{T+1}} \sum_{t=0}^{T} V_t \ldots V_1 |x\rangle |\psi\rangle |0\ldots0\rangle \left|\hat{t}\right\rangle.$$

We take φ to be the map that transforms $|t\rangle \mapsto \left|\hat{t}\right\rangle$ for $t \in \{0, 1, \ldots, T\}$ and acts on $\mathcal{H}_{\text{comp}}$ as the identity operator. We now define a new representation of the clock operators in the unary-encoded space. This transformation of the clock operators can be understood as defining a map

$$\Phi : \mathrm{L}(\mathcal{H}_{\text{comp}} \otimes \mathcal{H}_{\text{clock}}) \longrightarrow \mathrm{L}(\mathcal{H}_{\text{comp}} \otimes \mathcal{H}_{\text{clock,new}}),$$

which maps each logical clock operator to a new operator acting on the unary-encoded space. As we are only interested in how Φ transforms the operators in

[4] Choosing a smaller value of $\alpha(n)$ implicitly makes T larger, but it will remain polynomial as long as $\alpha(n) \geq 1/\exp(n)$.

H_{FK}, the following specification of the transformation suffices:

$$\begin{aligned}
|0\rangle\langle 0| &\mapsto |0\rangle\langle 0|_1 \,, \\
|0\rangle\langle 1| &\mapsto |0\rangle\langle 1|_1 \otimes |0\rangle\langle 0|_2 \,, \\
|t\rangle\langle t| &\mapsto |1\rangle\langle 1|_t \otimes |0\rangle\langle 1|_{t+1} \,, \\
|t-1\rangle\langle t| &\mapsto |1\rangle\langle 1|_{t-1} \otimes |0\rangle\langle 1|_t \otimes |0\rangle\langle 0|_{t+1} \,, \\
|T-1\rangle\langle T| &\mapsto |1\rangle\langle 1|_{T-1} \otimes |0\rangle\langle 1|_T \,, \\
|T\rangle\langle T| &\mapsto |1\rangle\langle 1|_T \,.
\end{aligned}$$

We let $\bar{H}_{\mathrm{FK}} = \Phi\left(H_{\mathrm{FK}}\right)$, and consider the Hamiltonian

$$H_{\mathrm{FK},5} = \bar{H}_{\mathrm{FK}} + H_{\mathrm{stab}}, \tag{3.5}$$

with

$$H_{\mathrm{stab}} \coloneqq \sum_{t=1}^{T-1} |0\rangle\langle 0|_t \otimes |1\rangle\langle 1|_{t+1} \,,$$

so that $H_{\mathrm{stab}} \left|\hat{t}\right\rangle = 0$ for all $t \in \{0, \ldots, T\}$. Note that, when represented using qubits, all the newly defined operators on $\mathcal{H}_{\mathrm{clock,new}}$ are at most 3-local. Since we we use a 2-local universal gate set $\mathcal{G}$, this means that $H_{\mathrm{FK},5}$ is indeed 5-local.

Finally, we let $\mathcal{L} \subseteq \mathcal{H}_{\mathrm{comp}} \otimes \mathcal{H}_{\mathrm{clock,new}}$ be the image of $\mathcal{S}$ under φ, i.e., $\varphi(\mathcal{S}) = \mathcal{L}$, and define $\mathcal{L}^\perp$ such that $\mathcal{L} \oplus \mathcal{L}^\perp = \mathcal{H}_{\mathrm{comp}} \otimes \mathcal{H}_{\mathrm{clock,new}}$. In other words, the subspace $\mathcal{S} \subseteq \mathcal{H}_{\mathrm{comp}} \otimes \mathcal{H}_{\mathrm{clock}}$ is mapped isomorphically onto the subspace $\mathcal{L} \subseteq \mathcal{H}_{\mathrm{comp}} \otimes \mathcal{H}_{\mathrm{clock,new}}$. Note that $H_{\mathrm{FK},5}$ acts invariantly on $\mathcal{L}$ (and thus also on $\mathcal{L}^\perp$, as it is Hermitian), as it never mixes valid and invalid clock strings.

We now have everything in place to explain why going from H_{FK} to $H_{\mathrm{FK},5}$ improves the locality whilst preserving the desired spectral properties. The key idea of the construction is that, restricted to $\mathcal{L}$, the new Hamiltonian $H_{\mathrm{FK},5}$ acts in exactly the same way as H_{FK} acted on $\mathcal{S}$ before, whilst states in $\mathcal{L}^\perp$ are given an additional energy penalty. Thus, for any state $|\psi\rangle \in \mathcal{S}$, we have $H_{\mathrm{FK},5}\, \varphi(|\psi\rangle) = \varphi\left(H_{\mathrm{FK}} |\psi\rangle\right)$, and we have that completeness follows immediately. For soundness, we will use that $H_{\mathrm{FK},5}$ acts invariantly on $\mathcal{L}$. When $H_{\mathrm{FK},5}$ is restricted to $\mathcal{L}^\perp$, the added term H_{stab} ensures that $H_{\mathrm{FK},5}\big|_{\mathcal{L}^\perp} \succeq 1$. From this we can conclude that the subspace of all eigenstates having energy < 1 has no intersection with $\mathcal{L}^\perp$, and we have that in this low-energy subspace, $H_{\mathrm{FK},5}$ also acts exactly like H_{FK}.[5] Hence, Theorem 3.2.2 and Lemma 3.2.1 still apply even

[5]For a short proof of this general fact: let $A, B \in \mathrm{Herm}(\mathcal{H})$, and let $\mathcal{L} \subseteq \mathcal{H}$ be a subspace invariant under both A and B, so $\mathcal{L}^\perp$ is also invariant under both. Suppose that $B|_{\mathcal{L}} = 0$, $B|_{\mathcal{L}^\perp} \succ \alpha > 0$, and $A \succeq 0$. Let $|\psi\rangle$ be a normalised eigenvector of $A+B$ with eigenvalue $\lambda < \alpha$, and write $|\psi\rangle = a\,|\phi\rangle + b\left|\phi^\perp\right\rangle$, with $|\phi\rangle \in \mathcal{L}$, $\left|\phi^\perp\right\rangle \in \mathcal{L}^\perp$, and $|a|^2 + |b|^2 = 1$. Since both A and B preserve $\mathcal{L}$ and $\mathcal{L}^\perp$, their sum does as well. Thus, $\lambda(a\,|\phi\rangle + b\left|\phi^\perp\right\rangle) = (A+B)(a\,|\phi\rangle + b\left|\phi^\perp\right\rangle) = aA\,|\phi\rangle + b(A+B)\left|\phi^\perp\right\rangle$, using that $B\,|\phi\rangle = 0$. Matching the $\mathcal{L}^\perp$-components gives $(A+B)b\left|\phi^\perp\right\rangle = \lambda b\left|\phi^\perp\right\rangle$. But $A \succeq 0$ and $B|_{\mathcal{L}^\perp} \succ \alpha$, so $\left\langle\phi^\perp\right|(A+B)\left|\phi^\perp\right\rangle > \alpha$, contradicting $\lambda < \alpha$ unless $b = 0$. Thus $|\psi\rangle \in \mathcal{L}$ and is an eigenvector of A.

in this new representation with H_{stab} added.[6]

Improved constructions. Further improvements have reduced the Hamiltonian to 3-local [KR03] and ultimately to 2-local [KKR06], which the smallest locality for which one can obtain QMA-hardness since the 1-local Hamiltonian problem is trivially in P (the ground state is simply the tensor product of the ground states of each 1-local term). We will not be interested in the precise details of these constructions, as we will only use them in a black-box fashion throughout the rest of the dissertation. We will denote the 3-local Hamiltonian of [KR03] by $H_{\text{FK},3}$.

The small-penalty clock construction. The next modification to the original construction that we consider involves changing the relative weighting of the H_{out} term compared to the H_0 component, a technique known as the *small-penalty clock construction* [DGF22]. In this construction, a small penalty $\epsilon > 0$ is applied to the H_{out} term, yielding the Hamiltonian

$$H_{\text{SPCC}} = H_{\text{in}} + H_{\text{prop}} + \epsilon H_{\text{out}}. \tag{3.6}$$

We will usually write $H_{\text{SPCC},k}$ to indicate a small-penalty construction Hamiltonian which uses a k-local circuit-to-Hamiltonian mapping, where the value k depends on the used construction (which will be made explicit in the text).

The core contribution of [DGF22] is the use of tools from the Schrieffer-Wolff transformation to derive precise bounds on the intervals in which the low-energy eigenvalues must lie. This provides control over the relationship between the acceptance probabilities of the circuit and the low-energy subspace of the Hamiltonian. The following lemma, adapted from [DGF22], will serve as a key technique in Chapters 4 and 5 of this thesis.

3.2.3. Lemma (Small-penalty clock construction [DGF22, Lemma 26]). *Let $\mathcal{U} = \{U_n : n \in \mathbb{N}\}$ be a* P*-uniform family of* QMA *verification circuits. Let n be the input size and consider an input $x \in \{0,1\}^n$. Suppose U_n consists of $T = \text{poly}(n)$ gates from some universal gate-set using at most 2-local gates. Denote $P(\psi)$ for the probability that U_n accepts $(x, |\psi\rangle)$, and let $H_{\text{SPCC},3}$ be the corresponding 3-local Hamiltonian from the circuit-to-Hamiltonian mapping in [KR03] with an ϵ-factor in front of H_{out}. Then there exists a constant $c > 0$, such that for all $0 < \epsilon \leq c/T^3$, we have that within the low-energy subspace $\mathcal{S}_\epsilon$ of $H_{\text{SPCC},3}$, i.e.,*

$$\mathcal{S}_\epsilon = \text{span}\{|\Phi\rangle : \langle\Phi| H_{\text{SPCC},3} |\Phi\rangle \leq \epsilon\}$$

[6] This generally holds for any type of clock construction that has the property that its low-energy subspace coincides with that of H_{FK}.

the eigenvalues λ_i *satisfy*

$$\lambda_i \in \left[\epsilon\frac{1-P(|\psi_i\rangle)}{T+1} - \mathcal{O}(T^3\epsilon^2), \epsilon\frac{1-P(|\psi_i\rangle)}{T+1} + \mathcal{O}(T^3\epsilon^2)\right], \tag{3.7}$$

where $\{|\psi_i\rangle\}$ *are the eigenstates of the Mariott-Watrous operator of the circuit* U_n*, given by*

$$Q_n = \left(\langle x| \otimes \mathbb{I} \otimes \langle 0|^{\otimes q(n)}\right) U_n^\dagger \left(|0\rangle\langle 0| \otimes \mathbb{I}\right) U_n \left(|x\rangle \otimes \mathbb{I} \otimes |0\rangle^{\otimes q(n)}\right).$$

3.2.4. Remark. Different circuit-to-Hamiltonian constructions typically rely on using different forms of H_{stab} and the encoding of the clock space. As long the history states span all 0-eigenstates of H_0, and the smallest non-zero eigenvalue of H_{stab} is at least $1/4$, and the entire Hamiltonian acts invariantly on the subspace with the "correct" clock states, the proof of Lemma 3.2.3 in [DGF22] should translate to any such "Feynman-Kitaev"-type construction. In particular, the lemma can be reformulated in terms of the standard 5-local circuit-to-Hamiltonian mapping described in Section 3.2.1.

3.2.2 QMA-hardness: a simple proof

We now return to the 5-local construction of Section 3.2.1 and combine it with the idea of the small-penalty factor to give a simple proof that the 5-local Hamiltonian problem is QMA-hard for some $b - a = 1/\text{poly}(n)$. Kitaev's original proof relies on a lemma that lower bounds the smallest eigenvalue of the sum of two positive semi-definite operators A_1 and A_2, which is shown to depend on the angle between their respective null spaces [KSV02]. As we observed in the previous subsection, Theorem 3.2.2 only gives a strictly positive (and thus valid) promise gap if error reduction is applied to the QMA circuit being reduced from, which is slightly unsatisfactory, as it means that the circuit-to-Hamiltonian mapping cannot be directly applied to an arbitrary QMA verifier.

Before we proceed, a small remark: the small-penalty clock construction of Lemma 3.2.3 can in fact be directly applied to show that the 3-local Hamiltonian problem is QMA-hard (even without applying error reduction to the QMA verifier). In the YES-case, there exists a quantum proof $|\psi_i\rangle$ such that $P(|\psi_i\rangle) \geq 2/3$, whereas in the NO-case we have that $P(|\psi_i\rangle) \leq 1/3$ for all quantum proofs $|\psi_i\rangle$. Hence, picking $\epsilon \coloneqq cT^{-4}$ for some sufficiently small $c > 0$ results in a strictly positive promise gap of $\Omega(T^{-5}) = 1/\text{poly}(n)$. However, making the proof fully self-contained would require proving Lemma 3.2.3, which is arguably more involved than returning to Kitaev's original proof.

We will show that one can also use the small-penalty factor to give a direct proof that does not rely on Lemma 3.2.3. The ideas developed here will also be important in Chapter 5.

We first prove the following lemma, which shows that states with sufficiently low energy with respect to the Hamiltonian of Eq. (3.6) must have a large fidelity with the space spanned by history states.

3.2.5. Lemma. *Let $|\Psi\rangle$ be a state such that $\langle\Psi| H_{\mathrm{FK},5} |\Psi\rangle \leq \delta$ and let Δ be the spectral gap of H_0. Write Π_{hist} for the projector onto the subspace spanned by history states. Then,*

$$\|\Pi_{\mathrm{hist}} |\Psi\rangle\|^2 \geq 1 - \frac{\delta}{\Delta}.$$

Proof:
Recall that $H_0 \coloneqq H_{\mathrm{FK},5} - H_{\mathrm{out}}$. Since H_{out} is positive semi-definite, we have that $\langle\Psi| H_0 |\Psi\rangle \leq \delta$ must hold as well. Let $\{|\psi_i\rangle\}$ be the eigenbasis of H_0, which consists of history states (with energy 0) and non-history states (with energy at least Δ by Lemma 3.2.1). We can write $H_0 = A_0 + A_{\geq\Delta}$, where A_0 are all the terms in the spectral decomposition of H with eigenvalues exactly zero and $A_{\geq\Delta}$ those with eigenvalues $\geq \Delta$. We have

$$\begin{aligned}
\delta &\geq \langle\Psi| H_{\mathrm{FK},5} |\Psi\rangle \\
&\geq \langle\Psi| H_0 |\Psi\rangle \\
&\geq \langle\Psi| A_0 |\Psi\rangle + \langle\Psi| A_{\geq\Delta} |\Psi\rangle \\
&\geq 0 + \langle\Psi| \sum_{i:\lambda_i\geq\Delta} \lambda_i |\psi_i\rangle\langle\psi_i| |\Psi\rangle \\
&\geq \Delta \langle\Psi| \sum_{i:\lambda_i\geq\Delta} |\psi_i\rangle\langle\psi_i| |\Psi\rangle \\
&\geq \Delta \langle\Psi| (\mathbb{I} - \Pi_{\mathrm{hist}}) |\Psi\rangle \\
&= \Delta \left(1 - \|\Pi_{\mathrm{hist}} |\Psi\rangle\|^2\right),
\end{aligned}$$

where we used that the history states span the ground space of H_0. The statement follows directly from rearranging the inequality. □

It can easily be checked that Lemma 3.2.5 also holds in the case when we use H_{SPCC}, $H_{\mathrm{SPCC},3}$ or $H_{\mathrm{SPCC},5}$, as we only require the property that H_{out} is positive semidefinite and that Lemma 3.2.1 holds for the used construction.

Having Lemma 3.2.5 in hand, it becomes easy to pick an $\epsilon > 0$ to show QMA-hardness for the 5-local Hamiltonian problem.

3.2.6. Theorem. *Let $A = (A_{\mathrm{YES}}, A_{\mathrm{NO}})$ be any promise problem in* QMA*, let $\mathcal{U} = \{U_n : n \in \mathbb{N}\}$ be a corresponding* P*-uniform family of* QMA *verifiers with completeness c and soundness s such that $c - s \geq 1/\mathrm{poly}(n)$. Then there exists a polynomial-time reduction from the corresponding circuit verification problem to a 5-local Hamiltonian problem with inverse polynomial promise gap that does not use error reduction on the circuit.*

Proof:
We will analyse a Hamiltonian $H_{\mathrm{SPCC},5}$ of the form of Eq. (3.6), which is just as $H_{\mathrm{FK},5}$ but with a small penalty ϵ in front of H_{out}. Clearly, the Hamiltonian $H_{\mathrm{SPCC},5}$ can be constructed in polynomial time from a description of U_n and x, since each term is efficiently computable. We will specify ϵ later, but assume for now that it is lower bounded by some polynomial in c,s and $1/T$ and that it satisfies $\epsilon \leq \Delta$, where Δ is the spectral gap of $H_0 = H_{\mathrm{SPCC},5} - \epsilon H_{\mathrm{out}}$. When $x \in A_{\mathrm{YES}}$, we have that there exists a history state with energy at most $\epsilon(1-c)/(T+1)$. Now for $x \in A_{\mathrm{NO}}$. We can write any state $|\Psi\rangle$ in the eigenbasis of H_0 as $|\Psi\rangle = \beta\,|\mathrm{hist}\rangle + \sqrt{1-\beta^2}|\mathrm{hist}^{\perp}\rangle$, for some real $\beta \in [0,1]$, where $|\mathrm{hist}\rangle$ lives in the space spanned by the history states and $|\mathrm{hist}^{\perp}\rangle$ in the space orthogonal to it. In its eigenbasis, H_0 is diagonal. We consider the following two cases:

Case (i): $\langle\Psi|\, H_{\mathrm{SPCC},5}\,|\Psi\rangle > \epsilon$ for all states $|\Psi\rangle$. In this case, the promise gap is simply lower bounded by

$$\epsilon\left(1 - \frac{1-c}{T+1}\right) = \Omega(\epsilon),$$

which is inverse polynomial in n when ϵ is some inverse polynomial in n.

Case (ii): There exists a state $|\Psi\rangle$ such that $\langle\Psi|\, H_{\mathrm{SPCC},5}\,|\Psi\rangle \leq \epsilon$. By Lemma 3.2.5, setting $\delta \coloneqq \epsilon$, we have that for such a $|\Psi\rangle$ it must hold that

$$\beta \geq \sqrt{1 - \frac{\epsilon}{\Delta}},$$

provided $\epsilon \leq \Delta$ (which holds by assumption). Therefore,

$$\begin{aligned}
\langle\Psi|\, H_{\mathrm{SPCC},5}\,|\Psi\rangle &= \left(\beta\,\langle\mathrm{hist}| + \sqrt{1-\beta^2}\langle\mathrm{hist}^{\perp}|\right) H_{\mathrm{SPCC},5}\left(\beta\,|\mathrm{hist}\rangle + \sqrt{1-\beta^2}|\mathrm{hist}^{\perp}\rangle\right) \\
&= \beta^2\,\langle\mathrm{hist}|\, H_{\mathrm{SPCC},5}\,|\mathrm{hist}\rangle + \beta\sqrt{1-\beta^2}\,\langle\mathrm{hist}|\, H_{\mathrm{SPCC},5}|\mathrm{hist}^{\perp}\rangle \\
&\quad + \beta\sqrt{1-\beta^2}\langle\mathrm{hist}^{\perp}|H_{\mathrm{SPCC},5}\,|\mathrm{hist}\rangle + (1-\beta^2)\langle\mathrm{hist}^{\perp}|H_{\mathrm{SPCC},5}|\mathrm{hist}^{\perp}\rangle.
\end{aligned}$$

Let us consider the four terms separately. For the first one we have

$$\langle\mathrm{hist}|\, H_{\mathrm{SPCC},5}\,|\mathrm{hist}\rangle \geq \epsilon\frac{1-s}{T+1},$$

using the promise on the QMA problem. For the second term,

$$\begin{aligned}
\langle\mathrm{hist}|\, H_{\mathrm{SPCC},5}|\mathrm{hist}^{\perp}\rangle &= \langle\mathrm{hist}|\, H_0 + \epsilon H_{\mathrm{out}}|\mathrm{hist}^{\perp}\rangle \\
&= \epsilon\,\langle\mathrm{hist}|\, H_{\mathrm{out}}|\mathrm{hist}^{\perp}\rangle \\
&\geq -\epsilon,
\end{aligned}$$

using that $\|H_{\text{out}}\| \leq 1$ and that $\langle \text{hist}| H_0 |\text{hist}^\perp\rangle = 0$, which holds since $|\text{hist}\rangle$ and $|\text{hist}^\perp\rangle$ are linear combinations of eigenstates of H_0. Similarly, for the third term, it must also hold that $\left\langle \text{hist}^\perp \right| H_{\text{SPCC},5} |\text{hist}\rangle \geq -\epsilon$. Finally, for the last term,

$$\langle \text{hist}^\perp | H_{\text{SPCC},5} | \text{hist}^\perp \rangle \geq \Delta \geq 0,$$

using again that H_{out} is positive semidefinite. Combining all of the above, we find

$$\begin{aligned}\langle \Psi| H_{\text{SPCC},5} |\Psi\rangle &\geq \beta^2 \epsilon \frac{1-s}{T+1} - 2\beta\sqrt{1-\beta^2}\epsilon \\ &\geq \left(1 - \frac{\epsilon}{\Delta}\right)\epsilon\frac{1-s}{T+1} - 2\sqrt{\frac{\epsilon}{\Delta}}\epsilon.\end{aligned}$$

Using Lemma 3.2.1 to lower bound Δ (which also holds for this particular H_0), the promise gap becomes at least

$$\begin{aligned}&\geq \left(1 - \frac{\epsilon}{\Delta}\right)\epsilon\frac{1-s}{T+1} - 2\sqrt{\frac{\epsilon}{\Delta}}\epsilon - \epsilon\frac{1-c}{T+1} \\ &\geq \epsilon\frac{c-s}{T+1} - \frac{\epsilon}{\Delta}\epsilon\frac{1}{T+1} - 2\sqrt{\frac{\epsilon}{\Delta}}\epsilon \\ &= \epsilon\frac{c-s}{T+1} - \frac{\epsilon^2}{\Delta}\frac{1}{T+1} - 2\frac{\epsilon^{3/2}}{\sqrt{\Delta}} \\ &= \epsilon\frac{c-s}{T+1} - \mathcal{O}\left(\epsilon^2 T^2\right) - \mathcal{O}\left(T^{3/2}\epsilon^{3/2}\right) \\ &= \Omega\left(1/T^7\right)\end{aligned}$$

for some choice of $\epsilon = \Omega((c-s)^2/T^6)$, which is $1/\text{poly}(n)$ since $c - s \geq 1/\text{poly}(n)$ and $T = \text{poly}(n)$. □

We note that this promise gap has worse scaling in T than what is achieved in Theorem 3.2.2. Finally, containment in QMA can easily be shown by Kitaev's energy estimation protocol, which we will study in more detail in Section 6.3.2.

3.3 The quantum marginal problem

In the previous section, we showed that the ability to solve the local Hamiltonian problem up to inverse polynomial precision suffices to solve any problem in QMA. However, one might observe that, whereas determining the acceptance probability of a QMA verification circuit requires access to the entire quantum proof as a global quantum state, computing the energy of a state with respect to a local Hamiltonian only requires *local information* from the proof—that is, classical descriptions of its local reduced density matrices.

But, of course, there is a catch: given a set of local density matrices, one must still verify whether they are consistent with a global quantum state. The question of whether a collection of reduced density operators for subsystems of a multipartite quantum system is compatible with a common global density operator is known as the *quantum marginal problem*. This problem can be formally defined as a promise problem as follows:

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

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

Promise: One of the following two cases holds:

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

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

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

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

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

If $\alpha = 0$, we write CLDM(q, β).

If (CLDM(q, α, β)) were solvable in QCMA for any $\beta - \alpha = \Omega(1/\text{poly}(n))$, it would imply that QCMA = QMA: the prover could simply provide the reduced density matrices corresponding to the ground state of a QMA-hard local Hamiltonian, along with a classical proof. The verifier could then compute the energy and use the classical proof to verify that the density matrices are consistent with a global quantum state (to sufficiently high precision). However, the following theorem shows that CLDM is QMA-hard.

3.3.2. Theorem ([BG22]). *There exists a constant $q = \mathcal{O}(1)$ and a polynomial $p(n)$ such that* CLDM(q, α, β) *with $\beta - \alpha \geq 1/p(n)$ is* QMA*-hard.*

Liu proved hardness under Turing reductions and showed that CLDM is in QMA for parameters satisfying $\beta/4^q =: \epsilon \geq 1/\text{poly}(n)$ and $\alpha = 0$ [Liu06]. Broadbent and Grilo established hardness under standard polynomial-time Karp reductions, and extended the containment result to arbitrary α (with $\epsilon = \beta/4^k - \alpha > 0$),

although their containment proof contains a minor flaw [BG22].[7] Nevertheless, the question remains open as to whether CLDM(q, α, β) is in QMA for any $\beta - \alpha \geq 1/\text{poly}(n)$.

We now show that a different proof technique can be used to lift this restriction on β. Our protocol is based on two key ideas: (i) with a sufficient number of copies, one can learn the local marginals of any given state; and (ii) using a specific formulation of a quantum de Finetti theorem, these marginals can be shown to be close to those of an actual quantum proof. We briefly outline both concepts below.

3.3.1 Full state tomography of marginals

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

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

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

with probability at least $1 - \delta$, using

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

copies of ρ.

[7] The proof relies on Hoeffding's inequality to establish soundness. However, the random variables it is applied to are generally not independent in this context, as the proof may involve highly entangled states.

Proof:
The measurement consists of applying a tensor product of different Pauli measurements M_j across the many copies of ρ to obtain estimates $\tilde{c}_{j,i} \approx \mathrm{tr}[P_j \sigma_i]$, from which all the marginals ρ_i can be approximately reconstructed. Since each measurement only acts on a single density matrix of a single copy of ρ, this ensures the tensor product structure of the overall measurement. Let $d_i = 2^{|C_i|}$, so that $\rho_i = \mathrm{tr}_{\overline{C}_i}[\rho]$ is the d_i-dimensional density matrix corresponding to the reduced density matrix of ρ that has indices from C_i. We have $\rho_i = \sum_{j \in [d_i^2]} c_{j,i} P_j$, where $c_{j,i} = \mathrm{tr}[P_j \rho_i]$. Using the measurement M_j corresponding to P_j, where each measurement outcome is bounded, standard mean estimation (see for example [Lee20]) gives an estimate $\tilde{c}_{j,i}$ such that $|\tilde{c}_{j,i} - c_{j,i}| \leq \epsilon / d_i^2$ with probability $1 - \delta/(m d_i^2)$, using $\mathcal{O}(d_i^4 \log(m d_i^2/\delta)/\epsilon^2)$ copies of ρ. We set $\tilde{\rho}_i = \sum_{j \in [d_i^2]} \tilde{c}_{j,i} P_j$. By a union bound and converting a bound on the max norm to the trace norm, we must have that for each $i \in [m]$, $\|\tilde{\rho}_i - \rho_i\|_1 \leq \epsilon$ holds with probability at least $1 - \delta/m$ for each i. Another union bound shows that the probability of $\|\tilde{\rho}_i - \rho_i\|_1 \leq \epsilon$ holding for all $i \in [m]$ simultaneously is at least $1 - \delta$. Using the upper bound $d_i \leq 2^q$ for all $i \in [m]$, the total number of copies (and thus measurements in the tensor product) can be upper bounded as

$$\mathcal{O}\left(mq 16^q \log(m/\delta)/\epsilon^2\right).$$

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

3.3.2 Quantum de Finetti under local measurements

Before explaining the ideas behind the variant of the quantum de Finetti theorem we use, it is helpful to introduce some additional notation. First, we will use superscripts instead of subscripts to indicate different parts of multipartite states $\rho^{A_1 \cdots A_l}$. For bipartite states ρ^{XY}, we stick to the convention that omitting superscripts corresponds to taking the partial trace over those systems; for example, $\rho^X = \mathrm{tr}_Y[\rho^{XY}]$. We say that $\rho^{A_1 \cdots A_k}$ is permutation symmetric if $\rho^{A_{\pi(1)} \cdots A_{\pi(k)}} = \rho^{A_1 \cdots A_k}$ for any permutation $\pi \in \mathrm{S}_k$ (recall that S_k is the symmetric group of order k). We associate to any POVM $\{M_k\}$ a map $\Lambda(X) = \sum_k \mathrm{Tr}(M_k X) |k\rangle\langle k|$, where $\{|k\rangle\}$ is an orthonormal basis. Thus, the $\Lambda(X)$ are so-called quantum-classical channels, as they map density operators to other density matrices that are diagonal in the basis $\{|k\rangle\}$. This implies that for two states ρ and σ, we have

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

where $p_k = \mathrm{tr}[M_k \rho]$ and $q_k = \mathrm{tr}[M_k \sigma]$ (see also Section 2.3.2).

Informally, the classical de Finetti Theorem states that if the joint probability distribution of a sequence of random variables is invariant under any permutation of the variables, then the marginal probability distribution of a subset of $l \ll k$ variables from such a k-variable sequence will be close to a convex combination of i.i.d. variables [DF80]. Quantum versions of the de Finetti Theorem posit that an l-partite quantum state $\rho^{A_1 \cdots A_l}$, which is the reduced state of a permutation-symmetric state $\rho^{A_1 \cdots A_k}$ on $k \gg l$ subsystems, is close to a convex combination of i.i.d. quantum states, i.e., $\rho^{A_1 \cdots A_l} \approx \int d\mu(\sigma)\, \sigma^{\otimes l}$, with μ a probability measure on quantum states. Different quantum de Finetti theorems consider different notions of "closeness", e.g., [Ren07], [CKMR07], and [BCY11]. We will focus on closeness with respect to local measurements performed on the subsystems, for which a quantum de Finetti theorem was proven by Brandão and Harrow [BH13a].

3.3.4. Lemma ([BH13a]). *Let $\rho^{A_1 \cdots A_k} \in \mathrm{D}(A^{\otimes k})$ be a permutation-invariant state. Then for every $0 \leq l \leq k$, there is a measure ν on $\mathrm{D}(A)$ such that*

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

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

3.3.3 The quantum marginal problem is in QMA

We now have all ingredients to give our QMA protocol for $\mathrm{CLDM}(q, \alpha, \beta)$, which is given in Protocol 3.3.1.

The core idea is that, with enough copies of a state, the verifier can estimate its local marginals via tomography. Then, using Lemma 3.3.4, we argue that any permutation-invariant state (which includes the verifier's post-processed state) is close to a separable state. Hence, from the verifier's perspective, the tomography is effectively performed on a state that is nearly separable, even if the prover were to send a highly entangled state. From this, we can show that the acceptance probability of the protocol is very close to that of an idealised version in which the prover sends a mixture of actual tensor-product copies, from which soundness follows.

Protocol 3.3.1: QMA protocol for CLDM(q, α, β)

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

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

Protocol:

1. The prover sends a state $\hat{\rho}^{A_1\dots A_k} \in \mathrm{D}(A^{\otimes k})$.
2. The verifier randomly permutes the index labels and traces out the last $k-l$ registers, creating the state $\rho^{A_1\dots A_l}$.
3. The verifier performs the measurement of Lemma 3.3.3, with desired precision ϵ and success probability $1-\delta$, on the registers $A_1,\dots,A_l$ and uses the measurement outcome to construct $\{\tilde{\rho}_i\}_{i\in[m]}$.
4. Accept if $\|\tilde{\rho}_i - \rho_i\|_1 \leq \alpha+\epsilon$ for all $i\in[m]$, and reject otherwise.

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

Proof:
We will show the correctness of Protocol 3.3.1.

Completeness. In this case, the prover sends the state $\sigma^{\otimes k}$, which is already permutation-invariant. After Step 2, the resulting state is $\sigma^{\otimes l}$. Hence, we have access to l perfect copies of σ, meaning that by Lemma 3.3.3, the estimates $\tilde{\sigma}_i$ of all m density matrices σ_i will be retrieved up to precision ϵ with high probability. Since we are in a YES-instance, this means that, conditioned on the estimation procedure succeeding, we have $\|\tilde{\sigma}_i - \rho_i\|_1 \leq \alpha+\epsilon$ for all $i\in[m]$, so we will accept with probability 1. Thus, the overall success probability is lower bounded by the success probability of the estimation procedure, which is $\geq 1-\delta \geq 2/3$ for our choice of δ.

Soundness. For any state $\hat{\rho}^{A_1\dots A_k}$, Step 2 ensures that the density matrix description of the state after Step 2, i.e., $\rho^{A_1\dots A_l}$, is permutation-invariant. Let

$P(\rho^{A_1\dots A_l})$ be the probability that the overall protocol in Protocol 3.3.1 accepts, given that the state after Step 2 is described by $\rho^{A_1\dots A_l}$. We will first argue that if the prover was honest and indeed provided a tensor product of multiple copies of some state, then we reject with high probability. Next, we will argue that by Lemma 3.3.4, Step 2 ensures that any arbitrary (so entangled) state must be close to such a state, and thus will also be rejected with high probability. We will now make this formal. Just as in the completeness case, suppose the state after Step 2 is again of the form

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

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

Recall that the $\Lambda_j(\cdot)$ are quantum-classical channels, mapping quantum states to discrete probability distributions, where each channel implements a single measurement $M^{(j)} = \{M^{(j)}_{+1}, M^{(j)}_{-1}\}$ corresponding to the one in Lemma 3.3.3 (note that the index j here does not represent the Pauli word, but labels the measurement instead). Step 4 can potentially distinguish between the two probability distributions described by the following two density matrices which are diagonal in some orthonormal basis

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

and

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

for some measure ν as per Lemma 3.3.4. Write p_1, p_2 for the distributions associated with measuring ρ_1, ρ_2 in their eigenbasis, respectively. By Lemma 3.3.4, and using the fact that the total variation distance upper bounds the maximum bias in the output by a single-sample distinguisher (see Section 2.3.2), we have

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

where

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

for our choice of parameters in Protocol 3.3.1. Here we also used that the trace distance is non-increasing under the application of a quantum channel. Thus, for our choice of δ, we have $P(\rho^{A_1 \ldots A_l}) \leq 2\delta = 1/3$, as desired.

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

3.4 States with useful succinct representations

By a simple counting argument, one can show that for n qubits, there exist *doubly exponentially* many quantum states, each with pairwise small fidelity. Consequently, the vast majority of quantum states cannot have efficient classical descriptions, as polynomially sized bit strings can describe at most exponentially many distinct objects. However, one could argue that we do not care about most Hamiltonians or quantum states in most scenarios. For instance, if we restrict attention to families of Hamiltonians with only local terms, we claim that (up to inverse exponential error in some suitable distance measure on quantum states) the ground state of *any* such Hamiltonian on n qubits *can* be specified using a polynomial number of bits. Namely,

$$\{|\psi\rangle\} = \{\text{ground states of } \mathrm{desc}(H)\}, \tag{3.8}$$

where $\mathrm{desc}(H)$ denotes a classical description of the Hamiltonian H. This description is efficient because locality ensures that $\mathrm{desc}(H)$ has polynomial size (up to exponential precision).

Still, one might justifiably argue that Eq. (3.8) is not *useful*, as the description does not allow us to extract any properties of the ground states. The same

holds for density matrices: while they allow us to compute properties of local observables, as we learned in Section 3.3, it remains a hard problem to verify whether they are consistent with a global state.

Besides having local terms, many other properties of Hamiltonians put restrictions on the family of possible Hamiltonians, like spectral gaps, geometric locality, specific families of terms, etc. It is believed (and sometimes even proven) that such structural constraints on Hamiltonians impose constraints on the structures of their ground states. There are families of states that do have useful classical descriptions and that are believed (or even proven) to capture some of these structural properties, some of which will be introduced in the next subsection. These families are commonly referred to as *ansätze* in the physics literature, and will be central to the remaining sections of the second half of this chapter.

3.4.1 Some examples of ansätze

We begin by reviewing some ansätze that are frequently encountered in practice. These will later serve to motivate our definitions of more abstract classes of ansätze that conform to specific access models.

Product states. Perhaps the simplest ansatz—commonly referred to as mean-field theory in quantum many-body physics or the Hartree-Fock method in quantum chemistry—is that of *product states*. For simplicity, we define these states in terms of clusters of qubits.

3.4.1. Definition (k-clustered product states). We say that a state $|u\rangle \in \mathbb{C}^{2^n}$ is a k-clustered product state if there exists a number $m \leq n$ such that

$$|u\rangle = |u_1\rangle_{A_1} \otimes \cdots \otimes |u_m\rangle_{A_m} ,$$

where each $|u_i\rangle_{A_i} \in \mathbb{C}^{2^{|A_i|}}$, and $\{A_1, \ldots, A_m\}$ is a partition of the set $[n]$ such that $|A_i| \leq k$ for all $i \in [m]$.

It is also possible to define k-clustered product states more generally by allowing each $|u_i\rangle$ in the product to be a state over a qudit of dimension d_i. Notably, there is no entanglement between the different components of the product.

Constant-depth quantum circuits. Another important class of ansätze consists of constant-depth quantum circuits.

3.4.2. Definition (Constant-depth circuit state). A *constant-depth circuit state* on n qubits is specified by a fixed initial state (typically $|0^n\rangle$), a fixed gate set $\mathcal{G}$ of a finite number of gates, an optional topological constraint that restricts which qubits may interact via multi-qubit gates, and a maximum circuit depth. The depth refers to the number of sequential layers of gates that may be applied in parallel.

It is known that when two gapped ground states of lattice Hamiltonians are in the same quantum phase of matter (or topological phase), they can be connected by a constant-depth quantum circuit [KR24]. Moreover, constant-depth quantum circuits form the class of states central to the NLTS theorem [ABN23].

Tensor network states. A *tensor network state* is a type of ansatz frequently encountered in quantum many-body physics, where a quantum state is described by a network of interconnected tensors. Formally, for a system of n qubits, any wavefunction $|\psi\rangle$ can be written as

$$|\psi\rangle = \sum_{i_1,i_2,\ldots,i_n} T^{i_1 i_2 \ldots i_n} |i_1 i_2 \ldots i_n\rangle ,$$

where $T^{i_1 i_2 \ldots i_n}$ is a rank-n tensor, and the indices $i_1, i_2, \ldots, i_n$ run over the physical degrees of freedom. The tensor T is typically decomposed into a network of smaller tensors connected by contracted indices, encoding the entanglement structure of the state.

Common examples of tensor network states include Matrix Product States (MPS) [KSZ93], Projected Entangled Pair States (PEPS) [VC04], and the Multiscale Entanglement Renormalisation ansatz (MERA) [Vid08]. It is well known that MPS efficiently approximate the ground states of gapped one-dimensional systems [Has07], and in such cases, it is even possible to compute these approximations in polynomial time [LVV15].

3.4.3. Definition (Matrix Product States). A *Matrix Product State* (MPS) is a quantum state of the form

$$|u\rangle = \sum_{\{s\}} \mathrm{Tr}\left[A_1^{(s_1)} A_2^{(s_2)} \ldots A_n^{(s_n)}\right] |s_1, \ldots, s_n\rangle ,$$

where $s_i \in \{0, 1, \ldots, d-1\}$ for physical dimension d, and $A_i^{(s_i)}$ are complex square matrices of bond dimension D. We say the MPS is bounded if $D \leq \mathrm{poly}(n)$ and $d = \mathcal{O}(1)$.

PEPS are a natural generalisation of MPS to two-dimensional systems.

3.4.4. Definition (Projected Entangled Pair State). A *Projected Entangled Pair State* (PEPS) is any (unnormalised) state that can be obtained by the following procedure: consider a graph and associate to each vertex v as many D-dimensional spins as there are edges incident to v. Assume that the spins associated with the endpoints of an edge form maximally entangled states, $|\mathrm{EPR}_D\rangle = \sum_{i=1}^{D} |i\rangle|i\rangle$. The PEPS is then obtained by applying a linear map $P_v : \mathbb{C}^D \otimes \cdots \otimes \mathbb{C}^D \to \mathbb{C}^d$ at each vertex v.

There are other equivalent definitions of PEPS, from which it can be readily seen that in the 1D case, the definition of MPS is recovered [CPGSV21]. It is known that computing the expectation value of local observables on a PEPS is #P-hard, and that preparing PEPS on a quantum computer is PostBQP-hard [SWVC07].

Another notable class of tensor network states is that of isometric tensor network states (isoTNS), which are known to capture two-dimensional string-net liquids [SSB+20], a family that includes many topological phases such as discrete gauge theories. IsoTNS form a subclass of PEPS. The following is a (slightly informal, as we will not use it directly) definition of a 2D isoTNS on a square lattice:

3.4.5. Definition (2D isoTNS on a square lattice). Let $\Lambda = (V, E)$ be a two-dimensional square lattice, and let $\chi \in V$ be the *orthogonality centre*. A PEPS is an *isometric tensor network state* (isoTNS) if, under some contraction order toward χ (e.g., row-by-row then column-by-column), each local tensor defines an isometry in that direction.

For more detailed definitions and constructions, see [ZP20, MT25].

3.5 Access models and families of states

We now turn to the question, "What makes a good ansatz?", particularly in the context of ground states. By "good," we mean that the ansatz satisfies certain desirable properties that can be efficiently exploited in computation.

In [GL22], the following access model to complex vectors was proposed.

3.5.1. Definition (Modified from [GL22]). We say that we have *query and ξ-sampling access* to a vector $u \in \mathbb{C}^N$ if the following three conditions are satisfied:

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

(ii) We have access to an $\mathcal{O}(\mathrm{poly}(\log N))$-time classical algorithm $\mathcal{Q}_u^{\mathrm{samp}}$ 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 such that $|m - \|u\|| \leq \xi\|u\|$.

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

Following [GL22], we specified Definition 3.5.1 in terms of complex vectors rather than quantum states (which is slightly less general), though this distinction will not be important for our purposes.

Given such an access model, [GL22] defines families of vectors (or states) that allow this type of access:

3.5.2. Definition (Modified from [GL22]). For a parameter $\xi \in [0,1]$, we say that a family of complex vectors $\mathcal{F} = \{u^{(l)}\}_{l\in\mathbb{N}}$ has a succinct representation allowing ξ-sampling access if there exist an injective function $\mathrm{rep} : \mathcal{F} \to \{0,1\}^*$, a polynomial q, and a polynomial-time algorithm A such that:

(i) For all $u \in \mathcal{F}$, the length of the binary string $\mathrm{rep}(u)$ is at most $q(\log(\dim u))$, where $\dim u$ denotes the dimension of u.

(ii) Algorithm A receives a binary string as input. If the string does not correspond to $\mathrm{rep}(u)$ for any $u \in \mathcal{F}$, then A outputs an error message. Otherwise, on input $\mathrm{rep}(u)$, A outputs the classical descriptions of the algorithms $\mathcal{Q}_u^{\mathrm{query}}$ and $\mathcal{Q}_u^{\mathrm{samp}}$, along with a real number m satisfying the specifications in Definition 3.5.1.

If $\xi = 0$, we say that $\mathcal{F}$ has a succinct representation allowing *perfect-sampling access*. If such an algorithm A exists for any $\xi \geq 1/\mathrm{poly}(n)$, we say that $\mathcal{F}$ is *samplable*.

The definition in [GL22] of families of states that allow succinct representation allowing ξ-sampling access was inspired by a line of work focused on the dequantization of quantum machine learning algorithms [Tan19, CGL$^+$22, JLGS20]. While this can be a somewhat powerful model [CHM21], it is closely related to the assumption of QRAM access (see [GLM08]) to classical data. 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. This helps rule out quantum speed-ups that arise merely from having access to quantum states constructed from exponentially large classical data.

For quantum chemistry and quantum many-body applications, however, it seems more natural to assume a different class of states. In particular, if one is interested in ground-state energies of local Hamiltonians, it is only necessary to extract local information from the description of the states. To capture this, we propose a new type of access model in the form of (ϵ-)local access, which requires only that one can efficiently (approximately) compute expectation values of local observables classically. To (i) make the notion of locality well-defined, (ii) ensure a fixed bound on the allowed operator norm of observables relative to the norm of the vector, and (iii) since we are only interested in quantum states anyway, we restrict to normalised vectors $u \in \mathbb{C}^{2^n}$.

3.5.3. Definition. For $\epsilon > 0$, we say that we have *ϵ-local access* to a normalised vector $u \in \mathbb{C}^{2^n}$ if there exists a classical probabilistic algorithm $\mathcal{Q}_u^{\text{local}}$ which, given query access to the matrix elements of some k-local observable O with $\|O\| \leq 1$, computes an estimate $\hat{z}$ such that $|\hat{z} - \langle u| O |u\rangle| \leq \epsilon$ in time $\text{poly}(n, 1/\epsilon, 2^k)$, with success probability at least $2/3$. If we have ϵ-local access for any $\epsilon \geq 1/\text{poly}(n)$, we say that we have *approximate local access*. If instead the algorithm $\mathcal{Q}_u^{\text{local}}$ is deterministic and computes $\langle u| O |u\rangle$ (i.e., up to a fixed polynomial number of bits of machine precision) in time $\text{poly}(n, 2^k)$, we say that we have *local access* to u.

3.5.4. Remark. For any $\delta > 0$, the success probability in Definition 3.5.3 can be amplified to $1 - \delta$ by performing $\mathcal{O}(\log(1/\delta))$ repetitions of $\mathcal{Q}_u^{\text{local}}$ and taking the median of the outcomes as the final estimate.

Let $\mathcal{P}_2 = \{\mathbb{C}^{2^n} : n \in \mathbb{Z}_+\}$. Given our definition of local access, we can again associate families of states.

3.5.5. Definition (Classically evaluatable families of states). For $\epsilon > 0$, we say that a family of normalised complex vectors $\mathcal{F} = \{u^{(l)} \mid u^{(l)} \in \mathcal{P}_2\}_{l \in \mathbb{N}}$ is *ϵ-classically evaluatable* if there exist an injective function $\text{rep} : \mathcal{F} \to \{0,1\}^*$, a polynomial q, and a polynomial-time algorithm A satisfying the following conditions:

(i) For all $u \in \mathcal{F}$, the length of the binary string $\text{rep}(u)$ is at most $q(\log(\dim u))$, where $\dim u$ denotes the dimension of u.

(ii) The algorithm A takes a binary string as input. If the string does not correspond to $\text{rep}(u)$ for any $u \in \mathcal{F}$, then A outputs an error message. Otherwise, on input $\text{rep}(u)$, the algorithm outputs the classical description of an algorithm $\mathcal{Q}_u^{\text{local}}$ satisfying the specifications of Definition 3.5.3.

If the algorithm $\mathcal{Q}_u^{\text{local}}$ provides local access, we say that $\mathcal{F}$ is *classically evaluatable*. If $\mathcal{F}$ is ϵ-classically evaluatable for any $\epsilon \geq 1/\text{poly}(n)$ (i.e., it allows approximate local access), we say that $\mathcal{F}$ is *approximately classically evaluatable*.

We will also frequently rely on the following two useful properties of families of classically evaluatable states, which will become particularly important in Chapter 4.

3.5.6. Proposition (Properties of local access). *Let $n \in \mathbb{Z}_+$, and let $|u\rangle \in \mathbb{C}^{2^n}$ be a quantum state to which we have ϵ-local access, for some $0 < \epsilon \leq 1$. Then the following hold:*

1. *(Tensor products). Let $m \in \mathbb{Z}_+$, and let $|w\rangle \in \mathbb{C}^{2^m}$ be a quantum state to which we have ϵ-local access. Then we have $3 \cdot 4^k \epsilon$-local access to $|u\rangle \otimes |w\rangle$.*

2. *(Local isometries). Let $A_1, A_2, \ldots, A_m$ be a partition of the set $[n]$ with $m \leq n$, such that each A_i consists of q_i qubits for some constant $q_i \in \mathbb{Z}_+$. Associate to each subset A_i a Hilbert space $\mathcal{H}_i = \mathbb{C}^{2^{q_i}}$. Let $\{V_i\}_{i\in[m]}$ be a collection of isometries such that each $V_i : \mathcal{H}_i \to \mathcal{H}_i'$ with $\dim(\mathcal{H}_i') = 2^{q_i'}$ for some constant $q_i' \geq q_i$. Then we have ϵ-local access to the state $V_1 \otimes \cdots \otimes V_m |v\rangle$.*

If the states allow for perfect (or approximate) local access, then the same statements hold with ϵ-local access replaced accordingly.

Proof:
We verify that the conditions of Definition 3.5.3 are satisfied in both cases.

Tensor products. Let A (resp. B) be the set of qubits from the Hilbert space of $|u\rangle$ (resp. $|w\rangle$) on which the observable O acts, with $|A| + |B| \leq k$. Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be the corresponding Hilbert spaces. Using the Schmidt decomposition for operators on k qubits, we can decompose the non-trivial part of O as:

$$\sum_{i=1}^{r} \lambda_i C_i \otimes D_i \tag{3.9}$$

where $\lambda_i \geq 0$, and $\{C_i\} \subset \mathrm{Herm}(\mathbb{C}^{2^{|A|}})$ and $\{D_i\} \subset \mathrm{Herm}(\mathbb{C}^{2^{|B|}})$ are orthonormal under the Hilbert–Schmidt inner product and r the operator Schmidt rank. Since tensoring with identities does not increase the Schmidt rank, the global matrix representation of Eq. (3.9) has the same decomposition except for tensoring with identity matrices and the specification of a permutation as per Definition 3.1.1. In a slight abuse of notation, extend each C_i, D_i into their n-, m-qubit matrix representations, respectively. We have

$$(\langle u| \otimes \langle w|) O (|u\rangle \otimes |w\rangle) = \sum_{i=1}^{r} \lambda_i \langle u| C_i |u\rangle \langle w| D_i |w\rangle .$$

Since O has operator norm at most 1, its Frobenius norm satisfies $\|O\|_2 = \sqrt{\sum_{i=1}^{r} \lambda_i^2} \leq \sqrt{r}$. Therefore,

$$\sum_{i=1}^{r} \lambda_i \leq \sqrt{r} \cdot \sqrt{\sum_{i=1}^{r} \lambda_i^2} \leq r.$$

We can efficiently estimate each $\langle u| C_i |u\rangle$ and $\langle w| D_i |w\rangle$ to ϵ precision with probability $1 - \delta$, for any $\delta \geq 1/\exp(n)$ (Remark 3.5.4). Hence, conditioning on all

estimations succeeding (which can be achieved with high probability),

$$\begin{aligned}|\hat{z} - (\langle u| \otimes \langle v|)\,(O \otimes \mathbb{I})\,(|u\rangle \otimes |v\rangle)| &\leq \left|\sum_{i=1}^{r} \lambda_i \left(\langle C_i\rangle \pm \epsilon\right)\left(\langle D_i\rangle \pm \epsilon\right) - \sum_{i=1}^{r} \lambda_i \langle C_i\rangle\langle D_i\rangle\right| \\ &\leq \sum_{i=1}^{r} \lambda_i \left(2\epsilon + \epsilon^2\right) \\ &\leq 3r\epsilon\end{aligned}$$

assuming $\epsilon \leq 1$, so we obtain $3r\epsilon$-local access. Since $r \leq 4^k$, the statement follows.

Local isometries. We want to evaluate:

$$\langle v| \left(V_1^\dagger \otimes \cdots \otimes V_m^\dagger\right) O \left(V_1 \otimes \cdots \otimes V_m\right) |v\rangle .$$

Let $B \subseteq [n]$ with $|B| = k$ be the set of qubits on which O acts non-trivially. For each i such that V_i acts only on qubits outside B, we have $V_i^\dagger V_i = \mathbb{I}$. Therefore, only the isometries acting on B contribute non-trivially. Define $I \subseteq [m]$ to be the indices of those V_i acting on B. Then we can write (again omitting identities and the permutation as per Definition 3.1.1)

$$\langle v| V^\dagger O V |v\rangle ,$$

where $V \coloneqq \bigotimes_{i\in I} V_i$. Since $q_i, q_i' \in \mathcal{O}(1)$ for each $i \in I$ and $I = \mathcal{O}(1)$, defining $Q \coloneqq V^\dagger O V$, we see that Q is k'-local with $k' = \mathcal{O}(k)$ and we can compute its description efficiently in terms of 2^k. Since $|v\rangle$ is classically evaluatable, we can compute $\langle v| Q |v\rangle$ to ϵ precision, which implies that have ϵ-local access to $V_1 \otimes \cdots \otimes V_m |v\rangle$.

For both properties, it is clear that the same proof structure shows that perfect or approximate local access is preserved. □

Finally, we consider one more type of access, which we refer to as quantumly preparable-access.

3.5.7. Definition (Quantumly-preparable access). Let $\mathcal{G}$ be a fixed universal gate set. For $\epsilon \geq 0$, we say that we have *ϵ-quantumly-preparable access* to a normalised vector $u \in \mathbb{C}^{2^n}$ if there exists a quantum circuit V consisting of at most $\mathrm{poly}(n, 1/\epsilon)$ gates from $\mathcal{G}$, acting on $m = \mathrm{poly}(n, 1/\epsilon)$ qubits initialized in $|0^m\rangle$, such that

$$\|(|u\rangle\langle u| \otimes \mathbb{I})\, V\, |0^m\rangle\|^2 \geq 1 - \epsilon.$$

If $\epsilon = 0$, we say that we have *perfect quantumly-preparable access* to $|u\rangle$. If $|u\rangle$ allows ϵ-quantumly-preparable access for any $\epsilon \geq 1/\mathrm{poly}(n)$, we say that we have *quantumly-preparable access* to $|u\rangle$.

Ansatz family	Samplable?	Approximately classically evaluatable?	Uniformly quantumly preparable?
$\mathcal{O}(\log n)$-clustered PS	✓	✓*	✓
Bounded MPS	✓	✓*	✓
Constant depth qc	✗[1]	✓*	✓
2D isoTNS	✗[2]	✗[2]	✓
PEPS	✗[3]	✗[3]	✗[4]

Table 3.1: A '*' means that the ansatz state is also classically evaluatable. Complexity-theoretic assumptions indicated by the superscripts: [1] $\mathsf{BQP} \not\subset \mathsf{AM}$, [2] $\mathsf{BQP} \not\subset \mathsf{BPP}$, [3] $\#\mathsf{P} \not\subset \mathsf{FBPP}$; [4] $\mathsf{PP} \not\subset \mathsf{BQP}$. All ansätze in this table that are samplable also allow perfect-sampling access.

The notion of quantumly-preparable access is robust to the choice of universal gate set, by the Solovay–Kitaev theorem (Theorem 2.4.2). As before, we consider an analogue of Definition 3.5.1 and Definition 3.5.5 for families of states that are preparable by uniformly generated quantum circuits.

3.5.8. Definition (Uniformly quantumly preparable families of states). Let $\epsilon \geq 0$. We say that a family of normalised complex vectors $\mathcal{F} = \{u^{(l)} \mid u^{(l)} \in \mathcal{P}_2\}_{l\in\mathbb{N}}$ is *uniformly ϵ-quantumly preparable* if there exist an injective function $\mathrm{rep} : \mathcal{F} \to \{0,1\}^*$, a polynomial q, and a polynomial-time algorithm A satisfying the following conditions:

(i) For all $u \in \mathcal{F}$, the length of the binary string $\mathrm{rep}(u)$ is at most $q(\log(\dim u))$, where $\dim u$ denotes the dimension of u.

(ii) Algorithm A receives a binary string as input. If the string does not correspond to $\mathrm{rep}(u)$ for any $u \in \mathcal{F}$, then A outputs an error message. Otherwise, on input $\mathrm{rep}(u)$, A outputs the classical description of a quantum circuit V satisfying the specification of Definition 3.5.7.

If $\epsilon = 0$, we say that $\mathcal{F}$ is *perfectly uniformly quantumly-preparable*. If $\mathcal{F}$ is uniformly ϵ-quantumly-preparable for any $\epsilon = \Omega(1/\mathrm{poly}(n))$, we say that $\mathcal{F}$ is *uniformly quantumly-preparable*.

3.5.1 Examples classification and relations amongst classes

Table 3.1 shows for the examples given in Section 3.4.1 whether they are samplable, classically evaluatable and uniformly quantumly-preparable. For each example, the statement underlying the entry in Table 3.1 is proven in Appendix B.

As Table 3.1 already hints at, it seems that approximately classically evaluatable states capture a larger class of states compared to samplable states. In fact,

this is easy to show by adapting the main technique from [GL22], which shows that one can estimate the following quantities involving samplable vectors.

The precise definitions of types of query access in the next lemma can be found in [GL22], and are not that important to us. All we need to know is that a matrix is s-sparse if each row and column has at most s non-zero entries, which is trivially polynomially bounded for $\mathcal{O}(\log n)$-local observables.

3.5.9. Lemma (Adapted from Theorem 4.1 in [GL22]). *Let $P \in \mathbb{R}[x]$ be an even polynomial of degree $2d$ with $|P(x)| \leq 1$ for all $x \in [-1, 1]$. There is an $\mathrm{poly}(s^d, 1/\epsilon)$-time classical procedure that, given:*

- *query access to an s-sparse matrix $A \in \mathbb{C}^{M\times N}$ with $\|A\| \leq 1$,*
- *query access to a vector $u \in \mathbb{C}^N$ such that $\|u\| \leq 1$,*
- *ξ-sampling access to a vector $v \in \mathbb{C}^N$ such that $\|v\| \leq 1$,*

outputs an estimate $\hat{z} \in \mathbb{C}$ such that

$$|\hat{z} - v^\dagger P(\sqrt{A^\dagger A})u| \leq \epsilon$$

with probability $\geq 1 - 1/\mathrm{poly}(N)$, for any $\xi \leq \epsilon/8$.

Lemma 3.5.9 readily implies our desired corollary.

3.5.10. Corollary. *If a family of states $\mathcal{F}$ is samplable, then it is also approximately classically evaluatable.*

Proof:
Let $N = 2^n$ and consider any $u \in \mathcal{F}$, with $u \in \mathbb{C}^N$ a normalised vector. Let O be any k-local Hermitian observable, given in terms of its matrix entries, with operator norm $\|O\| \leq 1$ and locality $k = \mathcal{O}(\log n)$. Since each term acts on at most k qubits, O has sparsity at most $s = 2^k$. Define the observable $Q = \frac{O+\mathbb{I}}{2}$, which satisfies $0 \preceq Q \preceq \mathbb{I}$. Since O may consist of only off-diagonal entries, the sparsity of Q is at most $2s$. Observe that $\langle u| O |u\rangle = 2 \langle u| Q |u\rangle - 1$, so estimating $\langle u| Q |u\rangle$ up to additive error $\epsilon' = \epsilon/2$ suffices to estimate $\langle u| O |u\rangle$ to within error ϵ. Because Q is positive semidefinite, we have $\sqrt{Q^\dagger Q} = Q$, and thus $\langle u| Q |u\rangle$ is a polynomial in $\sqrt{Q^\dagger Q}$ of degree $d = 1$. Since $\mathcal{F}$ is samplable, for any target accuracy $\epsilon' = \Omega(1/\mathrm{poly}(n))$, we may pick $\xi \leq \epsilon'/8$ and invoke Lemma 3.5.9 to obtain an efficient additive-error estimate of $\langle u| Q |u\rangle$ with high probability. Since the runtime of the algorithm in Lemma 3.5.9 is $\mathrm{poly}(s^d, 1/\epsilon)$, it follows that $\langle u| O |u\rangle$ can be estimated to within any additive error ϵ in time $\mathrm{poly}(n, 1/\epsilon, 2^k)$ and thus $\mathcal{F}$ is approximately classically evaluatable (the dependence of n is implicit in the algorithm A as per Definition 3.5.2). □

We conjecture that this is the only inclusion relation that holds among the state classes defined above. In the next and final section, we explore how these classes can be separated in the context of ground state energy estimation.

3.6 Connection to ground states

In this final section, we connect the previously introduced classes of ansätze to ground states of Hamiltonians. Specifically, we prove that, under reasonable complexity-theoretic assumptions, there exist:

- a family of local Hamiltonians that has a corresponding family of ground states that is both perfectly classically evaluatable and uniformly quantumly-preparable, but does not have a succinct representation allowing ξ-sampling access for any $0 \leq \xi < 1/3$;
- a family of local Hamiltonians that has a corresponding family of ground states that is uniformly is quantumly-preparable, but does not have any *any* efficiently describable[8] family that is approximately classically evaluatable or samplable.

Our first construction relies on the concept of *light cones* in quantum circuits.

Light cones. For a k-qubit operator O and a circuit U acting on $n \geq k$ qubits, we say that the light cone of O with respect to U is the set of qubits on which the operator $UOU^\dagger$ acts non-trivially. For a set of qubits $I = \{i_1, \ldots, i_k\}$, the light cone of I is the union of light cones over all operators O supported on qubits with indices from I. Suppose that U is a constant-depth circuit of depth t, consisting only of 2-qubit gates. If $|I| = k$, it can easily be shown that the size of the light cone of qubits from I with respect to circuit U is at most $2^t k$ (see, for example, [TD04, page 9]).

3.6.1. Proposition. *There exists a family of local Hamiltonians $H = \{H_n : n \in \mathbb{Z}_+\}$ whose ground states are part of a family $\mathcal{F}$, such that*

- *$\mathcal{F}$ is perfectly classically evaluatable and uniformly quantumly-preparable;*
- *$\mathcal{F}$ does not have ξ-sample access for any $0 \leq \xi < 1/3$, unless $\mathsf{BQP} \subseteq \mathsf{AM}$.*

Proof:
We will use a proof by contradiction. Assume $\mathsf{BQP} \not\subseteq \mathsf{AM}$. Recall from Appendix B that in [TD04] it is shown that the ability to perform approximate weak sampling from the output of a P-uniform family of constant depth quantum circuits $\mathcal{U} = \{V_n : n \in \mathbb{Z}_+\}$, that uses only 2-local gates acting on n qubits initialized in $|0^n\rangle$, up to relative error $0 < \xi < 1/3$ implies that $\mathsf{BQP} \subseteq \mathsf{AM}$. Let $V_n \in \mathcal{U}$ be a circuit of depth $t = \mathcal{O}(1)$. Now consider the 1-local Hamiltonian

$$H_n = \frac{1}{n} \sum_{i \in [n]} |1\rangle\langle 1|_i \otimes \mathbb{I}.$$

[8] See Definition 3.6.2.

We have that $|0^n\rangle$ is the unique ground state of H_n and every other eigenstate has energy $\geq 1/n$. Now consider the 2^t-local Hamiltonian H'_n defined by

$$H'_n = V_n H_n V_n^\dagger = \frac{1}{n}\sum_{i\in[n]} V_n \left|1\right\rangle\!\left\langle 1\right|_i V_n^\dagger = \frac{1}{n}\sum_{i\in[n]} \left(\prod_{j\in L^{(i)}} V_{n,j}\right) \left|1\right\rangle\!\left\langle 1\right|_i \left(\prod_{j'\in L^{(i)}} V_{n,j'}^\dagger\right),$$

where $L^{(i)} = \{j : V_{n,j} \text{ in light-cone of qubit } i\}$ and $\{V_{n,j}\}$ are the 2-qubit gates used in the circuit V_n. The ground state of H'_n is given by $V_n|0^n\rangle$, which is classically evaluatable and uniformly quantumly-preparable as it is prepared by a constant-depth quantum circuit (see Appendix B). Hence, the family of states $\mathcal{F} = \{V_n|0^n\rangle : V_n \in \mathcal{U}\}$ contains all ground states from the family of Hamiltonians $H = \{H_n : n \in \mathbb{Z}_+\}$. Now suppose that there exists a $0 \leq \xi < 1/3$, such that $\mathcal{F}$ also allows ξ-sample access. Since ξ-sampling from $\mathcal{F}$ would be identical to ξ-sampling from the output of the circuits V_n from the family U, this implies $\mathsf{BQP} \subseteq \mathsf{AM}$, which is in contradiction with our assumption. Hence, there exists no such $0 \leq \xi < 1/3$. □

We will now show that an even stronger statement can be made with respect to uniformly quantumly-preparable states. In Proposition 3.6.1, our statement concerned a Hamiltonian for which the verifier itself could efficiently find the exact form of the ground state. From this ground state description alone, it was impossible to construct the algorithm $\mathcal{Q}_u^{\text{samp}}$ to sample basis states from the ground state according to the Born rule. However, in this construction, it was essential to restrict the family of considered states to describe the ground state. In the next setting, we can show that it holds for all possible families $\mathcal{F}$ that allow for either approximately local or query and sample access.

First, there is one technicality we have to overcome. As we will have a prover-verifier setting, we will need a way to efficiently specify a certain family $\mathcal{F}_x$. This allows the prover to provide two classical witnesses: one indicating the family $\mathcal{F}_x$ (which tells the verifier what algorithm A to use), and the other to specify $\text{rep}(u)$ which describes the vectors to be used by A. Hence, we need one final definition, which is that of efficiently describable families that allow an access model as defined in Section 3.5. We will state the definition only explicitly for classically evaluatable states:

3.6.2. Definition. Consider a fixed Turing machine M. We say that a collection $\{\mathcal{F}_z\}_{z\in\{0,1\}^*}$ of classically evaluatable families of vectors is efficiently describable (by M) if, for every $z \in \{0,1\}^*$, the machine M outputs a classical description of the algorithm A satisfying Definition 3.5.5 for the family $\mathcal{F}_z$ in time $\text{poly}(|x|)$.

Similarly, one can give definitions for the existence of such Turing machines for samplable and uniformly quantumly preparable states, referring to the algorithm A in Definition 3.5.1 or the circuits V in Definition 3.5.8. One can view M as

a universally agreed-upon compiler that maps string encodings to efficient access procedures for well-defined families of states. For example, it could be that the string $z = 00100100$ indicates the family consisting of all matrix product states with bounded bond dimension and a fixed local dimension, and running $M(z)$ returns a classical description of the algorithm A (which could as a subroutine have the algorithm given in Appendix B). There might be many classes of vectors that we do not yet know of that allow sampling access or (approximately) local access, but this will not be important for the proof of the next proposition.

3.6.3. PROPOSITION. *There exists a family of local Hamiltonians* $H = \{H_x : x \in \{0,1\}^*\}$ *whose ground states:*

- *are part of a family* $\mathcal{F}$ *which is uniformly quantumly-preparable;*
- *are not part of any efficiently describable family of states* $\mathcal{F}'$ *that is either approximately classically evaluatable or samplable, unless* $\mathsf{QCMA} \subseteq \mathsf{MA}$.

Proof:
We will again use a proof by contradiction, assuming that $\mathsf{MA} \subsetneq \mathsf{QCMA}$. In Appendix D of [WFC24], it is shown that the following problem is QCMA-hard:[9] One is given a 6-local Hamiltonian H which is promised to have either (i) a ground state which comes from a family of uniformly quantumly-preparable states and has energy 0; or (ii) ground state energy $\geq b = 1/p(n)$ for some polynomial $p(n)$, and the task is to decide which of the two holds. Consider the family of Hamiltonians $\{H_x : x \in \{0,1\}^*\}$ with uniformly quantumly-preparable ground states $\{|\psi_x\rangle\}$ in case (i), for which the above problem is QCMA-complete. Suppose that every $|\psi_x\rangle$ is also part of a family of states that is classically evaluatable and efficiently describable by the Turing machine M from Definition 3.6.2. Then for every x and every choice of $\epsilon \geq 1/\mathrm{poly}(n)$, there exists a family $\mathcal{F}_z$ of ϵ-classically evaluatable vectors, containing a normalised vector $u \in \mathcal{F}_z$, such that we can estimate $\langle\psi_x| O |\psi_x\rangle$ for any $\mathcal{O}(\log n)$-local observable O up to ϵ-error with probability $1-\delta$ in time $\mathrm{poly}(n, 1/\epsilon, \log(1/\delta))$ (see Remark 3.5.4), using only the description of u, with $n = \log\dim(|\psi_x\rangle)$. Since H_x is 6-local, it has at most $m = \mathcal{O}(n^6)$ 6-local terms. Hence, if $\epsilon \leq m/4p(n)$ and $\delta \leq 1/3m$, we can estimate $\langle\psi_n| H_x |\psi_n\rangle$ up to an additive error of $\leq 1/4p(n)$ with probability $\geq 2/3$.

The MA protocol is as follows: the prover provides the description of the family $\mathcal{F}_z$ as a string z, as well as string denoting $\mathrm{rep}(u)$ for some $u \in \mathcal{F}_z$ of the supposed ground state $|\psi_n\rangle$. The verifier uses M with input z to find the algorithm A, and then uses A to check whether $\mathrm{rep}(u)$ corresponds to some $|u\rangle \in \mathcal{F}_z$ and output $\mathcal{Q}_u^{\mathrm{local}}$. If the check is sound, the verifier uses the algorithm $\mathcal{Q}_u^{\mathrm{local}}$

[9] We have not included this result in this thesis, but the idea is to do a reduction similar to what is done in Section 4.4.1 but starting from QCMA_1 instead of UQCMA. The ground states are then history states for classical proofs that get accepted with probability 1, which are uniformly quantumly-preparable.

to approximately compute $\langle\psi_x| H_x |\psi_x\rangle$ using $\mathrm{rep}(u)$ and checks if it is smaller than $1/4p(n)$. Completeness follows from the above argument, and soundness follows from the variational principle. Hence, this would imply $\mathsf{QCMA} \subseteq \mathsf{MA}$, contradicting the assumption.

The same argument applies to families of samplable states by Lemma 3.5.9, using a modified definition of Definition 3.6.2 tailored to samplable states. □

3.6.1. Open problem. *Are there families of samplable states and/or classically evaluatable states that are provably not uniformly quantumly preparable? What about considering only query or sampling access?*

Chapter 4

Local Hamiltonians with guiding states

4.1 Introduction

Simulation of physical systems is one of the originally envisioned applications of quantum computing [Fey82, Fey86]. Quantum chemistry and quantum many-body physics, in particular, have seen much activity on this front in recent years, e.g. [Aar09, AGDLHG05, BBMC20, LBG+21, RWS+17, SBW+21]. As we discussed in Chapter 3, estimating ground state energies of local Hamiltonians plays a central role here; unfortunately, it is nowadays well-known that estimating ground state energies of local Hamiltonians is QMA-complete [KSV02], which persists even in the bosonic [WMN10] and fermionic settings [SV09]. Thus, assuming BQP $\neq$ QMA, one cannot hope for an efficient algorithm for the local Hamiltonian problem on *all* k-local Hamiltonians.

As we saw in Chapter 1, a commonly proposed strategy to sidestep worst-case hardness in practice is to use the following two-step approach for estimating ground state energies on a quantum computer:

- (Step 1: Ground state approximation) A classical (or quantum) heuristic algorithm is applied to obtain a "guiding state" $|\psi\rangle$, which is hoped to have "good" fidelity with a ground state.
- (Step 2: Ground state energy approximation) The guiding state $|\psi\rangle$ is used in Quantum Phase Estimation (QPE) [Kit95] to efficiently compute the corresponding ground state energy [AL99, AGDLHG05].

In [GL22], Gharibian and Le Gall initiated the formal study of Step 2 above, by introducing the *Guided k-Local Hamiltonian problem (k-GLH)*. Roughly, k-GLH can be stated as follows (see Definition 3.1.2 for the formal version): given a k-local Hamiltonian H, a suitable representation of a guiding state $|\psi\rangle$ with ζ-fidelity to the ground space of H, and real thresholds $b > a$, estimate the ground state energy of H. Here, we assume that the Hamiltonian has been renormalised to have operator norm at most 1.

Two main results were established in [GL22]:

- For any constant k, k-GLH can be solved efficiently *classically* to *constant* precision, i.e., for $b - a = \Theta(1)$ and $\zeta = \Theta(1)$.
- In contrast, 6-GLH is BQP-hard for *inverse polynomial* precision, i.e., for $b - a = 1/\text{poly}(n)$ and $\zeta = 1/2 - 1/\text{poly}(n)$.[1]

The latter regime of inverse-polynomial precision turns out to be relevant for practical quantum chemistry applications: the target "chemical accuracy" is around 1.6 millihartree, which, after renormalising the Hamiltonian, corresponds to the stated inverse-polynomial precision.[2] This BQP-hardness result therefore provides theoretical support for the advantage of quantum algorithms in chemistry.

Many important questions were left open in [GL22]:[3] Is k-GLH still BQP-hard with larger ζ, particularly when ζ is arbitrarily inverse polynomially close to 1? Is k-GLH still BQP-hard for $k < 6$? Is k-GLH still BQP-hard for estimating excited state energies? Is k-GLH still BQP-hard for physically motivated Hamiltonians? Finally, what happens to the complexity of the problem if we also incorporate Step 2? Can we use this to say something about the power of classical versus quantum heuristics in the context of the above two-step procedure for ground state energy estimation?

4.1.1 Results of this chapter

In this chapter, we resolve all of the open problems mentioned above. Specifically, we will prove the following results:

- First, we show that BQP-hardness continues to hold even for $\zeta = 1 - 1/\text{poly}(n)$, i.e., even when we are promised that the guiding state $|\psi\rangle$ is a remarkably good approximation to the ground state.
- Second, we extend the BQP-hardness results to the case where one is interested in estimating energies of excited states, rather than just the ground state.
- Third, we prove hardness results for *physically motivated* Hamiltonians. These include the 2-local XY model (constraints of the form $XX + YY$), the Heisenberg model (constraints of the form $XX + YY + ZZ$), the antiferromagnetic XY model, and the antiferromagnetic Heisenberg model.

[1] Such hardness statements should be read as: there exists a polynomial $p(n)$ such that the problem is hard when $b - a = 1/p(n)$.

[2] Though this corresponds to only inverse-linear precision relative to the operator norm. It remains an open question whether the general local Hamiltonian problem is QMA-hard in this regime.

[3] See Chapter 1 for additional motivation behind these questions.

- Fourth, we define "Merlinised" versions of the guided local Hamiltonian problem, which we call *guidable local Hamiltonian problems*. Unlike their guided counterparts, these problems do not come with a guiding state as part of the input, but only with the promise that one *exists*. We prove that the guidable local Hamiltonian problem is QCMA-hard, and show that the additional results above in the guided setting translate directly to the guidable setting.

We will focus only on hardness as containment in the respective quantum classes is already known to hold via quantum phase estimation (QPE)[4] or more recent techniques [LT20b, LT20a] based on the quantum singular value transformation [GSLW19].

As a complementary result to this chapter, we present an alternative to the dequantization algorithm proposed in [GL22] in Appendix C, showing that the above problems are classically (non-deterministically) solvable in the constant relative precision regime for classically evaluatable guiding states (assuming that ζ is also lower bounded by a constant).

4.2 The guided local Hamiltonian problem

Let us start by formally defining the *guided local Hamiltonian problem*, slightly modifying the definition as introduced by [GL22] by incorporating classically evaluatable states (Definition 3.5.5) as our guiding states.

4.2.1. Definition (Guided local Hamiltonian problem, GLH(k, a, b, ζ)).

Input: A description of a collection of k-local Hermitian operators $\{H_i\}_{i\in[m]}$, which define an n-qubit k-local Hamiltonian $H = \sum_{i=1}^{m} H_i$ with $\|H\| \leq 1$, where $m = \text{poly}(n)$; a description of a classically evaluatable state $|u\rangle \in \mathbb{C}^{2^n}$; two efficiently computable real numbers a, b such that $b > a$; and an efficiently computable fidelity parameter $\zeta \in [0, 1]$.

Promise: $\|\Pi_{\text{gs}}^{H} |u\rangle \|^2 \geq \zeta$, where Π_{gs}^{H} denotes the projection onto the subspace spanned by the ground states of H, and either $\lambda_0(H) \leq a$ or $\lambda_0(H) \geq b$ holds.

Goal: Decide whether $\lambda_0(H) \leq a$ or $\lambda_0(H) \geq b$.

When $|u\rangle$ is also uniformly quantumly preparable (Definition 3.5.8), we write GLH$^*(k, a, b, \zeta)$.

[4]Proving that QPE works in this guided setting is actually rather tedious; see Chapter 3 in Lin's lecture notes [Lin22]. The reason is that in the standard derivation of QPE, three assumptions are typically made: (i) the eigenphase θ can be exactly expressed in a finite number of bits; (ii) one is given the exact eigenstate; and (iii) the unitary can be implemented perfectly. While assumption (iii) is easy to relax, handling (i) and (ii) simultaneously requires some care.

Note that Definition 4.2.1 differs from Definition 3.1.2 not only in its extra guiding state input, but also in the norm constraints imposed on the local Hamiltonian and its terms. In Definition 3.1.2, we required only that each local term has operator norm at most one, which implies that the operator norm of the entire Hamiltonian is at most m. This allows us to enforce a global norm bound of one by simply rescaling the Hamiltonian by a factor of $1/m$, which correspondingly scales the promise gap. The choice of a global norm constraint in Definition 3.1.2 is primarily historical—following [GL22], where the problem was first defined—but also motivated by the fact that the guided (and guidable) local Hamiltonian problem has connections to the quantum PCP conjecture (see also Appendix C). As we will see in Chapter 6, quantum PCPs naturally induce local Hamiltonians that satisfy such a global norm constraint.

As a technical remark, we define $\mathrm{GLH}^*(H, a, b, \zeta)$ separately from $\mathrm{GLH}(H, a, b, \zeta)$ because the latter may not lie in BQP when the guiding state is not uniformly quantumly preparable. However, this distinction will not affect our hardness results, as all guiding states used in our constructions will be quantumly preparable.

All the upcoming constructions will rely on a specific type of guiding state, which we call *semi-classical encoded states*.

4.2.2. Definition (Semi-classical encoded state). We say that a normalized state $|u\rangle \in \mathbb{C}^{2^m}$, for some $m = \mathcal{O}(n)$, is a *semi-classical encoded state* if there exists a subset $S \subseteq \{0,1\}^n$ with $|S| = \mathrm{poly}(n)$ and a set of isometries $V_1, V_2, \ldots, V_n$, where each V_i maps a single-qubit state to an r_i-qubit state, with $1 \leq r_i \leq r_{\max}$ for some $r_{\max} = \mathcal{O}(1)$, such that

$$|u\rangle = \frac{1}{\sqrt{|S|}} \sum_{x \in S} V_1(|x_1\rangle) \otimes V_2(|x_2\rangle) \otimes \cdots \otimes V_n(|x_n\rangle).$$

When $V_1 = V_2 = \cdots = V_n = \mathbb{I}_2$, we say that $|u\rangle$ is a *polynomially-sized subset state*.

It is straightforward to show that semi-classical encoded states satisfy the criteria of all the state classes introduced in Chapter 3.

4.2.3. Lemma. *Semi-classical encoded states are classically evaluatable, uniformly quantumly-preparable and allow perfect-sampling access.*

Proof:
Let $|\bar{u}\rangle = \frac{1}{\sqrt{|S|}} \sum_{x \in S} |x\rangle$, so that $|u\rangle = (V_1 \otimes V_2 \otimes \cdots \otimes V_n)\,|\bar{u}\rangle$, where $|u\rangle$ is any m-qubit semi-classical encoded state as defined in Definition 4.2.2. Just as in Appendix B, semi-classical encoded states have a classical description in standard English text from which membership in the class can be easily verified, so we again take the function $\mathrm{rep}(\cdot)$ to be any standard text-to-binary converter, e.g., ASCII. This way, we only need to check whether the algorithms $\mathcal{Q}_u^{\mathrm{query}}$, $\mathcal{Q}_u^{\mathrm{samp}}$, and $\mathcal{Q}_u^{\mathrm{local}}$, corresponding to a given class of states, exist.

Classical evaluability. By Proposition 3.5.6, classical evaluability is preserved under the application of local isometries. It thus suffices to show that $|\bar{u}\rangle$, which is a polynomially-sized subset state, is classically evaluatable. Since Condition (i) of Definition 3.5.5 holds directly by definition, we only need to verify Condition (ii).

We only need to show that $\mathcal{Q}_u^{\text{local}}$ exists. Consider any k-local observable O. We have

$$\langle u| O |u\rangle = \frac{1}{|S|} \sum_{i,j \in S} \langle i| O |j\rangle = \frac{1}{|S|} \sum_{i,j \in S} O_{i,j},$$

with $O_{i,j} = \langle i| O |j\rangle$. Since $|S| = \text{poly}(n)$, this sum can be evaluated efficiently given query access to the matrix elements of O. By Definition 3.1.1, any k-local observable is given as

$$O = P_\pi(\bar{O} \otimes \mathbb{I})P_\pi^{-1},$$

where $\bar{O}$ is a $2^k \times 2^k$ Hermitian matrix and P_π is a permutation matrix defined by $P_\pi |x_1 \cdots x_n\rangle = \left|x_{\pi(1)} \cdots x_{\pi(n)}\right\rangle$. Hence,

$$O_{i,j} = \langle i| P_\pi \left(\sum_{p,q} \bar{O}_{p,q} |p\rangle\langle q| \otimes \mathbb{I} \right) P_\pi^{-1} |j\rangle$$

can be evaluated efficiently in 2^k, and therefore so can $\langle u| O |u\rangle$.

Samplability. To verify Condition (i) of Definition 3.5.1, consider an arbitrary m-bit string z. For each $i \in [n]$, let $z_i \in \{0,1\}^{r_i}$ denote the portion of y corresponding to qubits associated with the image of V_i. Then we can write

$$\langle z|u\rangle = \frac{1}{\sqrt{|S|}} \sum_{x \in S} \prod_{i=1}^{n} \langle z_i| V_i(|x_i\rangle).$$

Each inner product $\langle z_i| V_i(|x_i\rangle)$ is a complex number that can be computed in constant time, since V_i maps 1-qubit states to $r_i \leq r_{\max} = \mathcal{O}(1)$ qubits and has a known description. Since $|S| = \text{poly}(n)$, the total sum can be evaluated in $\text{poly}(n)$ time.

For Condition (ii), define

$$P(y_0, \ldots, y_{i-1}) = \left\| (|y_0, \ldots, y_{i-1}\rangle\langle y_0, \ldots, y_{i-1}| \otimes \mathbb{I}) |u\rangle \right\|^2,$$

the probability that measuring the first i qubits in the computational basis yields outcome $y_0, \ldots, y_{i-1}$. For each $i \in [m]$, we can efficiently calculate $P(y_0, y_1, ..., y_{i-1})$ because $|S| = \text{poly}(n)$ and $V_1(|x_1\rangle) \otimes V_2(|x_2\rangle) \otimes \cdots \otimes V_n(|x_n\rangle)$ is a product state of $\mathcal{O}(1)$-qubit states. Then, we can also efficiently calculate the conditional probability

$$P(z \mid y_0, y_1, ..., y_{i-1}) = \frac{P(y_0, y_1, ..., y_{i-1}, z)}{P(y_0, y_1, ..., y_{i-1})}.$$

If the bits $y_0, y_1, ..., y_{i-1}$ have already been sampled, we compute $P(z|y_0, y_1, ..., y_{i-1})$ and sample the next bit by tossing the coin with bias $P(0|y_0, y_1, ..., y_{i-1})$. In this way, we can classically efficiently sample from the probability distribution that outputs x with probability $|\langle x|u\rangle|^2$.

Condition (iii) holds by definition (since it is a normalised state).

Uniformly quantum-preparability. To show that $|u\rangle$ is uniformly quantumly preparable, we must construct it to within any $\epsilon \geq 1/\text{poly}(n)$ using a universal gate set. The state $|\bar{u}\rangle$ can be prepared efficiently using the Grover–Rudolph procedure [GR02]. Each isometry V_i maps a single qubit to a state on $r_i \leq r_{\max} = \mathcal{O}(1)$ qubits. To implement V_i, we introduce $r_i - 1$ ancilla qubits in the state $|0\rangle$ and apply a unitary U_i on the resulting r_i-qubit register. We construct U_i using the Gram–Schmidt process: we take $V_i|0\rangle$ and $V_i|1\rangle$ as the first two basis vectors of an orthonormal basis for the 2^{r_i}-dimensional Hilbert space, and complete the basis arbitrarily. These vectors become the first two columns of U_i, with the remaining columns chosen to make U_i unitary. Since $r_i = \mathcal{O}(1)$, each U_i acts on a constant number of qubits and can be approximated to precision $1/\exp(n)$ using $\text{poly}(n)$ gates from a universal gate set [BBC+95]. The overall preparation procedure is as follows: prepare $|\bar{u}\rangle|0\cdots0\rangle$, and then apply $U_1 \cdots U_n$ to obtain $|u\rangle$ (up to any desired precision $\geq 1/\exp(n)$). □

In [GL22] it was proven that $\text{GLH}(k, a, b, \zeta)$ is BQP-hard for the following parameter settings.[5]

4.2.4. Theorem (From [GL22]). *For any* $\zeta \in \left(0, 1/\sqrt{2} - \Omega\left(1/\text{poly}(n)\right)\right)$, *there exist parameters* $a, b \in [0,1]$ *with* $b - a \geq 1/\text{poly}(n)$ *such that* $\text{GLH}(6, a, b, \zeta)$ *is* BQP*-hard.*

We briefly sketch Gharibian and Le Gall's original construction [GL22] used to prove Theorem 4.2.4. Let $A = (A_{\text{YES}}, A_{\text{NO}})$ be a promise problem in BQP, and let $x \in \{0,1\}^n$ be an input. Let $V = V_T \dots V_1$ be a P-uniform generated quantum circuit consisting of 1- and 2-qubit gates V_i, deciding A. More precisely, V takes an n-qubit input register A and an $r = \text{poly}(n)$-qubit work register B, and outputs, upon measurement, a 1 on the first qubit with probability at least c (resp. at most s) if $x \in A_{\text{YES}}$ (resp. $x \in A_{\text{NO}}$). By standard error reduction, we can assume without loss of generality that $c = 1 - 2^{-n}$ and $s = 2^{-n}$.

We recall Kitaev's original 5-local clock Hamiltonian from Section 3.2.1 $H_{\text{FK},5}$ (see Eq. (3.5)), but *omit* the witness register, since we are dealing with a BQP verifier. Also, in a slight abuse of notation to simplify the presentation, we (i) write H_t in terms of the clock states before the operator transformation Φ and do the same for the time states $|t\rangle$ and the transformation ϕ (see again Section 3.2.1)

[5]The proof in [GL22] uses a polynomially-sized subset state as a guiding state, which is classically evaluatable, uniformly quantumly preparable and allows perfect sampling access.

and (ii) write H_{in} in terms of global projectors onto the basis state as it does not change any of the analysis. The Hamiltonian $H_{\text{FK},5}$ is then given as a sum of the following terms:

$$\begin{aligned}
H_{\text{in}} &:= (\mathbb{I} - |x\rangle\langle x|)_A \otimes (\mathbb{I} - |0\ldots0\rangle\langle 0\ldots0|)_B \otimes |0\rangle\langle 0|_C\,,\\
H_{\text{out}} &:= |0\rangle\langle 0|_{\text{out}} \otimes |T\rangle\langle T|_C\,,\\
H_{\text{stab}} &:= \sum_{j=1}^{T} |0\rangle\langle 0|_{C_j} \otimes |1\rangle\langle 1|_{C_{j+1}}\,,\\
H_{\text{prop}} &:= \sum_{t=1}^{T} H_t \qquad \text{where}\\
H_t &:= -\frac{1}{2}U_t \otimes |t\rangle\langle t-1|_C - \frac{1}{2}U_t^\dagger \otimes |t-1\rangle\langle t|_C + \frac{1}{2}\mathbb{I} \otimes |t\rangle\langle t|_C + \frac{1}{2}\mathbb{I} \otimes |t-1\rangle\langle t-1|_C\,.
\end{aligned} \tag{4.1}$$

Here C denotes the clock register consisting of $T = \text{poly}(n)$ qubits. From Chapter 3, we know that the ground state energy of the Hamiltonian $H_{\text{FK},5} = H_{\text{in}} + H_{\text{out}} + H_{\text{clock}} + H_{\text{prop}}$ satisfies the following:

- If U accepts with at least probability c, then $\lambda_0(H_{\text{FK},5}) \leq \frac{1-c}{T}$.
- If U accepts with at most probability s, then $\lambda_0(H_{\text{FK},5}) \geq \Omega(\frac{1-\sqrt{s}}{T^3})$.

$H_{\text{FK},5}$ can be split into two separate terms:

$$\begin{aligned}
H_0 &:= H_{\text{in}} + H_{\text{stab}} + H_{\text{prop}}\\
H_1 &:= H_{\text{out}}
\end{aligned}$$

for which we know that the history state (which is unique as there is no witness register)

$$|\eta\rangle = \frac{1}{\sqrt{T+1}} \sum_{t=0}^{T} U_t \ldots U_1 |x\rangle_A |0\ldots0\rangle_B |t\rangle_C\,, \tag{4.2}$$

spans the null space of H_0. Consider the following guiding state, which is a polynomial-sized subset state:

$$|u\rangle = \frac{1}{\sqrt{N+1}} \sum_{t=0}^{N} |x\rangle_A |0\ldots0\rangle_B |t\rangle_C\,.$$

In general, this guiding state has at most $\mathcal{O}(1/(TN))$ fidelity with the history state and therefore an even smaller fidelity with the actual ground state of $H_{\text{FK},5}$. The idea is to transform $H_{\text{FK},5}$ into a new Hamiltonian H' such that the guiding state $|u\rangle$ achieves fidelity at least ζ with the ground space in both the YES- and NO-cases. To achieve this, Gharibian and Le Gall use the following tricks:

- Since the history state in Eq. (4.2) uniquely spans the null space of H_0 (of which all terms are positive semi-definite) in the BQP setting, and a bound is known on the energy of all non-zero eigenstates, weighing H_0 with a large (but only polynomial) prefactor Γ allows one to increase the fidelity of the actual ground state with the history state.[6]

- By *pre-idling* the circuit V that is in the clock Hamiltonian—that is, applying M identity gates before the first actual gate—the fidelity between $|u\rangle$ and the history state can be increased. This increases the number of gates from T to $T' = T + M$. Denote the weighted and pre-idled Hamiltonian as $\hat{H}_{\mathrm{FK},5}$. Also, define

$$\hat{c} := \frac{1-c}{T'+1} \quad \text{and} \quad \hat{s} := \Omega\left(\frac{1-\sqrt{s}}{T'^3}\right).$$

- Finally, by block-encoding $\hat{H}_{\mathrm{FK},5}$ into a larger Hamiltonian H_6, which acts on $n+r+T+1$ qubits (adding another single-qubit register D), one can add another Hamiltonian (in their case a scaled identity term) in another block such that the ground space in case of a NO-case is trivial, only increasing the locality of the Hamiltonian in the construction by 1. By setting this specific qubit in the guiding state to the $|+\rangle$ state, one ensures that it has fidelity with both the NO- and YES-cases.

The final Hamiltonian before renormalisation is then

$$H_6 := \frac{\hat{c}+\hat{s}}{2}\mathbb{I}_{ABC} \otimes |0\rangle\langle 0|_D + \hat{H}_{\mathrm{FK},5} \otimes |1\rangle\langle 1|_D\,, \tag{4.3}$$

where $\hat{H}_{\mathrm{FK},5} = \Gamma\left(H_{\mathrm{in}} + H_{\mathrm{stab}} + H_{\mathrm{prop}}\right) + H_{\mathrm{out}}$. The guiding state becomes

$$|u\rangle := |x\rangle_A |0\ldots 0\rangle_B \left(\frac{1}{\sqrt{M}}\sum_{t=1}^{M}|t\rangle\right)_C |+\rangle_D\,. \tag{4.4}$$

Since the overall construction starts from a 5-local Hamiltonian, the block-encoding step increases the locality to 6 and restricts the fidelity to be at most $1/2 - 1/\mathrm{poly}(n)$.

4.3 Increasing the allowed fidelity

Our first goal is to improve Theorem 4.2.4 in terms of the allowed fidelity, i.e., ζ-parameter. Observe that the optimal guiding state, i.e., the state that achieves the maximum fidelity with the ground space in both the YES- and NO-cases, for

[6] This is a similar idea to the small-penalty clock construction of Lemma 3.2.3.

the Hamiltonian of Eq. (4.3) takes the form $|\phi\rangle \otimes |+\rangle$ for a certain choice of $|\phi\rangle$. This shows that, in this construction, the fidelity cannot exceed $1/2$. To overcome this limitation, we employ the perturbative techniques of [KKR06, BDL11].

In particular, we apply first-order perturbation theory, specifically using the general Schrieffer–Wolff transformation framework from [BDL11]. The main idea is to introduce a large energy penalty term that excludes all low-energy states not resembling "history states". We then show that a suitable guiding state can be chosen as the semi-classical subset state from Eq. (4.4), but without the $|+\rangle$-state in the D-register. This is made possible by the fact that the ground state of our Hamiltonian is both gapped and unique. The uniqueness arises because we are reducing from BQP (rather than QMA): in other words, there is no "QMA proof" to be inserted into the history state construction, so there is a unique low-energy history state. By employing perturbation theory, we are able to *directly* approximate the ground state with a guiding state in *both* the YES- and NO-cases. This contrasts with the block-encoding approach of [GL22], which used equally weighted orthogonal subspaces to *separately* encode the YES- and NO-cases.

4.3.1 Schrieffer-Wolff transformation for non-degenerate gapped ground spaces

We now briefly introduce the Schrieffer–Wolff transformation and its approximation [BDL11], which is the key technique in the proof. We consider only the case where the unperturbed Hamiltonian has a one-dimensional ground space.

Let H_0 be a Hamiltonian with a one-dimensional ground space spanned by $|g_0\rangle$, whose energy is zero. Assume that the smallest non-zero eigenvalue of H_0 is greater than one. Consider the following (perturbed) Hamiltonian for some $\Gamma > 0$:

$$H = \Gamma H_0 + V.$$

We assume throughout that $\|V\| \leq \Gamma/2$. Under this assumption, there exists a unique eigenvector (denoted by $|g\rangle$) of H whose eigenvalue lies in the interval $[-\Gamma/2, \Gamma/2]$ (Lemma 3.1 of [BDL11]).

The Schrieffer–Wolff (SW) transformation is defined as a unitary U_{SW} that maps the ground space of H to that of H_0; that is, $U_{\mathrm{SW}}|g\rangle = |g_0\rangle$. The Hamiltonian

$$H_{\mathrm{eff}} = \Pi_0 U_{\mathrm{SW}}(\Gamma H_0 + V)U_{\mathrm{SW}}^{\dagger}\Pi_0,$$

is called the effective low-energy Hamiltonian, where Π_0 is the projector onto the ground space of H_0.[7] The eigenvector of H_{eff} is $|g_0\rangle$, and its eigenvalue coincides with the eigenvalue of $|g\rangle$ under H.

Next, we show how to approximate U_{SW} and H_{eff}. We will only need the simplest first-order approximation for our purposes. Assuming that $\|V\| \leq \Gamma/16$,

[7] Since the ground state is unique, we automatically have $\Pi_{\mathrm{gs}} = \Pi_0$.

it is known that

$$\|\mathbb{I} - U_{\mathrm{SW}}\| = \mathcal{O}(\Gamma^{-1}\|V\|) \tag{4.5}$$

and

$$\|H_{\mathrm{eff}} - \Pi_0 V \Pi_0\| = \mathcal{O}(\Gamma^{-1}\|V\|^2) \tag{4.6}$$

hold (Lemma 3.4 of [BDL11], Lemma 4 of [BH17]). This means that $\mathbb{I}$ and $\Pi_0 V \Pi_0$ serve as first-order approximations to U_{SW} and H_{eff}, respectively. The derivation and expressions for the higher-order terms can be found in [BDL11]. From Eq. (4.5), it follows that

$$\| \, |g\rangle - |g_0\rangle \, \| = \left\| (\mathbb{I} - U_{\mathrm{SW}}^\dagger) \, |g_0\rangle \right\| = \mathcal{O}(\Gamma^{-1}\|V\|). \tag{4.7}$$

It also follows from Eq. (4.6) that the ground state energy of H differs by at most $\mathcal{O}(\Gamma^{-1}\|V\|^2)$ from the eigenvalue of $H_{\mathrm{eff},1} \coloneqq \Pi_0 V \Pi_0$ (restricted to the subspace spanned by $|g_0\rangle$).

4.3.2 Application to GLH

With the tools from Section 4.3.1 in hand, we are now ready to make the first improvement to Theorem 4.2.4.

4.3.1. Proposition. *For any $\zeta \in (0, 1 - 1/\mathrm{poly}(n))$, there exist $a, b \in [0,1]$ with $b - a \geq 1/\mathrm{poly}(n)$ such that GLH$(5, a, b, \zeta)$ is* BQP*-hard. Moreover, it remains* BQP*-hard under the following two additional promises:*

1. *H has a non-degenerate ground state, separated from the first excited state by a gap $\gamma \geq 1/\mathrm{poly}(n)$ in both the* YES- *and* NO*-cases.*

2. *The guiding state is restricted to be a semi-classical subset state.*

Proof:
Let $A = (A_{\mathrm{YES}}, A_{\mathrm{NO}})$ be a promise problem in BQP, and let $\mathcal{U} = \{U_n : n \in \mathbb{N}\}$ be a corresponding P-uniform family of BQP verifiers that decide A. Fix some input size n, and let $x \in \{0,1\}^n$ be an input. As always, we write $U = U_n \in \mathcal{U}$ for the corresponding BQP verifier to simplify notation. Let $U = V_T V_{T-1} \cdots V_1$ be a decomposition of U into $T = \mathrm{poly}(n)$ one- and two-qubit gates. The circuit U acts on $|x\rangle_A \otimes |0 \cdots 0\rangle_B$, where A denotes the n-qubit input register and B denotes the polynomial-sized ancilla register. By measuring the output qubit of $U \, |x\rangle_A \otimes |0 \cdots 0\rangle_B$, the quantum verifier outputs 1 with probability at least c if $x \in A_{\mathrm{YES}}$, and at most s if $x \in A_{\mathrm{NO}}$. We may assume $c = 1 - 2^{-n}$ and $s = 2^{-n}$ via standard error reduction for BQP.

Consider the pre-idled quantum verifier $\tilde{U} \coloneqq U\mathbb{I} \cdots \mathbb{I}$, where $\mathbb{I}$ is the identity gate. The circuit $\tilde{U}$ consists of $M \coloneqq T + N$ gates $\tilde{V}_t$, where $N = \mathrm{poly}(n)$ denotes the number of idling steps (chosen appropriately later). We now construct the

Hamiltonian using Kitaev's 5-local circuit-to-Hamiltonian construction, with an additional scaling factor (we will renormalise the final Hamiltonian at the end of the proof to meet the norm condition of Definition 4.2.1):

$$H \coloneqq \Gamma(H_{\text{in}} + H_{\text{prop}} + H_{\text{stab}}) + H_{\text{out}}, \tag{4.8}$$

where the terms are defined as in Eq. (4.1). The non-degenerate, zero-energy ground space of $H_0 \coloneqq H_{\text{in}} + H_{\text{prop}} + H_{\text{stab}}$ is spanned by the history state

$$|\eta\rangle \coloneqq \frac{1}{\sqrt{M+1}} \sum_{t=0}^{M} \tilde{V}_t \cdots \tilde{V}_1 \, |x\rangle_A \otimes |0 \cdots 0\rangle_B \otimes |t\rangle_C \,. \tag{4.9}$$

In [GY19, Lemma 2.2], it is shown that the smallest non-zero eigenvalue of H_0 is larger than $\pi^2/(64M^2)$.[8]

We apply the Schrieffer–Wolff transformation from Section 4.3.1 to H, taking Γ sufficiently large. Note that $H_{\text{out}} = |0\rangle\langle 0| \otimes \mathbb{I} \otimes |M\rangle\langle M|$, and $\|H_{\text{out}}\| = 1$. We take

$$\Gamma \geq \frac{16 \cdot 64M^2}{\pi^2}.$$

Then, H has a one-dimensional ground space spanned by a ground state $|g\rangle$. We now analyse the fidelity between $|g\rangle$ and $|\eta\rangle$, and the eigenvalue of $|g\rangle$ in the YES- and NO-cases.

Analysis of the fidelity. By Eq. (4.7), we have

$$\| \, |g\rangle - |\eta\rangle \, \| = \mathcal{O}\left(\frac{M^2}{\Gamma}\right).$$

Define the state

$$|u\rangle \coloneqq \frac{1}{\sqrt{N}} \sum_{t=1}^{N} |x\rangle_A \otimes |0 \cdots 0\rangle_B \otimes |t\rangle_C \,,$$

which is a semi-classical subset state. This satisfies

$$|\langle u|\eta\rangle|^2 = \frac{N}{T+N+1}.$$

Therefore, for any positive polynomial r, we can take sufficiently large $N, \Gamma \in \text{poly}(n)$ so that

$$|\langle u|g\rangle|^2 \geq 1 - \frac{1}{r(n)}.$$

[8] Strictly speaking, we do not need this reference as we already argued in Section 3.2.1 that the smallest non-zero eigenvalue of this H_0 also is lower bounded by Lemma 3.2.1. Of course, this changes nothing in the proof except for constant factors what we do not care about.

Analysis of energies. We now analyse the ground state energy of H in the YES- and NO-cases. The first-order effective Hamiltonian is

$$H_{\text{eff},1} = |\eta\rangle\langle\eta| \, H_{\text{out}} \, |\eta\rangle\langle\eta|$$

with $|\eta\rangle$ the history state from Eq. (4.9). We have

$$\langle\eta| \, H_{\text{out}} \, |\eta\rangle = \frac{1}{M+1} \langle x, 0| \, U^\dagger (|0\rangle\langle 0|_{\text{out}} \otimes \mathbb{I}) U \, |x, 0\rangle \, .$$

The eigenvalue of $H_{\text{eff},1}$ is given by $\langle\eta| \, H_{\text{out}} \, |\eta\rangle$, which is within $\mathcal{O}(M^2/\Gamma) = 1/\text{poly}(n)$ of the ground state energy of H by Eq. (4.6). It can be verified that

$$\langle\eta| \, H_{\text{out}} \, |\eta\rangle \leq \frac{1-c}{M+1} \quad \text{if } x \in A_{\text{YES}},$$

and

$$\langle\eta| \, H_{\text{out}} \, |\eta\rangle \geq \frac{1-s}{M+1} \quad \text{if } x \in A_{\text{NO}}.$$

As noted earlier, we assume $c = 1 - 2^{-n}$ and $s = 2^{-n}$. Hence, the ground state energy a of H lies in the range $0 \pm \mathcal{O}(M^2/\Gamma)$ in the YES-case, and the ground state energy b lies in the range $1/(M+1) \pm \mathcal{O}(M^2/\Gamma)$ in the NO-case.

The spectral gap. To estimate the spectral gap in the NO-case: the smallest non-zero eigenvalue of H is at least $\Gamma\pi^2/(64M^2) - 1$. The ground state energy lies within

$$\frac{1-2^{-n}}{M+1} \pm \mathcal{O}\left(\frac{M^2}{\Gamma}\right).$$

Therefore, the spectral gap is at least

$$\mathcal{O}\left(\frac{\Gamma}{M^2}\right) - 1 - \left(\frac{1-2^{-n}}{M+1} + \mathcal{O}\left(\frac{M^2}{\Gamma}\right)\right).$$

Since the ground state energy in the YES-case is lower, taking sufficiently large $\Gamma = \text{poly}(n)$ ensures that H has an inverse-polynomially bounded spectral gap and $b - a \geq 1/\text{poly}(n)$. Finally, we may renormalise H by a polynomial factor to satisfy the norm condition of Definition 4.2.1, which completes the proof. □

In terms of the order of scaling, showing hardness for $\zeta = 1 - 1/\text{poly}(n)$ is the best one can hope for, as any $\zeta = 1 - o(1/\text{poly}(n))$ would imply that, for sufficiently large n, the guiding state becomes close enough to the ground state that the $\text{GLH}(k, a, b, \zeta)$ instance could simply be solved by classically computing $\langle u| \, H \, |u\rangle$, which puts the problem in P.

Before improving the result in terms of its other input parameter, that is, the locality, we first make a detour to define some variants and generalisations of $\text{GLH}(k, a, b, \zeta)$. The reason is that the techniques used to improve the parameters (see Section 4.6) also apply to these related problems.

4.4 Guidable local Hamiltonian problems

In the guided local Hamiltonian problem, the guiding state is part of the *input* to the problem. A natural question then arises: how does the hardness of the problem change if one is only promised that a guiding state *exists*, but it is no longer provided as part of the input? This leads to a modification of the guided local Hamiltonian problem, which we refer to as the *guidable local Hamiltonian problem*. Additionally, we distinguish between two separate families of problems to capture the power of classical versus quantum heuristics.[9]

4.4.1. Definition (Guidable local Hamiltonian problems).

Input: A description of a collection of k-local Hermitian operators $\{H_i\}_{i\in[m]}$, which define an n-qubit k-local Hamiltonian $H = \sum_{i=1}^{m} H_i$ with $\|H\| \leq 1$, where $m = \text{poly}(n)$; two efficiently computable real numbers a, b such that $b > a$; and an efficiently computable fidelity parameter $\zeta \in [0, 1]$.

Promise: We have that either $\lambda_0(H) \leq a$ or $\lambda_0(H) \geq b$ holds.

Additional promise: For each problem class, at least one of the following additional promises holds:

1. There exists a classically evaluatable state $|u\rangle \in \mathbb{C}^{2^n}$ with the property that $\left\|\Pi_{\text{gs}}^{H} |u\rangle\right\|_2^2 \geq \zeta$. Then the problem is called the *classically guidable local Hamiltonian problem* ($\mathsf{CGaLH}(k, a, b, \zeta)$).
2. There exists a uniformly quantumly-preparable state $|u\rangle \in \mathbb{C}^{2^n}$ with the property that $\left\|\Pi_{\text{gs}}^{H} |u\rangle\right\|^2 \geq \zeta$. Then the problem is called the *quantumly guidable local Hamiltonian problem* ($\mathsf{QGaLH}(k, a, b, \zeta)$).

If $|u\rangle$ is both classically evaluatable and uniformly quantumly-preparable, we call the problem the *classically guidable and quantumly-preparable local Hamiltonian problem* ($\mathsf{CGaLH}^*(k, a, b, \zeta)$).

Goal: Decide whether $\lambda_0(H) \leq a$ or $\lambda_0(H) \geq b$.

We observe that $\mathsf{QGaLH}(k, a, b, \zeta)$ is closely related to the low-complexity low-energy states problem introduced in [WJB03]. In fact, proving QCMA-hardness for $\mathsf{QGaLH}(k, a, b, \zeta)$ for some $\zeta \geq 1/\text{poly}(n))$, $b - a \geq 1/\text{poly}(n)$, implies QCMA-hardness of the low-complexity low-energy states problem for the same value of k. However, the converse does not hold directly. The reason is that, in the low-complexity low-energy States problem, it may be the case that all low-energy

[9] Something similar could have been done for $\text{GLH}(k, a, b, \zeta)$, but this would lead to weaker complexity-theoretic implications when formulating certain statements regarding classically evaluatable states (see for example Appendix C.2.1).

states of the Hamiltonian H in the YES-case are preparable by polynomial-size quantum circuits, but nonetheless have only negligible overlap with the ground space.

4.4.1 Proof of QCMA-hardness

We follow a similar proof structure to that used in the BQP-hardness proofs of the Guided local Hamiltonian problem in Section 4.3. However, several obstacles prevent us from directly adopting the same proof in the QCMA setting, i.e., when starting with a verification circuit U_n from a P-uniform family $\mathcal{U} = \{U_n : n \in \mathbb{N}\}$ of QCMA verifiers. This is mainly due to the fact that U_n, unlike a BQP verification circuit, includes an additional input register for the witness. As a result, there are many valid history states (i.e., zero-eigenvectors of the Hamiltonian H_0), which gives us less control over (and less information about) the actual ground state of the Hamiltonian produced by the circuit-to-Hamiltonian construction.

To work around this, we use several techniques in the new construction. First, we apply the CNOT trick introduced in [WJB03] to "force" all witnesses to be classical. Second, we use a result from [ABOBS22], which shows that there exists a randomised reduction from a QCMA protocol with verification circuit U_n to one with a verification circuit $\tilde{U}_n$, such that in the YES-case, there is a unique accepting witness. Next, we apply the small-penalty circuit-to-Hamiltonian mapping from [DGF22] (see Section 3.2.1), which, together with error reduction on the verification circuit $\tilde{U}_n$, gives us fine control over the bounds on the energies in the low-energy subspace of the Hamiltonian.

Combining this with the aforementioned randomised reduction, we find that the ground space of H is now one-dimensional and can be made to have exponentially high fidelity with the history state corresponding to the uniquely accepting witness in the YES-case. This allows us to construct a corresponding polynomial-sized subset state (which we show to be classically evaluatable and quantumly preparable) that has good fidelity with this history state, and to use it as our guiding state. We also apply the pre-idling and block-encoding techniques from [GL22] to increase fidelity with the guiding state and to handle the NO-case, respectively.

For completeness, we begin by providing a proof of the CNOT-*trick*, which was used without proof in [WJB03].

4.4.2. Lemma (The CNOT-trick). *Let $p, q\colon \mathbb{N} \to \mathbb{N}$ be polynomially bounded functions. Let U 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, initialised to $|0\rangle^{\otimes q(n)}$. Denote Π_0 as the projection onto the subspace where the first qubit is zero. Let Q be the Marriott–Watrous operator of the circuit,*

defined as

$$Q = \left(\langle x| \otimes \mathbb{I}_B \otimes \langle 0|^{\otimes q(n)}\right) U^\dagger \Pi_0 U \left(|x\rangle \otimes \mathbb{I}_B \otimes |0\rangle^{\otimes q(n)}\right). \tag{4.10}$$

Consider an additional $p(n)$-qubit workspace register D, initialised to $|0\rangle^{\otimes p(n)}$, on which U does not act. Then, by prepending U with $p(n)$ CNOT *operations, each controlled by a qubit in register B and targeting the corresponding qubit in register D, the Marriott–Watrous operator corresponding to the new circuit becomes diagonal in the computational basis.*

Proof:
Let U_{CNOT} denote the $2p(n)$-qubit operation that acts on the two registers B and D, and that applies, for each $l \in [p(n)]$, a CNOT gate controlled by qubit l in register B and targeting qubit l in register D. Consider the new verifier circuit $\tilde{U} = UU_{\text{CNOT}}$, which acts on registers A, B, C and D, with corresponding Marriott–Watrous operator $\tilde{Q}$. Let $|i\rangle$ and $|j\rangle$, for $i, j \in \{0,1\}^{p(n)}$, be arbitrary computational basis states. Then,

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

where we used that U and Π_0 act trivially on register D. Hence, the operator $\tilde{Q}$ is diagonal in the computational basis, with diagonal entries taken from the diagonal of Q. □

We now recall the small-penalty clock construction lemma from Section 3.2.1, which we restate for completeness.

4.4.3. Lemma (Small-penalty clock construction [DGF22, Lemma 26]). *Let $\mathcal{U} = \{U_n : n \in \mathbb{N}\}$ be a* P*-uniform family of* QMA *verification circuits. Let n be the input size and consider an input $x \in \{0,1\}^n$. Suppose U_n consists of $T = \text{poly}(n)$ gates from some universal gate-set using at most 2-local gates. Denote $P(\psi)$ for the probability that U_n accepts $(x, |\psi\rangle)$, and let $H_{\text{SPCC,3}}$ be the corresponding 3-local Hamiltonian from the circuit-to-Hamiltonian mapping in [KR03] with an ϵ-factor in front of H_{out}. Then there exists a constant $c > 0$, such that for all $0 < \epsilon \leq c/T^3$, we have that within the low-energy subspace $\mathcal{S}_\epsilon$ of $H_{\text{SPCC,3}}$, i.e.,*

$$\mathcal{S}_\epsilon = \text{span}\{|\Phi\rangle : \langle\Phi| H_{\text{SPCC,3}} |\Phi\rangle \leq \epsilon\}$$

the eigenvalues λ_i satisfy

$$\lambda_i \in \left[\epsilon\frac{1-P(|\psi_i\rangle)}{T+1} - \mathcal{O}(T^3\epsilon^2), \epsilon\frac{1-P(|\psi_i\rangle)}{T+1} + \mathcal{O}(T^3\epsilon^2)\right], \tag{3.7}$$

where $\{|\psi_i\rangle\}$ are the eigenstates of the Mariott-Watrous operator of the circuit U_n, given by

$$Q_n = \left(\langle x| \otimes \mathbb{I} \otimes \langle 0|^{\otimes q(n)}\right) U_n^\dagger \left(|0\rangle\langle 0| \otimes \mathbb{I}\right) U_n \left(|x\rangle \otimes \mathbb{I} \otimes |0\rangle^{\otimes q(n)}\right).$$

We emphasise again that the proof of the small-penalty clock construction also relies on the Schrieffer–Wolff transformation (Section 4.3.1), and essentially uses the same ideas as in Proposition 4.3.2. We are now ready to present a theorem analogous to Proposition 4.3.1, but for the guidable local Hamiltonian problem.

4.4.4. Theorem. *For any $\zeta \in (0, 1 - 1/\mathrm{poly}(n))$, there exist $a, b \in [0,1]$ with $b - a \geq 1/\mathrm{poly}(n)$ such that $\mathsf{CGaLH}^*(4, a, b, \zeta)$ is QCMA-hard under randomised reductions. Moreover, it remains QCMA-hard under randomised reductions with the following two additional promises:*

1. *H has a non-degenerate ground state, separated from the first excited state by a spectral gap $\gamma \geq 1/\mathrm{poly}(n)$ in both the cases $\lambda_0(H) \leq a$ and $\lambda_0(H) \geq b$.*

2. *The guiding state is restricted to be a semi-classical subset state.*

Proof:
We begin by presenting a "basic reduction" which uses computational basis states as guiding states: these trivially satisfy the conditions of Definition 3.5.5. We first establish completeness and soundness for this construction, and then show how to improve the fidelity parameters.

The basic reduction. Let $\langle U_n, p_1, p_2\rangle$ be a trivial QCMA-hard circuit verification promise problem. By the result of [ABOBS22], there exists a randomised reduction to a UQCMA instance $\langle \hat{U}_n, \hat{p}_1, \hat{p}_2\rangle$, i.e., a QCMA instance with a unique accepting witness in the YES-case, with completeness-soundness gap at least $1/q(n)$ for some polynomial q. Let the witness $y \in \{0,1\}^{p(n)}$, and suppose the verifier uses at most $T = \mathrm{poly}(n)$ gates. We apply two transformations:

1. The CNOT trick: we "force" the witness to be classical by introducing an extra register and copying each bit of y via CNOT gates before executing the verifier. By Lemma 4.4.2, this diagonalises the Marriott–Watrous operator corresponding to the modified circuit in the computational basis.

2. Error reduction: we apply strong error reduction as in [MW05], which allows repeated verification with the same witness. It follows from [MW05, Theorem 3.3] that, for any quantum verifier $\hat{U}_n$ using $T = \text{poly}(n)$ two-qubit gates and a $p(n)$-qubit witness, there exists a circuit $\tilde{U}_n$ that also uses $p(n)$-qubit witnesses and achieves completeness $1 - 2^{-r}$ and soundness 2^{-r}, for any polynomially bounded function r, using $\tilde{T} = \mathcal{O}(q^2 r T)$ gates.

Let $\langle \tilde{U}_n, \tilde{c}, \tilde{s} \rangle$ denote the resulting verification protocol. The circuit $\tilde{U}_n$ acts on input register A, witness register W, and ancilla register B, and we denote the unique accepting witness (if one exists) by y^*. We define $P(y) := \Pr\left[\tilde{U}_n \text{ accepts } y\right]$. The number of gates is now $\tilde{T} = \mathcal{O}(q^2 r T)$. Consider the following 4-local Hamiltonian:

$$H = H_{\text{YES}} \otimes |0\rangle\langle 0|_D + H_{\text{NO}} \otimes |1\rangle\langle 1|_D ,$$

where $H_{\text{YES}} = H_{\text{SPCC},3}$ is the Hamiltonian from Lemma 3.2.3 with small penalty parameter ϵ, and H_{NO} is given by

$$H_{\text{NO}} = \sum_{i=0}^{R-1} |1\rangle\langle 1|_i + b\mathbb{I},$$

with R is the number of qubits in the registers A, W, B and the clock register C, and $b > 0$ a tunable parameter.

We observe that H_{NO} has a unique ground state with energy b given by the all zeros state, and the spectrum increases in steps of 1 (and so it in particular has a spectral gap of 1). We also have that $\|H_{\text{NO}}\| = R + b = \text{poly}(n)$. As a guiding state in the YES-case, we take:

$$|u_{\text{YES}}\rangle = |x\rangle_A |y^*\rangle_W |0 \cdots 0\rangle_B |0\rangle_C |0\rangle_D .$$

The fidelity with the history state corresponding to y^* is then

$$|(\langle \eta(y^*)| \langle 0|_D) |u_{\text{YES}}\rangle|^2 = \frac{1}{\tilde{T} + 1}.$$

In the NO-case, we will show that the state

$$|u_{\text{NO}}\rangle = |0 \ldots 0\rangle_{AWBC} |1\rangle_D , \tag{4.11}$$

will be in fact the ground state. We will now show that setting

$$b := \mathcal{O}(1/\tilde{T}^7), \quad \epsilon := \mathcal{O}(1/\tilde{T}^5),$$

our reduction achieves the desired result.

Completeness. From Lemma 3.2.3, the eigenvalue $\lambda(y)$ corresponding to the witness $y = y^*$ can be upper bounded as

$$\lambda(y^*) \leq \epsilon \frac{2^{-r}}{\tilde{T}+1} + \mathcal{O}(\tilde{T}^3\epsilon^2).$$

For our choice of ϵ, and assuming $r \geq 1$ (we will later take r to be much larger), we also have that for any other witness $y \neq y^*$,

$$\lambda(y) \geq \epsilon \frac{1-2^{-r}}{\tilde{T}+1} - \mathcal{O}(\tilde{T}^3\epsilon^2) = \Omega\left(\frac{1}{\tilde{T}^6}\right).$$

This implies that the ground state $|g_{\text{YES}}\rangle$ of H_{YES} is unique. The spectral gap of H_{YES} can be bounded as

$$\gamma(H_{\text{YES}}) \geq \epsilon \frac{1-2^{-r+1}}{\tilde{T}+1} - \mathcal{O}(\tilde{T}^3\epsilon^2) = \Omega\left(\frac{1}{\tilde{T}^6}\right). \tag{4.12}$$

Next, we analyse the fidelity of the history state $|\eta(y^*)\rangle$ with $|g_{\text{YES}}\rangle$. First, the energy of $|\eta(y^*)\rangle$ is upper bounded by

$$\langle\eta(y^*)| H_{\text{YES}} |\eta(y^*)\rangle \leq \epsilon \frac{2^{-r}}{\tilde{T}+1} = \mathcal{O}\left(\frac{2^{-r}}{\tilde{T}^6}\right), \tag{4.13}$$

which follows directly from the history state construction of [KR03] and the fact that $P(y^*) \geq 1-2^{-r}$. Let us now write $|\eta(y^*)\rangle$ in the eigenbasis of H_{YES} as

$$|\eta(y^*)\rangle = \alpha |g_{\text{YES}}\rangle + \sqrt{1-\alpha^2}\left|g_{\text{YES}}^{\perp}\right\rangle,$$

for some real number $\alpha \in [0,1]$, where $|\psi\rangle$ is (up to an irrelevant phase factor) the ground state of H_{YES}, and $\left|g_{\text{YES}}^{\perp}\right\rangle$ is orthogonal to $|g_{\text{YES}}\rangle$. As already established, the energy of $|\eta(y^*)\rangle$ is upper bounded by

$$\langle\eta(y^*)| H_{\text{YES}} |\eta(y^*)\rangle \leq \epsilon \frac{2^{-r}}{\tilde{T}+1} = \mathcal{O}\left(\frac{2^{-r}}{\tilde{T}^6}\right).$$

Meanwhile, since H_{YES} is positive semi-definite, we also have the lower bound

$$\langle\eta(y^*)| H_{\text{YES}} |\eta(y^*)\rangle = \alpha^2 \langle g_{\text{YES}}| H_{\text{YES}} |g_{\text{YES}}\rangle + (1-\alpha^2)\left\langle g_{\text{YES}}^{\perp}\right| H_{\text{YES}} \left|g_{\text{YES}}^{\perp}\right\rangle \geq \Omega\left(\frac{1-\alpha^2}{\tilde{T}^6}\right).$$

Combining these two bounds yields

$$\alpha^2 = |\langle\eta(y^*)|g_{\text{YES}}\rangle|^2 \geq 1 - \mathcal{O}(2^{-r}), \tag{4.14}$$

which can be made at least $1-2^{-\Omega(\sqrt{\tilde{T}})}$ by taking $r = \mathcal{O}(q^2T) \gg 1$ (recall that $\tilde{T} = \mathcal{O}(q^2rT)$). For this choice of r, we also have that the ground-state energy

$< b$ by Eq. (4.13) (the history state energy provides an upper bound), so the ground state $|g\rangle$ of H has only support on $|0\rangle_D$ in the final register.

We now turn to the fidelity between the guiding state $|u_{\text{YES}}\rangle$ and the ground state $|g\rangle$ of H. We find

$$\begin{aligned}|\langle u_{\text{yes}}|g\rangle|^2 &\geq 1 - \left(\sqrt{1 - |\langle u_{\text{yes}}| (|\eta(y^*)\rangle |0\rangle)|^2} + \sqrt{1 - |(\langle \eta(y^*)| \langle 0|) |g\rangle|^2}\right)^2 \\ &\geq 1 - \left(\sqrt{1 - \frac{1}{\tilde{T}+1}} + 2^{-\Omega(\sqrt{\tilde{T}})}\right)^2 \\ &\geq \Omega\left(\frac{1}{\tilde{T}}\right).\end{aligned}$$

Soundness. All witnesses have acceptance probability at most $2^{-\Omega(\sqrt{\tilde{T}})}$, so $H_{\text{YES}} \succeq \Omega(1/\tilde{T}^6)$. For our choice of b, the ground state is $|u_{\text{no}}\rangle$ with energy $b = \Omega(1/\tilde{T}^7)$. Hence, the promise gap becomes

$$\delta = \Omega(1/\tilde{T}^7) - \mathcal{O}\left(\frac{2^{-r}}{\tilde{T}^6}\right) = \Omega(1/\tilde{T}^7) = 1/\text{poly}(n).$$

We will now use similar tricks as in Proposition 4.3.1 to improve the basic construction in terms of the fidelity range.

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, which we know is classically evaluatable and uniformly quantumly-preparable. However, in the YES-case, the ground state is a history state with only inverse polynomial fidelity with the state $|u_{\text{YES}}\rangle$. To work around this, we apply the same trick as in Section 4.3 by pre-idling the circuit with a polynomial number of identity gates, of which we denote the total number by N, and guiding state to

$$|u_{\text{YES}}^{\text{new}}\rangle = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} |x\rangle_A |y^*\rangle_W |0\ldots0\rangle_B |t\rangle_C |0\rangle_D , \tag{4.15}$$

which satisfies

$$|\langle u_{\text{YES}}^{\text{new}}| (|\eta(y^*)\rangle |0\rangle)|^2 = \frac{N}{N+\tilde{T}+1}.$$

Since the history state itself has an exponentially close fidelity with the ground state by Eq. (4.14), we have that for large enough N the guiding state itself has an inverse polynomially close fidelity with the ground state $|g\rangle$ (note that this is a different ground state than the one we had before). For the new pre-idled circuit

we can re-derive all results in our construction by replacing $\tilde{T}$ by $M = \tilde{T} + N$, so the fidelity can be lower bounded as

$$\begin{aligned}
|\langle u_{\text{YES}}^{\text{new}}|g\rangle|^2 &\geq 1 - \left(\sqrt{1 - |\langle u_{\text{YES}}^{\text{new}}|\left(|\eta(y^*)\rangle|0\rangle\right)|^2} + \sqrt{1 - |(\langle\eta(y^*)|\langle 0|)|g\rangle|^2}\right)^2 \\
&\geq 1 - \left(1 - \frac{N}{N + \tilde{T} + 1} + 2^{-\mathcal{O}\left(\sqrt{\tilde{T}+N}\right)}\right)^2 \\
&\geq 1 - \frac{1}{r(n)},
\end{aligned}$$

for any positive polynomial r, for some choice of $N = \text{poly}(\tilde{T})$.

Spectral gap. Taking the minimum over the YES- and NO-case, the spectral gap of the total Hamiltonian H can be lower bounded as

$$\gamma(H) = \Omega\left(\frac{1}{M^7}\right),$$

recalling that $M = \tilde{T} + N = \text{poly}(n)$.

Finally, we can renormalise H by a polynomial prefactor to meet the norm condition of Theorem 4.4.1, which preserves inverse polynomial scaling of the promise gap and spectral gap. □

Since $\mathsf{QGaLH}(k, a, b, \zeta)$ is in QCMA for any $k = \mathcal{O}(\log n)$, $b - a \geq 1/\text{poly}(n)$ and $\zeta \geq 1/\text{poly}(n)$, a direct corollary of the above theorem is the following: when one has access to a quantum computer (and in particular, quantum phase estimation), then from a complexity-theoretic perspective the ability to prepare any quantum state preparable by a polynomial-size 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 $1/\text{poly}(n)$.

It should be noted that our result does *not* imply that all Hamiltonians admitting efficiently quantumly-preparable guiding states also necessarily admit guiding states that are classically evaluatable. What it does show is that for any instance of the guidable local Hamiltonian problem, under the promise that there exists a guiding state efficiently preparable by a quantum circuit, there exists an (efficient) *mapping* to another instance of the guidable local Hamiltonian problem, under the promise that there exists a guiding state which is both classically evaluatable and quantumly preparable.

Whilst this reduction is efficient in the complexity-theoretic sense, it may not be useful in practice, as it is likely to eliminate all physical structure present in the original Hamiltonian. Hence, the main implication of our result is not that

such reductions are of practical merit, but rather that—again, from a complexity-theoretic point of view—the classical–quantum hybrid approach from Section 4.1, where the guiding state is selected using *classical* heuristics and energy estimation is performed *quantumly*, is at least as powerful as a fully quantum approach that uses quantum heuristics for state preparation instead.

4.5 Extension to excited states

In this section, we extend both constructions to apply to excited states. To this end, we first generalise both the guided and guidable local Hamiltonian problems to include a parameter c (not to be confused with the completeness parameter c, which we will no longer use), which labels the cth eigenstate.

4.5.1. Definition (Guided local Hamiltonian Low Energy, GLHLE(k, c, a, b, ζ)).

Input: A description of a collection of k-local Hermitian operators $\{H_i\}_{i\in[m]}$, which define an n-qubit k-local Hamiltonian $H = \sum_{i=1}^{m} H_i$ with $\|H\| \leq 1$; a description of a classically evaluatable state $|u\rangle \in \mathbb{C}^{2^n}$; two efficiently computable real numbers a, b such that $b > a$; an efficiently computable fidelity parameter $\zeta \in [0, 1]$; and a constant $c \in \mathbb{Z}_{\geq 0}$.

Promise: $\|\Pi_c |u\rangle\|^2 \geq \zeta$, where Π_c denotes the projection onto the eigenspace corresponding to the cth eigenvalue of H, with eigenvalues ordered in non-decreasing order. Moreover, either $\lambda_c(H) \leq a$ or $\lambda_c(H) \geq b$ holds.

Goal: Decide whether $\lambda_c(H) \leq a$ or $\lambda_c(H) \geq b$.

When $|u\rangle$ is also uniformly quantumly preparable (Definition 3.5.8), we write GLHLE$^*(k, c, a, b, \zeta)$.

The generalisations to the guidable and quantumly guidable settings, namely, the problems GaLHLE(k, c, a, b, ζ), GaLHLE$(k, c, a, b, \zeta)^*$, QGaLHLE(k, c, a, b, ζ), and QGaLHLE$^*(k, c, a, b, \zeta)$ follow analogously from their respective definitions.

Again, we show BQP- and QCMA-hardness for a wide range of parameter settings:

4.5.2. Proposition. *For any $0 \leq c \leq \mathrm{poly}(n)$ and any $\zeta \in (0, 1 - 1/\mathrm{poly}(n))$, there exist $a, b \in [-1, 1]$ with $b - a \geq 1/\mathrm{poly}(n)$ such that* GLHLE$(6, c, a, b, \delta)$ *(resp.* GaLHLE$(6, c, a, b, \delta)$*) is* BQP*-hard (resp.* QCMA*-hard), even under the following two promises:*

1. *The cth eigenvalue $\lambda_c(H)$ is separated from both $\lambda_{c-1}(H)$ and $\lambda_{c+1}(H)$ by a gap $\gamma \geq 1/\mathrm{poly}(n)$. We refer to such instances as γ-gapped* GLHLE$(6, c, a, b, \delta)$ *(resp.* GaLHLE$(6, c, a, b, \delta)$*).*

2. The guiding state is restricted to be a semi-classical subset state.

Proof:
We reduce directly from the BQP-hard Hamiltonian H defined in Eq. (4.8). Pick any $1 \leq c \leq \text{poly}(n)$ (since it holds trivially for $c = 0$). Let $|u\rangle$ be a semi-classical guiding state such that $|\langle u|\psi_0\rangle| \geq \zeta$. Define the following 6-local Hamiltonian $H^{(c)}$ on $n+1$ qubits:[10]

$$H^{(c)} = H^{(z)} \otimes |0\rangle\langle 0| + H^{(s)} \otimes |1\rangle\langle 1|, \tag{4.16}$$

where

$$H^{(z)} = \sum_{i=0}^{d} 2^i |1\rangle\langle 1|_i + \sum_{i=d+1}^{n} 2^{d+1} |1\rangle\langle 1|_i - \left(c - \frac{1}{2}\right)\mathbb{I},$$
$$H^{(s)} = \frac{1}{2}\frac{H + \mathbb{I}/4}{\|H\| + 1/4} - \frac{\mathbb{I}}{4},$$

and where $d = \lceil \log_2(c) \rceil$. The Hamiltonian $H^{(z)}$ has exactly c states with negative energy. Its smallest eigenvalue is $-c + \frac{1}{2}$, and its largest eigenvalue is

$$\sum_{i=0}^{d} 2^i + \sum_{i=d+1}^{n} 2^{d+1} - \left(c - \frac{1}{2}\right) = 2^{d+1} + 2^{d+1}(n-d) - \frac{1}{2} - c = \mathcal{O}(cn).$$

The spectrum of $H^{(z)}$ increases in steps of 1, and the smallest (resp. largest) positive (resp. negative) eigenvalues, i.e., that is closest to 0, is $\frac{1}{2}$ (resp. $-\frac{1}{2}$). Since $\text{eig}(H^{(s)}) \in [-1/4, 1/4]$, the spectrum of $H^{(s)}$ lies entirely within the interval $[-1/4, 1/4]$ and hence is positioned at the cth eigenvalue level within $H^{(c)}$ (i.e., just above the $c-1$ negative eigenvalues of $H^{(z)}$).

Therefore, given a guiding state $|u\rangle$ for H such that $|\langle u|\psi_0\rangle| \geq \zeta$, we construct a new guiding state

$$\left|u^{(c)}\right\rangle = |u\rangle \otimes |1\rangle,$$

which is also semi-classical and satisfies

$$\left|\left\langle u^{(c)}\middle|\psi_c^{(c)}\right\rangle\right| \geq \zeta,$$

where $|\psi_c^{(c)}\rangle$ denotes the cth eigenstate of $H^{(c)}$. Since this construction constitutes a polynomial-time reduction from an instance of $\text{GLH}(k, a, b, \zeta)$ to one of $\text{GLHLE}(k, c, a, b, \zeta)$ (with $1 \leq c \leq \text{poly}(n)$), we conclude that $\text{GLHLE}(6, c, a, b, \zeta)$ is BQP-hard for $k \geq 6$. Moreover, the eigenvalue gaps in $H^{(c)}$ satisfy

$$\lambda_c(H^{(c)}) - \lambda_{c-1}(H^{(c)}) = \frac{1}{4}, \qquad \lambda_{c+1}(H^{(c)}) - \lambda_c(H^{(c)}) = \gamma,$$

[10]Note that this gadget can be trivially adapted to show that estimating the n highest energy states is BQP-hard.

with γ determined as in the base construction. Since we have $\|H^{(c)}\| = \mathcal{O}(cn) = \text{poly}(n)$, normalisation preserves the inverse-polynomial spectral and promise gaps.

The same construction applies for the QCMA-case when using the Hamiltonian from Theorem 4.4.4, showing QCMA-hardness of GaLHLE$(6, c, a, b, \zeta)$ under the stated conditions. □

4.6 Reductions via approximate Hamiltonian simulators

In this section, we use the technique of *approximate Hamiltonian simulation* to improve the previously established hardness results with respect to the locality parameter, and to extend them to more physically motivated families of Hamiltonians.

4.6.1 Approximate Hamiltonian simulators

While in the QMA-hardness reduction it suffices to focus solely on the eigenvalues in the simulation, in the reductions for the guided and guidable local Hamiltonian problems it is also important to understand how the eigenvectors change under perturbative simulation. To this end, it is convenient to first introduce the notion of *approximate Hamiltonian simulation*.

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

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

2. $\|H'_{\leq\Delta} - \tilde{\mathcal{E}}(H)\| \leq \epsilon$, where $H'_{\leq\Delta} \coloneqq P_{\leq\Delta(H')}H'$.

Here, V is a local isometry that can be written as $V = \bigotimes_i V_i$ where each V_i is an isometry acting on at most 1 qubit, and P and Q are locally orthogonal projectors (i.e. for all i there exist orthogonal projectors P_i and Q_i acting on the same subsystem as V_i such that $P_iQ_i = 0$, $P_iP = P$ and $Q_iQ = Q$) 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 $\text{poly}(n, \eta^{-1}, \epsilon^{-1}, \Delta)$, and the description of H' can be computable in $\text{poly}(n)$ time given the description of H.

We approximately simulate the original Hamiltonian H within the low-energy subspace of H'. The corresponding encoding of a quantum state can be taken as

$$\mathcal{E}_{\text{state}}(\rho) = V(\rho \otimes \sigma)V^\dagger, \tag{4.17}$$

where σ is any state satisfying $P\sigma = \sigma$ (assuming $P \neq 0$). If ρ is an eigenvector of H with eigenvalue α and the spectrum is sufficiently gapped, then $\mathcal{E}_{\text{state}}(\rho)$ is approximately an eigenvector of H' with eigenvalue $\alpha' \in [\alpha - \epsilon, \alpha + \epsilon]$. Similarly, we define

$$\tilde{\mathcal{E}}_{\text{state}}(\rho) = \tilde{V}(\rho \otimes \sigma)\tilde{V}^\dagger.$$

In a slightly abuse of notation, when $\rho = |g\rangle\langle g|$ and $\sigma = |e\rangle\langle e|$ for some pure states $|e\rangle$ and $|g\rangle$, we simply write $\mathcal{E}_{\text{state}}(|g\rangle)$ and $\tilde{\mathcal{E}}_{\text{state}}(|g\rangle)$, as the encoding is guaranteed to transform $|g\rangle)$ to another pure state. From now on, we will always implicitly assume that σ is a pure state.

In [ZA21], it is shown that there exist families of Hamiltonians that can efficiently simulate any $\mathcal{O}(1)$-local Hamiltonian (on any hypergraph interaction structure, so not only 2D lattices). Such families are referred to as *strongly universal Hamiltonians*. Formally, the notions of strong and weak universality are defined as follows:

4.6.2. Definition (Strong and weak universality [ZA21]). A family of Hamiltonians $\{H_m\}$ is *weakly universal* if, given any $\Delta, \eta, \epsilon > 0$, any $\mathcal{O}(1)$-local n-qubit Hamiltonian can be (Δ, η, ϵ)-simulated by some H_m. Such a family is *strongly universal* if the simulation is always efficient: H_m is efficiently computable in $\mathcal{O}(\text{poly}(n))$ time, requires $n' = \mathcal{O}(\text{poly}(n, \eta^{-1}, \epsilon^{-1}, \Delta))$ qubits, and satisfies $\|H_m\| = O(\text{poly}(n, \eta^{-1}, \epsilon^{-1}, \Delta))$.

The following lemma shows that simulating low-energy states whose energies are well-separated from other states approximately preserves overlap with the (encoding of) the original state.

4.6.3. Lemma (Simulation of the gapped excited state). *Suppose the cth excited state $|g\rangle$ of a n-qubit Hamiltonian H with $\|H\| \leq \text{poly}(n)$ is separated from both the $(c-1)$th and $(c+1)$th eigenvalue (if they exist) by $\gamma > 0$. Let H' be an efficient (Δ, η, ϵ)-simulation of H such that $2\epsilon < \gamma$. Then H' has a non-degenerate cth excited state $|g'\rangle$, and*

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

Proof:
This is a slight modification of Lemma 2 from [BH17]. By assumption, the cth excited state $|g\rangle$ of H is non-degenerate and separated by a gap γ from adjacent eigenvalues. Since $\|H\| \leq \text{poly}(n)$ and the simulation is efficient, we can take Δ to be a sufficiently large enough polynomial so that the simulation H'

approximates the entire spectrum of H up to error ϵ in its low-energy subspace (by Definition 4.6.1). Hence, for $2\epsilon < \gamma$, the non-degeneracy at position c is preserved in H'. Let us view $\tilde{\mathcal{E}}(H)$ as the unperturbed Hamiltonian and define the perturbation $Q \coloneqq H' - \tilde{\mathcal{E}}(H)$, where $\tilde{\mathcal{E}}$ is the approximate encoding from Definition 4.6.1. Then $H' = \tilde{\mathcal{E}}(H) + Q$ is a perturbation of $\tilde{\mathcal{E}}(H)$ with small norm $\|Q\| \leq \epsilon$. First-order perturbation theory for eigenvectors implies that the cth excited state of H', denoted $|g'\rangle$, is close to the image $\tilde{\mathcal{E}}_{\text{state}}(|g\rangle)$ of the cth excited state of H under the approximate encoding:

$$\left\|\tilde{\mathcal{E}}_{\text{state}}(|g\rangle) - |g'\rangle\right\| \leq \mathcal{O}(\gamma^{-1}\epsilon).$$

Now we bound the difference between the encoded state $\mathcal{E}_{\text{state}}(|g\rangle)$ and its approximate version:

$$\left\|\mathcal{E}_{\text{state}}(|g\rangle) - \tilde{\mathcal{E}}_{\text{state}}(|g\rangle)\right\| = \left\|V(|g\rangle \otimes |e\rangle) - \tilde{V}(|g\rangle \otimes |e\rangle)\right\| \leq \eta.$$

Finally, applying the triangle inequality, we find

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

as claimed. □

Since we will always use efficient Hamiltonian simulators and the Hamiltonians H always have operator norm at most $\text{poly}(n)$ and $\gamma \geq 1/\text{poly}(n)$, Lemma 4.6.3 allows us to make $\left\|\mathcal{E}_{\text{state}}(|g\rangle) - |g'\rangle\right\|$ arbitrarily inverse polynomially small in n.

4.6.2 General strategy

Using approximate Hamiltonian simulators, our general strategy for extending the hardness results proceeds as follows:

1. Construct a BQP- (resp. QCMA-)hard problem instance of GLHLE(k, c, a, b, ζ) (resp. GaLHLE(k, c, a, b, ζ)), where the cth eigenstate satisfies:
   - It is γ-gapped from both the $(c-1)$th and $(c+1)$th eigenvalues (if they exists) for some $\gamma \geq 1/\text{poly}(n)$.
   - It has ζ-overlap with a classically evaluatable state.
2. Use strongly universal Hamiltonian simulators to extend the hardness results to families of Hamiltonians satisfying desirable physical or structural properties.

To make this strategy work, it is crucial that the strongly universal simulators preserve (approximate) classical evaluability. In this section, we will restrict our attention to the exact setting. We define this formally as follows:

4.6.4. Definition. We say that a state encoding $\mathcal{E}_{\text{state}}$ *preserves classical evaluability* if, for any classically evaluatable state $|u\rangle$, the state $|w\rangle = \mathcal{E}_{\text{state}}(|u\rangle)$ is also classically evaluatable.

The Hamiltonian simulators we consider will typically use one of the following three types of encodings:

1. **Mediator qubits:** In this encoding, some simple ancilla states are appended to the original state.
2. **Subspace encoding:** A local isometry is applied to the original state to embed it into a larger Hilbert space.
3. **local unitaries:** A local unitary of the form $U \otimes U \otimes \cdots \otimes U$, where each U acts on a single qubit, is applied to the original state.

Note that local unitaries are a special case of local isometries. Therefore, by Proposition 3.5.6, all of the above encodings preserve classical evaluability—provided that, in the case of mediator qubits, the appended ancilla states themselves are classically evaluatable.

We will examine each of these encodings in more detail in the next subsection. First, we formalise our general hardness extension strategy in the following theorem.

4.6.5. Theorem. *Let $\{H'_m\}$ be a family of strongly universal k'-local Hamiltonians whose state encoding $\mathcal{E}_{\text{state}}$ preserves classical evaluability. Then, for every polynomial $p(n)$, there exist $a', b' \in [-1, 1]$ with $b'-a' \geq 1/\text{poly}(n)$, $\gamma' \geq 1/\text{poly}(n)$ such that γ'-gapped* GLHLE(k', c, a', b', ζ') *(resp.* GaLHLE(k', c', a', b', ζ)*) is* BQP*-hard (resp.* QCMA*-hard) with $\zeta' \geq 1 - 1/p(n)$ and the restriction that the Hamiltonians come from the family $\{H'_m\}$.*

Proof:
By Theorem 4.5.2, for any $0 \leq c \leq \text{poly}(n)$ and any polynomially-bounded $q(n)$, there exists $a, b \in [-1, 1]$ with $b-a \geq 1/\text{poly}(n)$, $\gamma \geq 1/\text{poly}(n)$ and $\zeta \in [0, 1]$ with $\zeta \geq 1-1/q(n)$, such that γ-gapped GLHLE(k, c, a, b, ζ) (resp. GaLHLE(k, c, a, b, ζ)) is BQP-hard (resp. QCMA-hard) for some family of 2-local Hamiltonians $\{H_n\}$ and a classically evaluatable state $|u\rangle$. Since $\{H'_m\}$ is a family of strongly universal simulators, we may choose any $\epsilon \geq 1/\text{poly}(n)$, $\eta \geq 1/\text{poly}(n)$, and $\Delta \leq \text{poly}(n)$, such that there exists a H'_m which is an efficient (Δ, η, ϵ)-simulation of any H_n. For simplicity, we omit the exact polynomials and instead refer to the constraints ϵ, η, Δ and $q(n)$ must satisfy. Let the new guiding state be $|w\rangle = \mathcal{E}_{\text{state}}(|u\rangle)$.

First, we require ϵ to satisfy $\epsilon < \gamma/2$ and $\epsilon < (b-a)/2$. Such a choice is possible because both γ and $b-a$ are inverse-polynomial. Given this, define

$$b' := b - \epsilon, \quad a' := a + \epsilon, \quad \text{and} \quad \Delta := 2\|H_n\|.$$

These choices ensure that non-degeneracy of the cth eigenstate is preserved and if $\lambda_c(H_n) \leq a$ then $\lambda_c(H'_m) \leq a'$, and similarly if $\lambda_c(H_n) \geq b$ then $\lambda_c(H'_m) \geq b'$. Next, we apply Lemma 4.6.3 to control the deviation in the eigenstate. We have that $|\langle u|g\rangle|^2 \geq \zeta \geq 1 - 1/q(n)$. Since fidelity is preserved under $\mathcal{E}_{\text{state}}$, we have $|\mathcal{E}_{\text{state}}(\langle u|)\mathcal{E}_{\text{state}}(|g\rangle)|^2 = |\langle w|\,\mathcal{E}_{\text{state}}(|g\rangle)|^2 \geq \zeta$. Since we have that the encoding is efficient and Lemma 4.6.3 gives a bound of $\eta + \mathcal{O}(\gamma^{-1}\epsilon)$, we can pick ϵ and η to be sufficiently small inverse polynomials such that $\|\mathcal{E}_{\text{state}}(|g\rangle) - |g'\rangle\|$ becomes arbitrarily polynomially small in n, where $|g'\rangle$ is the cth excited state of H'_m. Since we only care about overlap with some state in the ground space (so global phases are irrelevant), for any polynomial p, there exist choices of $\epsilon, \eta \in 1/\text{poly}(n), q \in \text{poly}(n)$ such that:

$$|\langle w|g'\rangle|^2 \geq 1 - \frac{1}{p(n)} =: \zeta'.$$

Because the simulation is efficient, both the number of qubits of H'_m and its operator norm are bounded by $\text{poly}(n)$, so we can meet the normalization condition and ensure that our new $a', b' \in [-1, 1]$ by adding a small identity term and by dividing by a trivial upper bound on the operator norm of H'_m (e.g., the sum of all the operator norms of the local terms). This also ensures that $\gamma' \geq 1/\text{poly}(n)$. By assumption $\mathcal{E}_{\text{state}}$ preserves classical evaluability, so the state $|w\rangle$ is classically evaluatable.

The same argument applies in the QCMA setting, starting from a QCMA-hard instance of GaLHLE(k, c, a, b, ζ), which completes the proof. □

In fact, all reductions we consider map a semi-classical encoded state to another semi-classical encoded state, so we directly have that all results hold with respect to defining the problems in terms of samplable, classically evaluatable, or quantumly-preparable states. Nevertheless, in extending these results to other Hamiltonians, the above strategy might prove useful.

4.6.3 Classical evaluability-preserving Hamiltonian simulators

We first introduce some families of Hamiltonians to be used in the reductions via approximate Hamiltonian simulators. Given a set of (at most) two-body interactions $\mathcal{S} = \{h_\alpha\}$, $\mathcal{S}$-Hamiltonian refers to the family of Hamiltonians that can be written in the form

$$H = \sum_{\langle i,j\rangle \in E} J_{i,j} h^{(i,j)}_{\alpha_{i,j}}, \tag{4.18}$$

where $J_{i,j} \in \mathbb{R}$, $h^{(i,j)}_{\alpha_{i,j}}$ is a two-local interaction chosen from $\mathcal{S}$ and E is the set of edges that represents the connectivity of interaction [CM16]. If the connectivity

of the two-body interactions is restricted to a 2D square lattice, we call such a family $\mathcal{S}$-Hamiltonian on a 2D square lattice. We also introduce the notion of 2SLD and non-2SLD:

4.6.6. Definition (2SLD interaction [CM16]). Suppose $\mathcal{S}$ is a set of two-qubit interactions. We say that $\mathcal{S}$ is 2SLD if there exists a single-qubit unitary $U \in \mathbb{SU}(2)$ such that for all $h_i \in \mathcal{S}$,

$$(U \otimes U)h_i(U^\dagger \otimes U^\dagger) = \alpha_i Z \otimes Z + A_i \otimes \mathbb{I} + \mathbb{I} \otimes B_i,$$

where $\alpha_i \in \mathbb{R}$ and A_i, B_i are single-qubit Hermitian operators.

A set $\mathcal{S}$ is non-2SLD if it is not 2SLD. In particular, such non-2SLD $\mathcal{S}$ includes the following physically motivated[11] Hamiltonians:

- $\{Z, X, ZZ, XX\}$ ($ZZXX$ interaction [BL08])
- $\{Z, X, ZX, XZ\}$ (ZX interaction [BL08])
- $\{XX + YY\}$ (general XY interaction)
- $\{XX + YY + ZZ\}$ (general Heisenberg interaction).

If there is only a single type of interaction (like $\mathcal{S} = \{XX + YY + ZZ\}$), the Hamiltonian is called *semi-translationally-invariant* (interaction strength can differ in each term.).

Restriction on the sign of the interaction. We also introduce an even more restricted class of $\mathcal{S}$-Hamiltonian in which all the signs of the coefficients are promised to be non-negative (i.e., all of $J_{i,j}$ in Eq. (4.18) must satisfy $J_{i,j} \geq 0$). We call such a family of Hamiltonians an $\mathcal{S}^+$-Hamiltonian following [PM17]. In [PM17], the following results are shown:

- $\{\alpha XX + \beta YY + \gamma ZZ\}^+$-Hamiltonian are QMA-complete if $\alpha + \beta > 0$, $\alpha + \gamma > 0$ and $\beta + \gamma > 0$ hold.
- $\{\alpha XX + \beta YY + \gamma ZZ\}^+$-Hamiltonian is QMA-complete when the interactions are restricted to the edges of a 2D triangular lattice, provided that $\alpha XX + \beta YY + \gamma ZZ$ is not proportional to $XX + YY + ZZ$ and that $\alpha + \beta > 0$, $\alpha + \gamma > 0$, and $\beta + \gamma > 0$.

[11] For clarity, in [CM16] and here, all hardness results require *non-uniform* weights on constraints. It is an open question whether one can obtain (say) QMA-hardness results with uniform (i.e., unit weight) constraints for such models. This remains an interesting open question, as many-body physicists typically utilise unit weights to model physical systems.

The first type of $\mathcal{S}^+$-Hamiltonian includes the antiferromagnetic Heisenberg model ($\{XX+YY+ZZ\}^+$-Hamiltonian) and the antiferromagnetic XY model ($\{XX+YY\}^+$-Hamiltonian) as important special cases. The antiferromagnetic XY model (unlike the antiferromagnetic Heisenberg model) remains QMA-complete if its geometric interaction is restricted to a 2D triangular lattice, as it is included in the second type of $\mathcal{S}^+$-Hamiltonian above.

We now sketch the construction of the strong Hamiltonian simulation introduced in [ZA21]. The simulation consists of two main components. First, they construct a spatially sparse 5-local Hamiltonian using a quantum phase estimation (QPE) circuit and a modified version thereof, following the approach of [OT08]. This procedure can be interpreted as a "Hamiltonian-to-circuit" transformation, followed by a return to a Hamiltonian via a circuit-to-Hamiltonian construction. In the second step, they perturbatively simulate the resulting spatially sparse Hamiltonian using established techniques from [OT08, CMP18, PM17].

Let us take a closer look at the individual steps of their construction.

(1) Arbitrary k-local Hamiltonian $\rightarrow$ Spatially sparse 5-local Hamiltonian ([ZA21]). Let H be a target $\mathcal{O}(1)$-local Hamiltonian. We write H in its eigendecomposition as $H = \sum_i E_i \left|\psi_i\right\rangle\!\left\langle\psi_i\right|$, where $\{E_i\}$ and $\{|\psi_i\rangle\}$ are the eigenvalues and eigenvectors, respectively. In [ZA21], the authors construct a spatially sparse quantum circuit $U_{\mathrm{PE}}^{\mathrm{sparse}}$ that approximately estimates the energies of H, i.e.,

$$U_{\mathrm{PE}}^{\mathrm{sparse}} \sum_i c_i \left|\psi_i\right\rangle \left|0^m\right\rangle \approx \sum_i c_i \left|\psi_i\right\rangle \left|\tilde{E}_i\right\rangle \left|\mathrm{other}\right\rangle ,$$

where $\{c_i\}$ are arbitrary coefficients, and $\{|\tilde{E}_i\rangle\}$ are approximations of the energy eigenvalues $\{E_i\}$. The circuit $U_{\mathrm{PE}}^{\mathrm{sparse}}$ is constructed from a 1D nearest-neighbour circuit $U_{\mathrm{NN}}^{\mathrm{sparse}}$, which is then converted into a spatially sparse circuit using ancilla qubits and swap gates. To prepare for a circuit-to-Hamiltonian construction, they define the composed circuit

$$U = (\mathrm{Idling})(U_{\mathrm{PE}}^{\mathrm{sparse}})^\dagger(\mathrm{Idling})U_{\mathrm{PE}}^{\mathrm{sparse}}.$$

Applying the circuit-to-Hamiltonian mapping to this U yields a spatially sparse 5-local Hamiltonian H_{circuit}. First-order perturbation theory is then used to show that H_{circuit} simulates H in its low-energy subspace. The effective encoding from the low-energy subspace of H_{circuit} to that of H can be approximated by the map $H \rightarrow H \otimes |\alpha\rangle\!\langle\alpha|$, where $|\alpha\rangle$ is a subset state supported on a poly(n)-sized subset S'. This state arises from the history state structure of the idling phase following uncomputation; see Proposition 2 in [ZA21] for details. The corresponding encoding of states takes the form:

$$|u\rangle \rightarrow |u\rangle \otimes |\alpha\rangle .$$

If $|u\rangle$ is a semi-classical subset state, then the encoded state is also semi-classical. Moreover, if $|u\rangle$ is classically evaluatable, then classical evaluability is preserved under this encoding by Proposition 3.5.6.

(2) Spatially sparse 5-local Hamiltonian → Spatially sparse 10-local real Hamiltonian (Lemma 22 of [CMP18]). In this simulation step, the state is encoded by appending a polynomial number of $|+_y\rangle$ states, where $|+_y\rangle$ denotes the $+1$ eigenvector of the Pauli Y matrix:

$$|u\rangle \to |u\rangle \otimes |+_y\rangle \otimes \cdots \otimes |+_y\rangle . \tag{4.19}$$

This encoding does not map a semi-classical subset state to another semi-classical subset state, but instead maps it to a semi-classical *encoded* state, because of the following: Let V_y be a unitary such that $|+_y\rangle = V_y |0\rangle$, and suppose $|u\rangle = \frac{1}{\sqrt{|S|}} \sum_{x\in S} |x\rangle$ is a semi-classical subset state. Then, the right-hand side of Eq. (4.19) can be rewritten as

$$|u\rangle \otimes |+_y\rangle \otimes \cdots \otimes |+_y\rangle = \frac{1}{\sqrt{|S|}} \sum_{x\in S\times\{0\cdots 0\}} (\mathbb{I}\otimes\cdots\otimes\mathbb{I}\otimes V_y\otimes\cdots\otimes V_y)\,|x\rangle .$$

This is a semi-classical encoded state with subset $S\times\{0\cdots 0\}$ and a local isometry given by $\mathbb{I}\otimes\cdots\otimes\mathbb{I}\otimes V_y\otimes\cdots\otimes V_y$ (note that in this case, the isometry is a local unitary). Once again, if $|u\rangle$ is classically evaluatable, classical evaluability is preserved under this encoding by Proposition 3.5.6.

(3) Spatially sparse 10-local real Hamiltonian → Spatially sparse 2-local Pauli interactions with no Y-terms ([OT08, CM16]). This step is accomplished by first simulating the 10-local real Hamiltonian with an 11-local Hamiltonian whose Pauli decomposition contains no Pauli Y terms [CMP18, Lemma 40]. In the associated encoding, $|1\rangle$ states are appended for the polynomially many mediator qubits introduced in the simulation. Next, subdivision gadgets and 3-to-2 gadgets [OT08] are applied to reduce the locality. This simulation introduces additional mediator qubits, for which the encoding of states simply appends $|0\rangle$ states. The resulting Hamiltonian takes the form

$$\sum_{i<j} \alpha_{ij} A_{ij} + \sum_k (\beta_k X_k + \gamma_k Z_k) ,$$

where each A_{ij} is one of the interactions $\{X_iX_j, X_iZ_j, Z_iX_j, Z_iZ_j\}$. Since only computational basis states are added in the encoding, classical evaluability is trivially preserved.

(4) Subspace encoding for spatially sparse $\mathcal{S}_0 = \{XX + YY + ZZ\}$ or $\{XX+YY\}$ Hamiltonian (Theorem 42 of [CMP18]). We have already obtained 2-local Hamiltonian in the form $\sum_{i<j} \alpha_{ij} A_{ij} + \sum_k (\beta_k X_k + \gamma_k Z_k)$. Then we show how to simulate this Hamiltonian with arbitrary non-2SLD $\mathcal{S}$-Hamiltonians. We first consider $\mathcal{S}_0$-Hamiltonians, where $\mathcal{S}_0 = \{XX + YY + ZZ\}$ or $\mathcal{S}_0 = \{XX + YY\}$. In this simulation, we use subspace encoding in which the *logical qubit* of the original Hamiltonian is encoded into four *physical qubits*. Consider the simulation by Heisenberg interaction $\{XX + YY + ZZ\}$ for example. Each logical qubit is encoded into a 4-qubit state by an isometry that is defined as

$$V|0\rangle = |0_L\rangle = \left|\Psi^-\right\rangle_{13} \left|\Psi^-\right\rangle_{24} \tag{4.20}$$

$$V|1\rangle = |1_L\rangle = \frac{2}{\sqrt{3}} \left|\Psi^-\right\rangle_{12} \left|\Psi^-\right\rangle_{34} - \frac{1}{\sqrt{3}} \left|\Psi^-\right\rangle_{13} \left|\Psi^-\right\rangle_{24}, \tag{4.21}$$

where $|\Psi^-\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$. For details, see [CMP18, Theorem 42]. The encoding of states for $\{XX + YY\}$ interaction is the same. A semi-classical encoded state is clearly mapped to a semi-classical encoded state by applying a local isometry of the corresponding subspace encoding. By Proposition 3.5.6, applying local isometries preserves classical evaluability.

(5) Spatially sparse $\mathcal{S}_0$-Hamiltonian $\to$ $\mathcal{S}_0$-Hamiltonians on a 2D square lattice (Lemma 47 of [CMP18]). This simulation is achieved using three types of perturbative gadgets: subdivision, fork, and crossing gadgets. Each application of a gadget introduces a mediator qubit. It suffices to apply $\mathcal{O}(1)$ rounds of these gadgets in parallel to simulate a spatially sparse $\mathcal{S}_0$-Hamiltonian with a $\mathcal{S}_0$-Hamiltonian defined on a 2D square lattice. This number of rounds ensures that interaction strengths remain polynomially bounded, rather than growing exponentially (in contrast, simulating general interaction graphs may require $\mathcal{O}(\log n)$ rounds of perturbative gadgets). Since the encodings in this step only append single-qubit computational basis states, which are trivially classically evaluatable, classical evaluability is again preserved.

(6) $\mathcal{S}_0$-Hamiltonian on a 2D square lattice $\to$ Arbitrary non-SLD $\mathcal{S}$-Hamiltonian on a 2D square lattice (Theorem 43 of [CMP18]). Finally, this simulation is performed using variants of mediator qubit gadgets, subspace encoding gadgets, and local unitaries, all of which preserve classical evaluability.[12]

Combining all of the above steps with Theorem 4.6.5, we arrive at the following corollary:

4.6.7. Corollary (Hardness of physically motivated Hamiltonians). *For any integer $0 \leq c \leq \mathrm{poly}(n)$, and any $\zeta \in (0, 1 - 1/\mathrm{poly}(n))$, there exist $a, b \in [-1, 1]$*

[12] Applying local unitaries means simulating H by $U^{\otimes n} H (U^\dagger)^{\otimes n}$, where U acts on a single qubit. The corresponding state encoding is $\mathcal{E}_{\mathrm{state}}(|\psi\rangle) = U^{\otimes n} |\psi\rangle$.

with $b-a \geq 1/\mathrm{poly}(n)$ *such that* $\mathrm{GLHLE}(k,c,a,b,\zeta)$ *(resp.* $\mathrm{GaLHLE}(k,c,a,b,\zeta)$*) is* BQP*-hard (resp.* QCMA*-hard) for the following classes of Hamiltonians:*

- *Non-2SLD* $\mathcal{S}$*-Hamiltonians on a 2D square lattice;*
- $\{XX+YY+ZZ\}^+$*-Hamiltonians;*
- $\{XX+YY\}^+$*-Hamiltonians on a 2D triangular lattice.*

It can also be verified that indeed each simulation step (1)-(6) above maps semi-classically encoded states to other semi-classically encoded states, so Corollary 4.6.7 also holds when the problem is formulated in terms of samplable, classically evaluatable, or quantumly preparable states (in fact, all three simultaneously). This directly implies containment (and hence completeness) for both $\mathrm{GLHLE}^*(k,c,a,b,\zeta)$ and $\mathrm{GaLHLE}^*(k,c,a,b,\zeta)$ under the same setup as in Corollary 4.6.7.

Chapter 5

Finding quantum partial assignments by search-to-decision

5.1 Introduction

Decision (or promise) problems are arguably the central objects of study in computational complexity theory. While resolving a decision problem provides information about the *existence* of a solution, it does not provide the solution itself. Fortunately, *search problems*, where the task is to output an actual solution, are often reducible to their related decision problems. In this context, one generally considers *Turing reductions*: here, one has access to an oracle capable of solving a class of decision problems, which is then used as a subroutine to solve the desired search problem.

As an example, consider a formula ϕ corresponding to a Boolean satisfiability (SAT) problem on n bits, and assume that we have access to an NP oracle capable of solving any problem in NP (or equivalently, a single NP-complete problem). Under the assumption that ϕ is satisfiable, one can find a solution x^* such that $\phi(x^*) = 1$ in the following way: one queries the NP oracle adaptively to ask whether ϕ is satisfiable under the extra constraint that a certain subset of variables takes on specific values, i.e., under a fixed partial assignment. Every query to the oracle yields one bit of information about some x^*, and thus, after n queries, the algorithm has found a solution.[1] This strategy generally works for any problem in NP and can also be used to calculate the optimal value of an optimisation problem up to exponential accuracy using binary search [Kre88].

In [INN+22], Irani, Natarajan, Nirkhe, Rao, and Yuen studied whether a similar result holds in a quantum setting, where the goal is to output a *quantum state* as a QMA witness, as opposed to a classical string. To extend the SAT example to the quantum case, one can consider its quantum generalisation in terms of the local Hamiltonian problem (see Chapter 3). Recall that in this problem

[1] It can return any satisfying assignment if the solution is not unique.

the input is a Hermitian operator H on n qubits that can be efficiently written down as a sum of local terms, each acting non-trivially on only a subset of the qubits, and two parameters a and b with $b > a$. The task is to decide whether the ground state energy (its smallest eigenvalue) is $\leq a$ or $\geq b$. For sufficiently small $b - a \geq 1/\text{poly}(n)$, the local Hamiltonian problem is QMA-complete [KSV02]. The question is now whether a quantum algorithm with access to a QMA oracle can prepare the ground state (that is, the eigenstate corresponding to the smallest eigenvalue) of H as a quantum state. Or in other words: does the ability to estimate the ground state energy also provide the ability to find the ground state itself?

As we recall from Chapter 1, [INN+22] pointed out that it seems difficult to adapt the above strategy for NP to QMA because of the following two issues:

(i) the description size complexity of a quantum state on n qubits is generally exponential in n;

(ii) there does not appear to be a natural way of conditioning a quantum state on a partial assignment.

It turns out that with a PP-oracle, one can avoid this partial assignment strategy and generate QMA witnesses by making only a single quantum query [INN+22]. However, [INN+22] shows that relative to a quantum oracle, QMA fails to have search-to-decision reductions, contrasting with some related classes where the witnesses are *classical*. For instance, NP, MA, and QCMA all have such search-to-decision reductions relative to all oracles [INN+22].

Going back to the local Hamiltonian problem, we observe that the full quantum state contains more information than strictly needed—since the Hamiltonian is local, it suffices to have sufficiently good approximations of all k-local *density matrices* of a low-energy state to compute the energy, provided we know that the density matrices are approximately *consistent* with some global state. Constant-locality density matrices do not suffer from point (i) above, as there are only a polynomial number of them and each has a polynomially-sized description (for inverse polynomial accuracy). However, from Chapter 3 we know that it is again QMA-complete to check if all density matrices are consistent with a global quantum state [Liu06, BG22]. Nevertheless, this should not necessarily pose a problem when one has access to a QMA oracle. Hence, perhaps if we relax the requirement of what it means to find the ground state—namely, to recovering its local description in terms of reduced density matrices—could something be possible with a QMA oracle?

5.1.1 Results of this chapter

In this chapter, we will show that with access to a QMA oracle, a quantum analogue of the adaptive partial assignment strategy is possible for density matrices

of low-energy states which can be ensured to be approximately consistent. This demonstrates that point (ii) has a natural quantum manifestation for the class QMA when density matrices of low-energy states of local Hamiltonians are concerned. In particular, we will prove the following results:

- For any k, q constant, we have that for any k-local Hamiltonian H there exists a polynomial-time classical algorithm that makes queries to a QMA oracle and outputs a set of q-local density matrices that are at least arbitrarily (inverse-polynomially) close in trace distance to the density matrices of a state with energy arbitrarily (inverse-polynomially) close to the ground state energy.

- Moreover, we show that there exists an *approximately witness-preserving reduction* from any problem in QMA to the above task, which means that the above result can be lifted to show that approximations of the density matrices of a near-optimal witness for *any* problem in QMA can be found.

We give the algorithm for the first point in Section 5.2, and the approximately witness-preserving reduction will be given in Section 5.3, and conclude with some open problems in Section 5.4

5.1.2 Related work

In [Amb14], Ambainis initiated the study of $\mathsf{P}^{\mathsf{QMA}[\log]}$, where he showed that the problem APX-SIM—which formalizes the problem of computing expectation values of local observables on the ground state—is complete for this class. This work was extended by Gharibian and Yirka [GY19], who gave a similar $\mathsf{P}^{\mathsf{QMA}[\log]}$-completeness result for estimating two-point correlation functions, as well as fixing a bug in the hardness proof of Ambainis' original work. In addition, Gharibian and Yirka showed that $\mathsf{P}^{\mathsf{QMA}[\log]} \subseteq \mathsf{PP}$. In [GPY20], these types of ground state observable problems were studied for Hamiltonians under more physically motivated constraints.

Next to the work mentioned in the introduction by [INN+22], Gharibian and Kamminga study oracle search-to-decision reductions for *classical* problems using *quantum* algorithms in [GK24]. Specifically, they examine this in the context of problems in NP where a quantum algorithm has access to an NP oracle. They show that $\mathsf{FNP} \subseteq \mathsf{FBQP}^{\mathsf{NP}[\log]}$, meaning that any witness to an NP-relation can be found using a quantum algorithm that makes $\mathcal{O}(\log n)$ NP queries.

As pointed out by Sevag Gharibian (private communication), a result similar to our Theorem 5.2.7 can be derived as a corollary of the proof that consistency is QMA-hard under Turing reductions, as given by Liu in [Liu06]. Liu's proof relies on techniques from convex optimisation while treating consistency as a black-box constraint, and also identifies the density matrices corresponding to a low-energy state of a Hamiltonian. We argue that our construction is simpler and more directly aligned with the idea of adaptively constructing partial assignments.

5.2 Finding low-energy marginals of local Hamiltonians

In this section, we present a simple algorithm that finds low-energy marginals of local Hamiltonians by making queries to a QMA oracle. We introduce a new QMA-complete promise problem, which we call the *low-energy density matrix verification* (LEDMV) problem, which will be called by the QMA oracle. This problem can be viewed as a combination of the local Hamiltonian problem and the Consistency of Local Density Matrices (CLDM) problem from Chapter 3. The key idea is that solving LEDMV provides simultaneous information on whether the input density matrices are approximately consistent with a global state *and* whether this global state has low energy. Combined with a brute-force search algorithm that searches for a specific density matrix on a constant-size subset of qubits, we can give a partial assignment strategy similar to what one would use for NP.

5.2.1 Low-energy Density Matrix Verification

Let us begin by giving a formal definition of the low-energy density matrix verification problem:

5.2.1. Definition (Low-energy density matrix verification, LEDMV$(k, q, \delta, \alpha, \beta)$).

Input: A classical description of a collection of k-local Hermitian operators $\{H_i\}_{i\in[m]}$, with $0 \preceq H_i \preceq 1$ for all $i \in [m]$, which define an n-qubit k-local Hamiltonian $H = \sum_{i=1}^{m} H_i$, along with efficiently computable numbers $a, \delta, \alpha, \beta \geq 0$ with $\beta - \alpha > 0$, and a classical description of a collection of q-local density matrices $D = \{\rho_j\}_{j\in[l]}$, where $l \leq \text{poly}(n)$, and each ρ_j acts on a subset $C_j \subseteq [n]$ with $|C_j| \leq q$. Let $\overline{C}_j = [n] \setminus C_j$ denote the complementary set.

Promise: One of the following two cases holds:

(i) There exists an n-qubit state ξ with $\text{tr}[H\xi] \leq a$ such that for all $j \in [l]$ we have $\left\|\text{tr}_{\overline{C}_j}[\xi] - \rho_j\right\|_1 \leq \alpha$;

(ii) For all n-qubit states ξ with $\text{tr}[H\xi] \leq a + \delta$ we have that there exists a $j \in [l]$ such that $\left\|\text{tr}_{\overline{C}_j}[\xi] - \rho_j\right\|_1 \geq \beta$.

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

We have that LEDMV$(k, q, \delta, \alpha, \beta)$ is trivially QMA-hard for $k, q \in \mathcal{O}(1)$, $\delta \geq 0$, and $\beta - \alpha = 1/\text{poly}(n)$, because one can choose the Hamiltonian to be the

identity operator $\mathbb{I}$ and set $a = m$, which turns it into a reformulation of the QMA-hard CLDM problem as defined in Definition 3.3.1. To demonstrate containment, we will show that LEDMV is in QMA for a wide range of parameters. The QMA protocol is given in Protocol 5.2.1.

Protocol 5.2.1: QMA protocol for LEDMV.

Input: Classical descriptions of $\{H_i\}$ and $D = \{\rho_j\}$, and efficiently computable numbers a, δ, α, β.

Set: $\gamma \coloneqq \min\{\frac{\delta}{m}, \beta - \alpha\}$, $r \coloneqq \max\{k, q\}$, $I \coloneqq \binom{[n]}{r}$.

Protocol:

1. The prover sends a classical description of the set $\Sigma \coloneqq \{\sigma_{i_1,\ldots,i_r}\}_{(i_1,\ldots,i_r)\in I}$ and a quantum proof $|\psi\rangle$.
2. Let $\{C_i^H\}$ be the set of indices of qubits that terms H_i acts on. The verifier performs the following four checks, and accepts if and only if all of them accept:
   - **Check 1:** it checks if all $\sigma_{i_1,\ldots,i_r}$ are valid density matrices.
   - **Check 2:** it checks if $\sum_{i\in[m]} \max \operatorname{tr}\left[H_i \operatorname{tr}_{\overline{C}_i^H}[\sigma_{i_1,\ldots,i_r}]\right] \leq a$, where the maximization is over all $\sigma_{i_1,\ldots,i_r}$ supported on the qubits with indices from C_i^H.
   - **Check 3:** it checks if $\max \left\|\rho_j - \operatorname{tr}_{\overline{C}_j}[\sigma_{i_1,\ldots,i_r}]\right\|_1 \leq \alpha$ for all $j \in [l]$, where the maximization is over all $\rho_{i_1,\ldots,i_r}$ supported on the qubits with indices from C_j.
   - **Check 4:** it uses the quantum proof $|\psi\rangle$ to verify CLDM(q, γ) with input Σ using Protocol 3.3.1.

Let us first clarify some of the notation and ideas behind Protocol 5.2.1. The input consists of a local Hamiltonian $H = \sum_{i\in[m]} H_i$ and a collection of local density matrices $D = \{\rho_j\}_{j\in[l]}$, both acting on an n-qubit Hilbert space. There are two notions of locality at play: the first is the locality of the Hamiltonian, where each term H_i acts non-trivially on at most k qubits; the second concerns the density matrices in D, where each ρ_j is defined on a subset of qubits $C_j \subseteq [n]$ of size at most q.

To track the qubits each Hamiltonian term acts on, we define the set $\{C_i^H\}_{i\in[m]}$,

where C_i^H contains the indices of the qubits on which H_i acts non-trivially, which is analogous to the role of C_j for each ρ_j. Let $r \coloneqq \max\{k, q\}$. We then consider all r-local reduced density matrices $\sigma_{i_1,\ldots,i_r}$ of an n-qubit state ξ, with indices drawn from the set $I \coloneqq \binom{[n]}{r}$. The Σ containing all such marginals is now by design *overcomplete*: this way, the set contains enough information to simultaneously estimate the energy of H and compare the individual trace distances with density matrices from D.

However, due to overlapping supports, the same subset of qubits may appear in multiple marginals $\sigma_{i_1,\ldots,i_r}$, which can lead to inconsistencies if these marginals are not even mutually consistent on their shared support. To address this, in Checks 2 (resp. Check 3) we simply compute the local expectation value (resp. trace distance) for all marginals $\sigma_{i_1,\ldots,i_r}$ whose set of supporting qubits contain C_i^H (resp. C_j), and keep only the maximum value to be used to check whether Check 2 (resp. Check 3) passes. This ensures soundness even when the prover submits an inconsistent collection of marginals, while being redundant in the honest case.

Finally, it is important to note that Check 4 is not intended to verify the consistency of the input set D itself. Rather, its purpose is to test whether the collection of marginals submitted by the prover is approximately consistent. Once this condition is met, the combination of *both* Checks 3 and 4 passing certifies that the marginals from the input set D are also approximately consistent with each other.

5.2.2. Lemma. *We have that* LEDMV$(k, q, \delta, \alpha, \beta)$ *is in* QMA *for* $k = \mathcal{O}(1)$, $q = \mathcal{O}(1)$, $\delta = \Omega(1/\mathrm{poly}(n))$ *and* $\beta - \alpha = \Omega(1/\mathrm{poly}(n))$.

Proof:
We prove the correctness of Protocol 5.2.1. First, observe that the protocol runs in polynomial time: the maximisation in Check 2 involves at most $\binom{n-k}{r-k}$ options for each term H_i, since each k-subset appears in that many r-sized subsets. This is polynomial in n for constant $r = \max\{k, q\}$. A similar argument applies to Check 3. All other checks are clearly polynomial-time for our choice of parameters.

Completeness. Suppose the YES-instance holds, and let ξ be the corresponding n-qubit state. The prover submits the full set of r-local reduced density matrices

$$\Sigma = \left\{\sigma_{i_1,\ldots,i_r} \mid \sigma_{i_1,\ldots,i_r} = \mathrm{tr}_{[n]\setminus\{i_1,\ldots,i_r\}}[\xi],\ (i_1,\ldots,i_r) \in I\right\}.$$

(*Note:* The reader might rightly point out that such an exact description may not admit a polynomial-size representation. However, an exponentially precise approximation can always be provided, which suffices for our purposes, since the overall acceptance and rejection probabilities will only be affected by an inverse-exponential additive error—preserving the inverse-polynomial completeness–soundness gap.)
Let us now verify that indeed all checks are passed with high probability:

- Check 1 succeeds because each $\sigma_{i_1,\dots,i_r}$ is a valid density matrix.
- Check 2 and Check 3 succeed because, by the promise, ξ yields energy $\leq a$ and each ρ_j is α-close to its marginal from ξ, and the trace distance cannot increase under partial trace.
- Check 4 accepts with high probability, since the prover gives consistent marginals from ξ, and by Theorem 3.3.5, the CLDM-protocol accepts with high probability.

Soundness. We will use proof by contradiction. Suppose that we are in a NO-instance and Checks 1–3 have all succeeded (if this is not the case, we are already done). In particular, this means that we have concluded:

$$\sum_{i\in[m]} \max \operatorname{tr}\left[H_i \operatorname{tr}_{\overline{C}_i^H}[\sigma_{i_1,\dots,i_r}]\right] \leq a, \quad \text{and} \quad \forall j \in [l] : \max \left\|\rho_j - \operatorname{tr}_{\overline{C}_j}[\sigma_{i_1,\dots,i_r}]\right\|_1 \leq \alpha,$$

where the maximisation is performed over the density matrices $\sigma_{i_1,\dots,i_r}$ with the relevant support, see Protocol 5.2.1. Now suppose that Check 4 accepts with probability $> 1/3$. Then there exists a state ξ' such that, for all $(i_1,\dots,i_r) \in I$,

$$\left\|\sigma_{i_1,\dots,i_r} - \operatorname{tr}_{[n]\setminus\{i_1,\dots,i_r\}}[\xi']\right\|_1 < \gamma.$$

However, this implies that, for our choice of γ,

$$\begin{aligned}
\operatorname{tr}[H\xi'] &= \sum_{i\in[m]} \operatorname{tr}\left[H_i \operatorname{tr}_{\overline{C}_i^H}[\xi']\right] \\
&= \sum_{i\in[m]} \max \operatorname{tr}\left[H_i \left(\operatorname{tr}_{\overline{C}_i^H}[\xi'] - \operatorname{tr}_{\overline{C}_i^H}[\sigma_{i_1,\dots,i_r}]\right)\right] + \sum_{i\in[m]} \max \operatorname{tr}\left[H_i \operatorname{tr}_{\overline{C}_i^H}[\sigma_{i_1,\dots,i_r}]\right] \\
&\leq \sum_{i\in[m]} \max \left\|\left(\operatorname{tr}_{\overline{C}_i^H}[\xi'] - \operatorname{tr}_{\overline{C}_i^H}[\sigma_{i_1,\dots,i_r}]\right)\right\|_1 + \sum_{i\in[m]} \max \operatorname{tr}\left[H_i \operatorname{tr}_{\overline{C}_i^H}[\sigma_{i_1,\dots,i_r}]\right] \\
&< m\gamma + a \\
&\leq a + \delta,
\end{aligned}$$

where we used the linearity of the trace, trace distance is nonincreasing under the partial trace, $\|H_i\| \leq 1$ for all i and the fact that the maximisation is performed over all $\sigma_{i_1,\dots,i_r}$ that contain all indices in C_i^H. At the same time, using that Check 2 succeeded, for all $\rho_j \in D$ we must have

$$\begin{aligned}
\left\|\rho_j - \operatorname{tr}_{\overline{C}_j}[\xi']\right\|_1 &\leq \max \left\|\rho_j - \operatorname{tr}_{\overline{C}_j}[\sigma_{i_1,\dots,i_r}]\right\|_1 + \max \left\|\operatorname{tr}_{\overline{C}_j}[\sigma_{i_1,\dots,i_r}] - \operatorname{tr}_{\overline{C}_j}[\xi']\right\|_1 \\
&< \alpha + \gamma \\
&\leq \beta.
\end{aligned}$$

This implies that there must exist a state ξ' with energy $< a + \delta$ such that all $\rho_j \in D$ are strictly less than β-consistent (in terms of trace distance) with ξ', contradicting the promise of a NO-instance. Hence, Check 4 must reject with probability at least 2/3, so the overall procedure is sound. □

Now that we have established QMA-completeness for LEDMV in the parameter regime relevant to our purposes, we turn to the task of constructing a brute-force search over the continuous set of constant-size density matrices. To do this, we introduce the notion of *covering sets of density matrices* in the next subsection.

5.2.2 Covering sets of density matrices

We begin by formally defining an h-covering set of density matrices, which can be viewed as a mixed-state analogue of h-nets for pure states.

5.2.3. Definition (h-covering set of density matrices). Let $\mathcal{H} = \mathbb{C}^d$ for $d \in \mathbb{Z}_+$. We say a finite set of density matrices $D_h^d = \{\rho_i\} \subseteq \mathrm{D}(\mathcal{H})$ is an h-covering of $\mathrm{D}(\mathcal{H})$ if for all $\sigma \in \mathrm{D}(\mathcal{H})$ there exists a $\rho_i \in D_h^d$ such that $\frac{1}{2}\|\rho_i - \sigma\|_1 \leq h$.

Next, we show that for any inverse-polynomially small value of h, one can construct an h-covering set of manageable size. To do so, we use the fact that every q-qubit density matrix admits a purification on $2q$ qubits. Hence, it suffices to construct an h-net over pure states in a $2q$-qubit system, and then trace out the purifying register. This net can be obtained by generating a sufficiently fine discretisation of $\mathbb{SU}(4^q)$, noting that global phases become irrelevant once pure states are expressed as density matrices.

5.2.4. Lemma. *Every $U \in \mathbb{SU}(2^n)$ can be implemented using $\mathcal{O}(n^2 4^n)$ CNOT and single-qubit gates.*

For a proof, see Nielsen and Chuang, Chapter 4 [NC10]. By the Solovay-Kitaev theorem, one can approximate $U \in \mathbb{SU}(2)$ up to error h in the diamond norm using at most $\mathcal{O}(\log^c(1/h))$ gates for some $c > 1$, using *any* inverse-closed universal gate set. However, for our purposes, we need the optimal scaling of $c = 1$ [HRC02]. Many gate sets are known to achieve this for $\mathbb{SU}(2)$, as shown in [HRC02, RS16, FGKM15, BRS15, KMM15, PS18]. Since all we care about is that the gates *can* optimally approximate a unitary in $\mathbb{SU}(2)$ (not necessarily that we can efficiently find the sequence), we simply use the gate set from [HRC02], which originates from [LPS86].

5.2.5. Lemma (Adapted from [HRC02]). *There exists a universal gate set $\mathcal{G}$ with $|\mathcal{G}| = 3$ such that for every $U \in \mathbb{SU}(2)$, there exists a circuit over gates from $\mathcal{G}$ that approximates U up to error ϵ in the operator norm using at most $\mathcal{O}\left(\log(1/\epsilon)\right)$ gates.*

We now have all the necessary ingredients to construct h-covering sets of density matrices in polynomial time for any constant number of qubits.

5.2.6. Lemma. *Let $q \in \mathbb{Z}_+$. Then for all $0 < h \leq 1$, there exists a polynomial-time algorithm that constructs an h-covering set of density matrices $D_h^{2^q}$ of size at most* $\mathrm{poly}(1/h)$ *in time* $\mathrm{poly}(1/h)$.

Proof:
For any q-qubit density matrix ρ, there exists a purification $|\psi\rangle$ in a $2q$-qubit system. By Uhlmann's Theorem (Lemma 2.2.3), the fidelity between two mixed states equals the maximum fidelity between their purifications. Hence, it suffices to create an h-net over 4^q-dimensional pure states, which we obtain by approximating $\mathbb{SU}(4^q)$.

Let $\mathcal{G}'$ be the gate set from Lemma 5.2.5, and let $\mathcal{G} = \mathcal{G}' \cup \{\mathrm{CNOT}\}$. Since the global phase is irrelevant when considering density matrices, we only need to approximate unitaries in $\mathbb{SU}(4^q)$. We construct the h-covering set $D_h^{2^q} = \{\rho_i\}$ by setting $\rho_i = \mathrm{tr}_B |\psi_i\rangle\langle\psi_i|$, where $|\psi_i\rangle = U_i |0 \ldots 0\rangle$ and each U_i is a circuit over $\mathcal{G}$ that approximates some $U \in \mathbb{SU}(4^q)$ up to error h. By Lemma 5.2.4, any such U can be implemented using at most $m := C_1 q^2 4^{2q}$ CNOT and 1-qubit gates, for some constant $C_1 > 0$. By Lemma 5.2.5, each 1-qubit gate can be approximated to precision h using $\mathcal{O}(\log(1/h))$ gates from $\mathcal{G}'$. Since the errors of each approximation of the single qubit gates accumulate linearly, any U can be approximated up to error h using a circuit of depth at most $M = C_2 m \log(m/h)$, for some constant $C_2 > 0$. Using that $|\mathcal{G}| = |\mathcal{G}'| + 1 = 4$, the total number of possible circuits on $2q$ qubits of depth M, using gates from $\mathcal{G}$, can be upper bounded as

$$\begin{aligned}\left(4\binom{2q}{2}\right)^{C_2 m \log(m/h)} &\leq \left(16q^2\right)^{C_2 m \log(m/h)} \\ &= \left(16q^2\right)^{C_2 m \log m} \cdot (1/h)^{C_2 m \log\left(16q^2\right)} = \mathrm{poly}(1/h),\end{aligned}$$

for constant q. Hence, we can efficiently enumerate over this set of circuits, which means we can generate $D_h^{2^q}$ in $\mathrm{poly}(1/h)$ time. This also implies that $|D_h^{2^q}| = \mathrm{poly}(1/h)$, as desired. □

5.2.3 The algorithm

We can now state the QMA query algorithm to find all q-qubit marginals of a low-energy state of a k-local Hamiltonian in Algorithm 5.2.1. It is important to stress that J is different from the set I from Protocol 5.2.1, which depends on r, i.e., the maximum of q and k. In particular, it does not have to depend on k, as the k-dependence is implicit in query to the LEDMV-instance (which has H

amongst its inputs).

Algorithm 5.2.1: Randomized QMA-query algorithm to find ϵ-approximations of the q-local density matrices of a low-energy state of a k-local Hamiltonian H.

Input: Classical descriptions of k-local terms $\{H_i\}$, a locality parameter q, and an accuracy parameter ϵ.

Set: $\alpha \coloneqq \epsilon/2$, $\beta \coloneqq \epsilon$, $J \coloneqq \binom{[n]}{q}$, $\delta \coloneqq \frac{a}{|J|+1}$, and $h \coloneqq \epsilon/2$.

Algorithm:

1. Perform binary search on the local Hamiltonian problem corresponding to H using the QMA oracle to obtain an estimate $\hat{\lambda}_0$ such that $\lambda_0(H) \in [\hat{\lambda}_0 - \delta, \hat{\lambda}_0 + \delta]$. Define the set $\{a_l \mid a_l = \hat{\lambda}_0 + l\delta\}_{l \in [|J|]}$.
2. Construct a h-covering set of q-qubit density matrices $D_h^{2^q}$.
3. For each $(i_1, \ldots, i_q) \in J$:

   Suppose we are at the lth step and have already obtained $\{\rho_{i_1,\ldots,i_q}\}_{(i_1,\ldots,i_q)\in S} =: \Sigma$ for some $S \subseteq J$ with $|S| = l$.

   (a) For each $\rho \in D_h^{2^q}$:

      i. **Partial assignment guess:** Set $\Sigma' \coloneqq \Sigma \cup \{\rho\}$.

      ii. **Partial assignment verification:** Make a single query to the QMA oracle on the LEDMV-instance with input H, Σ', q, a_l, δ, α, β. If the outcome is YES, exit the loop, update $\Sigma \leftarrow \Sigma'$, and increment $l + 1 \leftarrow l$.
4. Output Σ (and optionally $\hat{\lambda}_0$).

The key idea behind Algorithm 5.2.1 is that even density matrices corresponding to inputs *within* the promise set of an LEDMV instance maintain sufficient precision for our desired approximation. This effectively creates a decision problem where the soundness parameter serves as an upper bound on precision. This concept stems from the nature of making oracle queries to promise problems: when you conclude that something is a YES instance, all you can be certain of is that it is *not* a NO instance.

Since density matrices are constructed through partial assignments, each step introduces a potential error. Therefore, it is important to ensure that these errors

remain small enough so that the state, which the density matrices approximately represent, does not significantly increase in energy.

5.2.7. THEOREM. *Let $H = \sum_{i\in[m]} H_i$ be a k-local Hamiltonian on n qubits with $m = \mathrm{poly}(n)$ terms, where $0 \preceq H_i \preceq \mathbb{I}$. Let $J = \binom{[n]}{q}$ and let C_j denote the jth element of J under lexicographic ordering. Then for every $q = \mathcal{O}(1)$, every $\epsilon \geq 1/\mathrm{poly}(n)$ and every $a \geq 1/\mathrm{poly}(n)$, there exists a polynomial-time algorithm making queries to a* QMA *oracle that outputs a set of q-local density matrices $\{\rho_j\}$ such that there exists a state ξ with $\mathrm{tr}[H\xi] \leq \lambda_0(H) + a$, for which*

$$\left\|\rho_j - \mathrm{tr}_{\overline{C}_j}[\xi]\right\|_1 \leq \epsilon \quad \textit{for all } j \in [|J|].$$

Proof:
We prove correctness and analyse the complexity of Algorithm 5.2.1.

Correctness. The correctness of Step 1 follows from the techniques in [Amb14, GY19], which show that binary search over the promise gap of the local Hamiltonian problem can be implemented using a QMA oracle. Since LEDMV is QMA-complete for the stated parameters, the query in Step 1 can be implemented using a polynomial-time Karp reduction from LH to LEDMV.

At each step l, the algorithm attempts to extend a partial collection Σ of marginals to a new collection $\Sigma' = \Sigma \cup \{\rho\}$ by querying whether this partial assignment is consistent with a low-energy state. If the oracle returns YES, we are guaranteed the existence of a state ξ_l with energy $\mathrm{tr}[H\xi_l] \leq \lambda_0(H) + a_l + \delta$ that satisfies

$$\left\|\rho_j - \mathrm{tr}_{\overline{C}_j}[\xi_l]\right\|_1 \leq \epsilon$$

for all $j \in [l]$. Repeating this process for each of the $\mathrm{poly}(n)$ subsets in J, we obtain a full collection $\{\rho_j\}_{j\in[J]}$ such that for the final state $\xi \coloneqq \xi_{|J|}$, we have

$$\mathrm{tr}[H\xi] \leq \lambda_0(H) + a_{|J|} + \delta \leq \lambda_0(H) + a,$$

and

$$\left\|\rho_j - \mathrm{tr}_{\overline{C}_j}[\xi]\right\|_1 \leq \epsilon \quad \text{for all } j \in [|J|],$$

since each update was made using an oracle that verified this trace-distance consistency with error parameter $\beta = \epsilon$.

Complexity. Step 1 requires $\mathcal{O}(\log n)$ queries to the QMA oracle to locate the ground state energy to precision $\delta \geq 1/\mathrm{poly}(n)$. Step 2 runs in $\mathrm{poly}(n)$ time for constant q and $\epsilon \geq 1/\mathrm{poly}(n)$ by Lemma 5.2.6. Each iteration of the inner loop (Step 3a) runs in time polynomial in $1/\epsilon$ and the size of the covering set, which

is polynomially bounded when $\epsilon \geq 1/\text{poly}(n)$ and $q = \mathcal{O}(1)$. Since $|J| = \text{poly}(n)$, the total number of iterations is polynomial, and each query to the QMA oracle is made on an instance of LEDMV with inverse polynomial promise gap. Hence, the full algorithm runs in polynomial time. □

We note that it is straightforward to show that if the Hamiltonian has an inverse polynomially bounded spectral gap, we can ensure that the density matrices are guaranteed to be good approximations of the marginals of the ground state (the proof can be found in the full work [Weg24]).

5.3 Finding marginals of near-optimal QMA witnesses

In this section, we extend the results of Section 5.2 by showing that approximate descriptions of the local density matrices of a near-optimal witness can be obtained for *any* problem in QMA. More precisely, given a promise problem A in QMA, the goal is to produce approximate classical descriptions of all q-local marginals corresponding to a proof whose acceptance probability is within $1/\text{poly}(n)$ of the optimal witness. The key idea is to construct an *approximately witness-preserving reduction* from any QMA problem to the local Hamiltonian problem, such that recovering low-energy marginals for the resulting Hamiltonian is equivalent to recovering marginals of a nearly-optimal witness for the original problem.

5.3.1 Approximately witness-preserving reductions in QMA

To make this chapter more self-contained, we begin by recalling the small-penalty clock construction from Chapter 3. This construction transforms a quantum verification circuit U_n from a P-uniform family of QMA verification circuits $\mathcal{U} = \{U_n : n \in \mathbb{N}\}$, each consisting of T gates from a universal set of at most 2-local gates and acting on an input x and a quantum witness $|\psi\rangle \in (\mathbb{C}^2)^{\otimes \text{poly}(n)}$, into a k-local Hamiltonian of the form

$$H_{\text{SPCC},k} = H_{\text{in}} + H_{\text{stab}} + H_{\text{prop}} + \epsilon_{\text{penalty}} H_{\text{out}}, \tag{5.1}$$

where the locality k depends on the specific construction used, and $\epsilon_{\text{penalty}} > 0$ is a tunable parameter.[2] For our purposes, the exact form of these terms is not essential; however, precise descriptions of the 3-local version we use (i.e., $H_{\text{SPCC},3}$) can be found in [KR03]. As usual, to ease notation, we may omit the subscript n from U and assume n is fixed.

[2] We change the notation of the penalty parameter in this chapter, since ϵ is already used as a precision parameter.

As we know from Chapter 3, the ground state of the first three terms, $H_0 \coloneqq H_{\text{in}} + H_{\text{stab}} + H_{\text{prop}}$, is given by the *history states*, which have zero energy with respect to H_0 and are defined as

$$|\eta(\psi)\rangle = \frac{1}{\sqrt{T+1}} \sum_{t=0}^{T} V_t \dots V_1 |\psi\rangle |0\rangle |t\rangle , \tag{5.2}$$

where $|\psi\rangle \in (\mathbb{C}^2)^{\otimes \text{poly}(n)}$ is a quantum witness and t represents the time step of the computation. We have that if U accepts $(x, |\psi\rangle)$ with probability p, then the corresponding history state has energy

$$\langle \eta(\psi)| H^x_{\text{FK}} |\eta(\psi)\rangle = \epsilon \frac{1-p}{T+1}. \tag{5.3}$$

Moreover, we recall that, by linearity, we have

$$\alpha_1 |\eta(|\psi_1\rangle)\rangle + \alpha_2 |\eta(|\psi_2\rangle)\rangle = |\eta(\alpha_1 |\psi_1\rangle + \alpha_2 |\psi_2\rangle)\rangle ,$$

so any linear combination of history states is itself a history state.

We reformulate two lemmas from Chapter 3 to make the dependence on the spectral gap explicit, and we provide a lower bound on the spectral gap of H_0, taken from [DGF22], which applies to the 3-local clock construction of [KR03].

5.3.1. Lemma ([DGF22]). *$H_0 = H_{\text{SPCC},3} - \epsilon_{\text{penalty}} H_{\text{out}}$ has a spectral gap of $\Delta \geq C/T^3$ for some constant $C > 0$.*

5.3.2. Lemma (Reformulated Lemma 3.2.3, from [DGF22]). *Let $\mathcal{U} = \{U_n : n \in \mathbb{N}\}$ be a P-uniform family of QMA verification circuits. Fix an input size n and consider an input $x \in \{0,1\}^n$. Suppose U_n consists of $T = \text{poly}(n)$ gates from some universal gate-set using at most 2-local gates. Denote $P(\psi)$ for the probability that U_n accepts $(x, |\psi\rangle)$, and let $H_{\text{SPCC},3}$ be the corresponding 3-local Hamiltonian from the circuit-to-Hamiltonian mapping in [KR03] with an $\epsilon_{\text{penalty}}$-factor in front of H_{out}. Then there exists a constant $c_0 > 0$, such that for all $0 < \epsilon_{\text{penalty}} \leq \Delta/16$, we have that within the low-energy subspace $\mathcal{S}_{\epsilon_{\text{penalty}}}$ of $H_{\text{SPCC},3}$, i.e.,*

$$\mathcal{S}_{\epsilon_{\text{penalty}}} = \text{span}\{|\Phi\rangle : \langle \Phi| H_{\text{SPCC},3} |\Phi\rangle \leq \epsilon_{\text{penalty}}\}$$

the eigenvalues λ_i satisfy

$$\lambda_i \in \left[\epsilon_{\text{penalty}} \frac{1-P(\psi_i)}{T+1} - c_0 \frac{\epsilon^2_{\text{penalty}}}{\Delta}, \epsilon_{\text{penalty}} \frac{1-P(\psi_i)}{T+1} + c_0 \frac{\epsilon^2_{\text{penalty}}}{\Delta} \right], \tag{5.4}$$

where $\{|\psi_i\rangle\}$ are the eigenstates of the Mariott-Watrous operator of the circuit U_n, given by

$$Q = \left(\langle x| \otimes \mathbb{I} \otimes \langle 0|^{\otimes q(n)} \right) U_n^\dagger \left(|0\rangle\langle 0| \otimes \mathbb{I} \right) U_n \left(|x\rangle \otimes \mathbb{I} \otimes |0\rangle^{\otimes q(n)} \right).$$

5.3.3. Lemma (Reformulated Lemma 3.2.5). *Let $|\Psi\rangle$ be a state such that*

$$\langle\Psi| H_{\mathrm{SPCC},3} |\Psi\rangle \leq \delta$$

and let Δ be the spectral gap of H_0. Write Π_{hist} for the projector onto the subspace spanned by history states. Then,

$$\|\Pi_{\mathrm{hist}} |\Psi\rangle\|^2 \geq 1 - \frac{\delta}{\Delta}.$$

We now aim to precisely characterise the maximum acceptance probability of a witness encoded in a history state, as a function of the state's energy relative to the ground state energy of H. This is formalised in the following lemma.

5.3.4. Lemma. *Let p^* be the maximum acceptance probability of a* QMA *verification circuit. Let $H_{\mathrm{SPCC},3}$ be the Hamiltonian as in Eq.* (5.1) *resulting from the small-penalty clock construction for some $0 < \epsilon_{\mathrm{penalty}} < \Delta/16$, with ground state energy $\lambda_0(H_{\mathrm{SPCC},3})$. Suppose we are given a state $|\Psi\rangle$ with an energy at most*

$$\lambda_0(H_{\mathrm{SPCC},3}) + c_0 \frac{\epsilon_{\mathrm{penalty}}^2}{\Delta}.$$

Then we have that $|\Psi\rangle$ has fidelity at least

$$1 - \left(\frac{\epsilon_{\mathrm{penalty}}}{\Delta} \frac{1-p^*}{T+1} + 2c_0 \left(\frac{\epsilon_{\mathrm{penalty}}}{\Delta} \right)^2 \right)$$

with a history state $|\eta(\psi)\rangle$ for some witness $|\psi\rangle$ which has an acceptance probability $\tilde{p}$ satisfying

$$\begin{aligned} p^* - \tilde{p} \leq (T+1) 2c_0 \frac{\epsilon_{\mathrm{penalty}}}{\Delta} + 2(T+1) \sqrt{\frac{\epsilon_{\mathrm{penalty}}}{\Delta} \frac{1-p^*}{T+1} + 2c_0 \left(\frac{\epsilon_{\mathrm{penalty}}}{\Delta} \right)^2} \\ + \frac{\epsilon_{\mathrm{penalty}}}{\Delta} \frac{1-p^*}{T+1} + 2c_0 \left(\frac{\epsilon_{\mathrm{penalty}}}{\Delta} \right)^2. \end{aligned}$$

Proof:
By Lemma 5.3.2, we have that the ground state energy of $H_{\mathrm{SPCC},3}$ satisfies

$$\lambda_0(H_{\mathrm{SPCC},3}) \in \left[\epsilon_{\mathrm{penalty}} \frac{1-p^*}{T+1} - c_0 \frac{\epsilon_{\mathrm{penalty}}^2}{\Delta}, \epsilon_{\mathrm{penalty}} \frac{1-p^*}{T+1} + c_0 \frac{\epsilon_{\mathrm{penalty}}^2}{\Delta} \right].$$

Hence, we have that $|\Psi\rangle$ has an energy at most

$$\langle\Psi| H_{\mathrm{SPCC},3} |\Psi\rangle \leq \epsilon_{\mathrm{penalty}} \frac{1-p^*}{T+1} + 2c_0 \frac{\epsilon_{\mathrm{penalty}}^2}{\Delta} =: \delta. \tag{5.5}$$

Note the extra factor of 2 incurred because of the theorem's assumptions. We can write any state $|\Psi\rangle$ in the eigenbasis of H_0 as

$$|\Psi\rangle = \alpha\,|\text{hist}\rangle + \sqrt{1-\alpha^2}|\text{hist}^\perp\rangle, \tag{5.6}$$

for some real $\alpha \in [0,1]$, where $|\text{hist}\rangle$ lives in the space spanned by the history states and $|\text{hist}^\perp\rangle$ in the space orthogonal to it. In its eigenbasis, H_0 is diagonal. Note that $\alpha^2 = \|\Pi_{\text{hist}}\,|\Psi\rangle\|^2$. Hence, by Lemma 5.3.3 it must hold that

$$\alpha \geq \sqrt{1-\frac{\delta}{\Delta}} = \sqrt{1-\left(\frac{\epsilon_{\text{penalty}}}{\Delta}\frac{1-p^*}{T+1} + 2c_0\left(\frac{\epsilon_{\text{penalty}}}{\Delta}\right)^2\right)}.$$

We expand the energy using the decomposition of $|\Psi\rangle$ in the eigenbasis of H_0 using Eq. (5.6) as

$$\begin{aligned}\langle\Psi|\,H_{\text{SPCC},3}\,|\Psi\rangle &= \left(\alpha\,\langle\text{hist}| + \sqrt{1-\alpha^2}\langle\text{hist}^\perp|\right) H_{\text{SPCC},3}\left(\alpha\,|\text{hist}\rangle + \sqrt{1-\alpha^2}|\text{hist}^\perp\rangle\right)\\ &= \alpha^2\,\langle\text{hist}|\,H_{\text{SPCC},3}\,|\text{hist}\rangle + \alpha\sqrt{1-\alpha^2}\,\langle\text{hist}|\,H_{\text{SPCC},3}|\text{hist}^\perp\rangle\\ &\quad + \alpha\sqrt{1-\alpha^2}\langle\text{hist}^\perp|H_{\text{SPCC},3}\,|\text{hist}\rangle + (1-\alpha^2)\langle\text{hist}^\perp|H_{\text{SPCC},3}|\text{hist}^\perp\rangle.\end{aligned}$$

We now want to lower bound $\langle\Psi|\,H_{\text{SPCC},3}\,|\Psi\rangle$ in terms of $\langle\text{hist}|\,H_{\text{SPCC},3}\,|\text{hist}\rangle$ to compare with our upper bound in Eq. (5.5). To do this, we must find lower bounds on the other three terms in the expression. For the first one we have

$$\begin{aligned}\langle\text{hist}|\,H_{\text{SPCC},3}|\text{hist}^\perp\rangle &= \langle\text{hist}|\,H_0 + \epsilon_{\text{penalty}}H_{\text{out}}|\text{hist}^\perp\rangle\\ &= \epsilon_{\text{penalty}}\langle\text{hist}|\,H_{\text{out}}|\text{hist}^\perp\rangle\\ &\geq -\epsilon_{\text{penalty}},\end{aligned}$$

using that $\|H_{\text{out}}\| = 1$ and that $\langle\text{hist}|\,H_0|\text{hist}^\perp\rangle = 0$, which holds since $|\text{hist}\rangle$,$|\text{hist}^\perp\rangle$ live in separate eigenspaces of H_0. Similarly, for the second term it must also hold that $\left\langle\text{hist}^\perp\right|\,H_{\text{SPCC},3}\,|\text{hist}\rangle \geq -\epsilon_{\text{penalty}}$. And finally, for the third term we have $\langle\text{hist}^\perp|H_{\text{SPCC},3}|\text{hist}^\perp\rangle \geq \Delta \geq 0$. Putting this all together, we have

$$\langle\Psi|\,H_{\text{SPCC},3}\,|\Psi\rangle \geq \alpha^2\,\langle\text{hist}|\,H_{\text{SPCC},3}\,|\text{hist}\rangle - 2\alpha\sqrt{1-\alpha^2}\epsilon_{\text{penalty}}, \tag{5.7}$$

Suppose that $|\text{hist}\rangle$ encodes a witness with acceptance probability $\tilde{p}$ (recall that linear combinations of history states are also history states). We have that

$$\langle\text{hist}|\,H_{\text{SPCC},3}\,|\text{hist}\rangle = \epsilon_{\text{penalty}}\frac{1-\tilde{p}}{T+1}.$$

Plugging this into Eq. (5.7) and combining the resulting expression with Eq. (5.5) gives

$$\epsilon_{\text{penalty}}\frac{1-p^*}{T+1} + 2c_0\frac{\epsilon^2}{\Delta} \geq \alpha^2\epsilon_{\text{penalty}}\frac{1-\tilde{p}}{T+1} - 2\alpha\sqrt{1-\alpha^2}\epsilon_{\text{penalty}}$$

which after rearranging to get $p^* - \alpha^2\tilde{p}$ at the LHS of the inequality results in

$$p^* - \alpha^2\tilde{p} \leq (T+1)2c_0\frac{\epsilon_{\text{penalty}}}{\Delta} + 2\alpha(T+1)\sqrt{1-\alpha^2} + 1 - \alpha^2.$$

which gives, using our bounds on α and the fact that $p^* - \alpha^2\tilde{p} \geq p^* - \tilde{p}$ as $p^* \geq \tilde{p} \geq 0$ and $\alpha \in [0,1]$, as well as Lemma 5.3.1,

$$\begin{aligned} p^* - \tilde{p} \leq (T+1)2c_0\frac{\epsilon_{\text{penalty}}}{\Delta} + 2(T+1)\sqrt{\frac{\epsilon_{\text{penalty}}}{\Delta}\frac{1-p^*}{T+1} + 2c_0\left(\frac{\epsilon_{\text{penalty}}}{\Delta}\right)^2} \\ + \frac{\epsilon_{\text{penalty}}}{\Delta}\frac{1-p^*}{T+1} + 2c_0\left(\frac{\epsilon_{\text{penalty}}}{\Delta}\right)^2, \end{aligned}$$

which completes the proof. □

However, being close to a history state is not sufficient for our purposes: we require closeness to an actual witness state $|\psi\rangle$ tensored with an auxiliary state that we do not care about. To achieve this, we make again use of the pre-idling technique from Chapter 4. This technique ensures that all history states become close to product states of the form $|\psi\rangle \otimes |\Phi\rangle$, where $|\Phi\rangle$ is a fixed, known state (typically encoding the clock and ancilla registers).

5.3.5. Lemma. *Let $U = V_T \dots V_1$ be a* QMA *verification circuit consisting of T gates. Define $\tilde{U} = \tilde{V}_{T+M} \dots \tilde{V}_1$ to be the circuit obtained by prepending M identity gates to V, and let $H_{\text{SPCC},3}$ be the Hamiltonian from Eq. (5.1) which uses $\tilde{U}$. Then for any history state $|\eta(\psi)\rangle$ corresponding to a witness $|\psi\rangle$, there exists a state of the form $|\psi\rangle \otimes |\Phi\rangle$, where $|\Phi\rangle$ is a fixed state depending only on M and the size of the ancilla register of U, such that*

$$|\langle\eta(\psi)|\,(|\psi\rangle \otimes |\Phi\rangle)|^2 = \frac{M}{M+T+1}.$$

Proof:
Let $|\eta(\psi)\rangle$ be the history state associated with $\tilde{U}$, i.e,

$$|\eta(\psi)\rangle = \frac{1}{\sqrt{M+T+1}} \sum_{t=0}^{M+T} \tilde{V}_t \dots \tilde{V}_1 |\psi\rangle |0^a\rangle |\bar{t}\rangle,$$

where a denotes the total number of ancilla qubits and $\bar{t}$ a unary encoding of t. Now define

$$|\Phi\rangle := \frac{1}{\sqrt{M}} \sum_{t=0}^{M-1} |0^a\rangle |\bar{t}\rangle,$$

and consider the state $|\psi\rangle\otimes|\Phi\rangle$. A direct calculation shows $|\langle\eta(\psi)|(|\psi\rangle\otimes|\Phi\rangle)|^2 = M/(M+T+1)$. □

We are now ready to combine all of the above to construct our approximately witness-preserving reduction. This reduction allows us to recover a witness close to the optimal (i.e., highest-accepting) witness for a given QMA verification circuit by finding a low-energy state of an associated local Hamiltonian instance instead.

5.3.6. THEOREM. *Let $A = (A_{\text{YES}}, A_{\text{NO}})$ be a promise problem in* QMA *with a* P*-uniform family of* QMA *verification circuits $\{U_n : n \in \mathbb{N}\}$, where each U_n uses $T = \text{poly}(n)$ gates and has a witness register denoted by W. Let $x \in \{0,1\}^n$ be an input, and let p^* be the optimal probability that U_n accepts x, i.e., maximised over all quantum proofs in W. For any polynomially bounded functions $p_1(n), p_2(n) \geq 1$, define*

$$M := (4p_2(n))^2 (T+1), \quad \epsilon_{\text{penalty}} := \frac{\min\{C,1\}}{100(c_0+1)(\tilde{T}+1)^7 (p_1(n)\cdot p_2(n))^2},$$

where $\tilde{T} = M + T$. Then, there exists a polynomial-time reduction from a M-pre-idled verification circuit $\tilde{U}_n$ with $\tilde{T} = M + T$ gates, to a local Hamiltonian $H_{\text{SPCC},3}$ such that any state with $|\Psi\rangle$ that satisfies

$$\langle\Psi| H_{\text{SPCC},3} |\Psi\rangle \leq \lambda_0(H_{\text{SPCC},3}) + c_0\epsilon_{\text{penalty}}^2 \tilde{T}^3$$

it holds that $\|\text{tr}_{\overline{W}} |\Psi\rangle\langle\Psi| - |\psi\rangle\langle\psi|\|_1 \leq 1/2p_2(n)$ for some quantum witness $|\psi\rangle$, which satisfies the property that U_n accepts $(x, |\psi\rangle)$ with probability at least $p^* - 1/p_1(n)$.

Proof:
By Lemma 5.3.5, we can use pre-idling with M gates, creating a new circuit $\tilde{U}_n$ with $\tilde{T} = M + T$ gates such that

$$\begin{aligned} \| |\eta(\psi)\rangle\langle\eta(\psi)| - |\psi\rangle|\Phi\rangle\langle\psi|\langle\Phi| \|_1 &= \sqrt{1 - |\langle\eta(\psi)|(|\psi\rangle\otimes|\Phi\rangle)|^2} \\ &= \sqrt{1 - \frac{M}{M+T+1}} \\ &\leq \frac{1}{4p_2(n)} \end{aligned}$$

if $M \geq (4p_2(n)-1)(T+1)$, which is satisfied by our choice of M. The statement in the theorem then consequently holds since the trace distance is nonincreasing upon taking the partial trace (taken over the non-witness registers). Hence, we can take $\tilde{T} = T + M = \text{poly}(n)$ in the new circuit. By Lemma 5.3.1, we have that the spectral gap Δ for our H_0 corresponding to the new circuit $\tilde{U}_n$ satisfies $\Delta \geq C/\tilde{T}^3$. By Lemma 5.3.4 we have that for our choice of $\epsilon_{\text{penalty}}$ that if we

are given a state $|\Psi\rangle$ with energy at most $\lambda_0(H_{\mathrm{SPCC},3}) + 2c_0\epsilon_{\mathrm{penalty}}^2\tilde{T}^3$, it has trace distance at most

$$\begin{aligned}\left\| |\Psi\rangle\langle\Psi| - |\eta(\psi)\rangle\langle\eta(\psi)| \right\|_1 &= \sqrt{1 - |\langle\Psi|\eta(\psi)\rangle|^2} \\ &\leq \sqrt{\frac{\epsilon_{\mathrm{penalty}}}{\Delta}\frac{1-p^*}{\tilde{T}+1} + 2c_0\left(\frac{\epsilon_{\mathrm{penalty}}}{\Delta}\right)^2} \\ &\leq \frac{1}{4p_2(n)},\end{aligned}$$

with a history state $|\eta(\psi)\rangle$ for some witness $|\psi\rangle$ with acceptance probability $\tilde{p}$ which satisfies

$$\begin{aligned}p^* - \tilde{p} &\leq (\tilde{T}+1)2c_0\frac{\epsilon_{\mathrm{penalty}}}{\Delta} + 2(\tilde{T}+1)\sqrt{\frac{\epsilon_{\mathrm{penalty}}}{\Delta}\frac{1-p^*}{\tilde{T}+1} + 2c_0\left(\frac{\epsilon_{\mathrm{penalty}}}{\Delta}\right)^2} \\ &\quad + \frac{\epsilon_{\mathrm{penalty}}}{\Delta}\frac{1-p^*}{\tilde{T}+1} + 2c_0\left(\frac{\epsilon_{\mathrm{penalty}}}{\Delta}\right)^2 \\ &\leq \frac{1}{p_1(n)}.\end{aligned}$$

Hence, by the triangle inequality,

$$\begin{aligned}\left\| |\Psi\rangle\langle\Psi| - |\psi\rangle|\Phi\rangle\langle\psi|\langle\Phi| \right\|_1 &\leq \left\| |\Psi\rangle\langle\Psi| - |\eta(\psi)\rangle\langle\eta(\psi)| \right\|_1 \\ &\quad + \left\| |\eta(\psi)\rangle\langle\eta(\psi)| - |\psi\rangle|\Phi\rangle\langle\psi|\langle\Phi| \right\|_1 \\ &\leq \frac{1}{2p_2(n)}.\end{aligned}$$

The desired result directly follows since the trace distance is non-increasing under the partial trace. □

In the above theorem we have left the dependence on the $\epsilon_{\mathrm{penalty}}$-parameter explicit in the energy bound, since we do not know in advance what $\lambda_0(H_{\mathrm{SPCC},3})$ will be (even after fixing $\epsilon_{\mathrm{penalty}}$), as it depends on the maximum acceptance probability p^*. However, in the next section, we will see that this is not an issue, as we can estimate the ground state energy using QMA oracle access, as shown in Section 5.2.

Finally, we need to show that Theorem 5.3.6 continues to hold in the mixed-state setting, which is important since Algorithm 5.2.1 only returns density matrices that are promised to be approximately consistent with some global density matrix (rather than a pure state).

5.3.7. Corollary. *Under the same assumptions and parameter choices as Theorem 5.3.6, let ξ be a mixed state such that*

$$\mathrm{tr}\left[H_{\mathrm{SPCC},3}\,\xi\right] \leq \lambda_0(H_{\mathrm{SPCC},3}) + c_0\epsilon_{\mathrm{penalty}}^2\tilde{T}^3.$$

Then there exists a quantum witness ξ_{proof} *such that*

$$\|\text{tr}_{\overline{W}}[\xi] - \xi_{\text{proof}}\|_1 \leq \frac{1}{2p_2(n)},$$

and the original verifier U *accepts* (x, ξ_{proof}) *with probability at least* $p^* - 1/p_1(n)$.

Proof:
We omit the subscript n in the verification circuit U, and assume U is already pre-idled as in Theorem 5.3.6. Suppose $H_{\text{SPCC},3}$ acts on $q(n)$ qubits. Extend U to $U_{\text{ext}} = U \otimes \mathbb{I}$ acting on the extended proof register $W_{\text{ext}} = W \cup W'$, where W' consists of $q(n)$ additional qubits and W is our original proof register. The corresponding circuit-to-Hamiltonian mapping becomes $H_{\text{ext}} = H_{\text{SPCC},3} \otimes \mathbb{I}$, and acts on $2q(n)$ qubits. Now let ξ be a $q(n)$-qubit mixed state satisfying the energy condition. Then there exists a purification $|\Phi\rangle \in \mathbb{C}^{2^{2q(n)}}$ such that $\text{tr}_{W'}[|\Phi\rangle\langle\Phi|] = \xi$ and

$$\text{tr}[H_{\text{ext}} |\Phi\rangle\langle\Phi|] \leq \lambda_0(H_{\text{SPCC},3}) + c_0 \epsilon_{\text{penalty}}^2 \tilde{T}^3.$$

Since H_{ext} corresponds to U_{ext}, Theorem 5.3.6 implies that there exists a $(p(n) + q(n))$-qubit proof state $|\Psi\rangle$ such that

$$\left\|\text{tr}_{\overline{W_{\text{ext}}}}[|\Phi\rangle\langle\Phi|] - |\Psi\rangle\langle\Psi|\right\|_1 \leq \frac{1}{2p_2(n)},$$

and V_{ext} accepts $(x, |\Psi\rangle\langle\Psi|)$ with probability at least $p^* - 1/p_1(n)$. Taking the partial trace over the remaining registers to isolate W, and using that $\text{tr}_{\overline{W}}[\xi] = \text{tr}_{\overline{W}}[|\Phi\rangle\langle\Phi|]$, we obtain

$$\|\text{tr}_{\overline{W}}[\xi] - \text{tr}_{\overline{W}}[|\Psi\rangle\langle\Psi|]\|_1 \leq \frac{1}{2p_2(n)},$$

where we used the monotonicity of the trace distance under partial trace. Setting $\xi_{\text{proof}} := \text{tr}_{\overline{W}}[|\Psi\rangle\langle\Psi|]$ completes the proof. □

We now have everything in place to leverage our approximately witness-preserving reduction and construct an algorithm that, for any problem in QMA, produces approximations of all q-local density matrices corresponding to a nearly-optimal accepting witness in the final subsection.

5.3.2 Finding marginals of high-accepting QMA witnesses

Let J be the set of all q-element subsets of the indices of the qubits on which H_{FK}^x is defined, and $J_W \subset J$ the set of all q-element index combinations of indices corresponding to the proof register. After we pre-idle the circuit U_n and

construct the corresponding $H_{\mathrm{SPCC},3}$ for the some choice of $\epsilon_{\mathrm{penalty}}$, we simply run the Algorithm 5.2.1 for $H_{\mathrm{SPCC},3}$ to obtain all density matrices with indices from the set J and finally keep only those with indices from J_W. The full algorithm is given in Algorithm 5.3.1.

Algorithm 5.3.1: QMA query algorithm to find approximations of the q-local density matrices of high-accepting witnesses

Input: Classical descriptions of a QMA verification circuit U_n, functions $p_1(n)$, $p_2(n)$, and locality parameter q.

Set: Define M, $\tilde{T}$, and $\epsilon_{\mathrm{penalty}}$ as in Theorem 5.3.6, and set $a \coloneqq c_0 \tilde{T}^3 \epsilon_{\mathrm{penalty}}^2$, $\epsilon \coloneqq 1/(2p_2(n))$.

Algorithm:

1. Construct the M-pre-idled circuit $\tilde{U}_n$ from U_n.
2. Generate the Hamiltonian $H_{\mathrm{SPCC},3}$ using the small-penalty clock construction from Eq. (5.1) with parameter $\epsilon_{\mathrm{penalty}}$. Let J be the set of all q-element subsets of qubit indices for $H_{\mathrm{SPCC},3}$, and let $J_W \subseteq J$ be the subsets restricted to the witness register W of $\tilde{U}$.
3. Run Algorithm 5.2.1 on $H_{\mathrm{SPCC},3}$ with parameters a and ϵ, obtaining $\{\rho_{i_1,\ldots,i_q}\}_{(i_1,\ldots,i_q)\in J}$ and an estimate $\hat{\lambda}_0(H_{\mathrm{SPCC},3})$.
4. Output the marginals $\{\rho_{i_1,\ldots,i_q}\}_{(i_1,\ldots,i_q)\in J_W}$ and the estimate

$$\hat{p} \coloneqq 1 - \frac{\hat{\lambda}_0(H_{\mathrm{SPCC},3}) \cdot (\tilde{T}+1)}{\epsilon_{\mathrm{penalty}}}.$$

5.3.8. Theorem. *Let $A = (A_{\mathrm{YES}}, A_{\mathrm{NO}})$ be any problem in* QMA *with a* P*-uniform family of verifier circuits $\{U_n\}$, and let $x \in \{0,1\}^n$ be the input. Then for any polynomially bounded functions $p_1(n), p_2(n) \geq 1$, and any constant $q = \mathcal{O}(1)$, there exists a polynomial-time algorithm that makes queries to a* QMA *oracle and outputs:*

- *A value $\hat{p}$ satisfying $|p^* - \hat{p}| \leq 1/p_1(n)$, where p^* is the maximum acceptance probability of U_n on input $(x, |\psi\rangle)$, over all quantum witnesses $|\psi\rangle$.*
- *A set of q-local density matrices $\{\rho_{i_1,\ldots,i_q}\}$, each within trace distance $1/p_2(n)$*

of the corresponding reduced density matrix of some state ξ_{proof} that is accepted by U_n with probability at least $\tilde{p} \geq p^* - 1/p_1(n)$.

Proof:
We will prove that Algorithm 5.3.1 satisfies the criteria of the theorem.

Correctness. Suppose $H_{\text{SPCC},3}$ acts on $p_3(n) = \text{poly}(n)$ qubits. By Theorem 5.2.7, the density matrices $\{\rho_{i_1,\ldots,i_q}\}_{(i_1,\ldots,i_q)\in J}$ returned by Algorithm 5.2.1 correspond to a state ξ satisfying

$$\text{tr}[H_{\text{SPCC},3}\xi] \leq \lambda_0(H_{\text{SPCC},3}) + a = \lambda_0(H_{\text{SPCC},3}) + c_0\tilde{T}^3\epsilon^2_{\text{penalty}},$$

which meets the assumptions of Theorem 5.3.6 (and thus of Corollary 5.3.7). Therefore, Corollary 5.3.7 tells us there exists a state ξ_{proof}, $\|\text{tr}_{\overline{W}}[\xi] - \xi_{\text{proof}}\|_1 \leq 1/2p_2(n)$, that gets accepted with probability $\tilde{p}$ satisfying

$$\tilde{p} \geq p^* - \frac{1}{p_1(n)}.$$

Next, by Lemma 5.3.2, the estimated ground state energy satisfies

$$\hat{\lambda}_0(H_{\text{SPCC},3}) \in \left[\epsilon_{\text{penalty}}\frac{1-p^*}{T+1} \pm \left(c_0\frac{\epsilon^2_{\text{penalty}}}{\Delta} + \frac{a}{|J|+1}\right)\right],$$

which implies

$$p^* \in \left[1 - \frac{\hat{\lambda}_0(H_{\text{SPCC},3})\cdot(\tilde{T}+1)}{\epsilon_{\text{penalty}}} \pm 2\frac{c_0}{C}\epsilon_{\text{penalty}}\tilde{T}^3\right],$$

where we used the choice of a, the fact that $|J| \geq 1$, and the spectral gap bound $\Delta \geq C/\tilde{T}^3$ from Lemma 5.3.1. Since

$$\hat{p} = 1 - \frac{\hat{\lambda}_0(H_{\text{SPCC},3})\cdot(\tilde{T}+1)}{\epsilon_{\text{penalty}}},$$

it follows that for our choice of $\epsilon_{\text{penalty}}$,

$$|p^* - \hat{p}| \leq 2\frac{c_0}{C}\epsilon_{\text{penalty}}\tilde{T}^3 \leq 1/p_1(n).$$

By Theorem 5.2.7, the algorithm returns all q-local density matrices for $J \supseteq J_W$, and each satisfies

$$\left\|\rho_{i_1,\ldots,i_q} - \text{tr}_{[p_3(n)]\setminus\{i_1,\ldots,i_q\}}[\xi]\right\|_1 \leq 1/(2p_2(n)).$$

Combining this with Corollary 5.3.7 and applying the triangle inequality yields

$$\begin{aligned}\left\|\rho_{i_1,\ldots,i_q} - \text{tr}_{[p_3(n)]\setminus\{i_1,\ldots,i_q\}}[\xi_{\text{proof}}]\right\|_1 &\leq \left\|\rho_{i_1,\ldots,i_q} - \text{tr}_{[p_3(n)]\setminus\{i_1,\ldots,i_q\}}[\xi]\right\|_1 \\ &\quad + \left\|\text{tr}_{[p_3(n)]\setminus\{i_1,\ldots,i_q\}}[\xi] - \text{tr}_{[p_3(n)]\setminus\{i_1,\ldots,i_q\}}[\xi_{\text{proof}}]\right\|_1 \\ &\leq 1/p_2(n),\end{aligned}$$

for each $(i_1,\ldots,i_q) \in J_W$.

Complexity. The complexity is polynomial in 2^q and $1/p_2(n)$. Since $q \in \mathcal{O}(1)$ and $p_2(n) = \text{poly}(n)$ by assumption, the algorithm runs in polynomial time. □

5.4 Open problems

We have shown that for any problem in QMA, it is possible to approximate all q-local density matrices of a high-acceptance witness, up to any constant locality, given access to an oracle that solves QMA promise problems. Naturally, it remains open whether QMA admits full search-to-decision reductions that produce the actual quantum witness states. Since Kitaev's circuit-to-Hamiltonian mapping does not relativise, approaching this question in the local Hamiltonian setting would be a natural direction, as it would sidestep the quantum oracle separation found in [INN+22].

5.4.1. Open problem. *Can a low-energy state of a local Hamiltonian be found by a polynomial-time classical algorithm with query access to a* QMA *oracle?*

In this direction, one could also investigate whether placing additional structure on the local Hamiltonians—such as enforcing a spectral gap, geometric locality, or other physical constraints—could make the problem more tractable. While such Hamiltonians may not be QMA-complete in the constrained setting, they may still serve as valuable test cases for understanding whether and how search-to-decision reductions can be achieved.

An interesting open question regarding our construction for finding the density matrices of QMA witnesses is whether the use of a circuit-to-Hamiltonian mapping is truly necessary, or if a more direct approach, based on the trivial QMA-complete problem of circuit verification, could suffice. At present, it is unclear whether such an approach would work. The main obstacle is that the acceptance probability of a quantum verification circuit is inherently a *global* observable, whereas we are only given local information in the form of reduced density matrices of a quantum proof. This stands in contrast to the local Hamiltonian setting, where the energy is naturally expressed as a sum of *local* observables.

Part II.

Quantum probabilistically checkable proof systems

Chapter 6

Adaptivity, multiple provers and reductions to Hamiltonians

6.1 Introduction

A probabilistically checkable proof system (PCP) has a polynomial-time probabilistic verifier with query access to a proof provided by a computationally unbounded prover. Usually, a PCP is specified by two parameters: the number of random coins the verifier is allowed to use (which also upper bounds the length of the proof as some exponential in the number of coins) and the number of queries it can make to the proof. The PCP theorem [ALM+98, AS98], which originated from a long line of research on the complexity of interactive proof systems, states that all problems in NP can be decided, with a constant probability of error, using only a logarithmic number of coin flips and a constant number of queries to the proof.

Equivalently, the PCP theorem can be formulated as the statement that it is NP-hard to decide whether a given CSP instance is fully satisfiable or whether no more than a constant fraction of its constraints can be satisfied. This is usually referred to as the *hardness of approximation* formulation of the PCP theorem. One can also prove the PCP theorem by reducing a CSP directly to another CSP that exhibits this property: this type of transformation, due to Dinur [Din07], is usually referred to as *gap amplification*, referring to the increase in the difference (the gap) in the fraction of constraints that can be satisfied in both the YES- and NO-instances.

As we discussed in Section 1.2, quantum complexity theorists have proposed proof-checking and hardness-of-approximation versions of a quantum PCP conjecture. The hardness-of-approximation formulation states that the local Hamiltonian problem with constant promise gap, relative to the total number of local terms, is QMA-hard. Evidence in favour of this conjecture includes the NLTS theorem [ABN23] and related constructions [CCNN23b, CCNN23a, HATH24, AGK24], which show that there exist Hamiltonians whose low-energy states—

those with energy up to a constant fraction of the total norm of the Hamiltonian—cannot be described by any class of states admitting efficient classical descriptions (which would serve as NP-witnesses).

On the other hand, there are also results giving evidence against the conjecture: various works suggest that the "quantum hardness" of the local Hamiltonian problem disappears in the constant promise gap regime under certain structural assumptions, such as high-degree and good expansion in the interaction graph [BH13b], or stoquasticity [ABG19]. This also includes our results from Chapter 4 and Appendix C, which show that imposing the condition that the ground space has non-negligible overlap with classes of states admitting certain classical access models makes a variation of the local Hamiltonian problem QCMA-hard in the inverse-polynomial precision regime, but contained in NP (or MA) when the precision is constant (see also [GL22]). Not only do these limitations rule out certain structural properties of "quantum PCP candidate" Hamiltonians, they also constrain the kinds of structural transformations a potential gap amplification—à la Dinur, but for Hamiltonians—could exhibit (again, see Appendix C).

The proof-checking formulation of the quantum PCP conjecture, which states that one can solve any promise problem in QMA using a quantum verifier that only accesses a constant number of qubits from a quantum proof, has received considerably less attention. A possible reason for this is that both conjectures are known to be equivalent under *quantum* reductions. This was already pointed out in the first work that proposed a quantum PCP formulation [AALV09].[1]

Perhaps that is why, even after more than two decades since the question of the existence of a quantum PCP was first posed [AN02], many basic questions regarding the proof-checking formulation have not been addressed. For example, as already raised in [AALV09], how robust are definitions of quantum PCPs under subtle changes, e.g., the choice of distribution over which the proof qubits are selected? Are adaptive queries to the proof more powerful than non-adaptive ones in the constant-query setting, or are they equivalent as in the classical setting? Are there non-equivalent variations of quantum PCPs, similar to the many natural variations of QMA?[2]

6.1.1 Results of this chapter

We define a general formulation of quantum PCPs in Section 6.2, capturing adaptivity and multiple unentangled provers, and provide a detailed construction of a quantum reduction to the local Hamiltonian problem with a constant promise

[1] Whilst this is widely known in the community, this reduction has (as far as the author is aware) never been written down in full detail except in the works of [Gri18] and [HATH24], which both consider more restricted formulations of quantum PCPs than considered in this chapter.

[2] See [Gha24] for a recent review.

gap. This reduction, presented in Section 6.3, turns out to be a versatile subroutine for proving structural properties of quantum PCP systems, allowing us to show in Section 6.4:

- Non-adaptive quantum PCPs can simulate adaptive quantum PCPs when the number of proof queries is constant. This holds even if the non-adaptive verifier samples proof indices uniformly at random from a subset of possible combinations, resolving an open question posed by Aharonov, Arad, Landau, and Vazirani [AALV09].

- If the $2q$-local Hamiltonian problem with constant relative promise gap lies in QCMA, then $\mathsf{QPCP}[q] \subseteq \mathsf{QCMA}$.[3]

- If $\mathsf{QMA}(2)$ admits a constant-query, unentangled k-prover, quantum PCP for any $2 \leq k \leq \mathrm{poly}(n)$, then $\mathsf{QMA}(2) = \mathsf{QMA}$, establishing a connection between two long-standing open problems in quantum complexity theory.

6.2 Quantum probabilistically checkable proofs

The bulk of this section is taken up a definition of a quantum PCP. Our definition is more detailed than that given in [AAV13], and more general than the definitions in [HATH24, Gri18], in that we also include the ability to make q queries to proof qubits adaptively and allow for the possibility of having k unentangled provers. Once we have this definition in place, we can define an associated complexity class $\mathsf{QPCP}[k, q]$. For convenience, we also explicitly define a non-adaptive version. For completeness, we also prove that weak error reduction is possible for quantum PCPs.

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

- an n-qubit input register A, initialised in input $|x\rangle$, $x \in \{0,1\}^n$;

- a p_1-qubit ancilla register B, initialised in $|0\rangle^{\otimes p_1(n)}$;

- a kp_2-qubit proof register C, initialised in $\xi = \otimes_{j=1}^{k} \xi_j$ for some quantum witnesses $\xi_j \in \mathrm{D}\left((\mathbb{C}^2)^{\otimes p_2(n)}\right)$ for all $j \in [k]$;

- a collection of PVMs $\Pi^t = \{\Pi_i^t\}$, $i \in [kp_2(n)]$, with $\Pi_i^t = |i\rangle\langle i| \otimes \mathbb{I}$ for all $t \in [q]$, $\Pi^{\mathrm{out}} = \{\Pi_0^{\mathrm{out}}, \Pi_1^{\mathrm{out}}\}$, with $\Pi_0^{\mathrm{out}} = |0\rangle\langle 0| \otimes \mathbb{I}$ and $\Pi_1^{\mathrm{out}} = |1\rangle\langle 1| \otimes \mathbb{I}$;

[3] The factor of 2 arises solely from our specific definition of the local Hamiltonian problem in Chapter 3, and does not appear under a slightly more general formulation.

- collection of circuits $V^{t'}$, $t' \in [q+1]$, where circuit $V^{t'}$ only acts on at most t' qubits of the proof register and consists of at most p_3 gates from some universal gate set.

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

1. For $t \in [q]$: it applies the circuit V^t to registers A, B and qubits I from C, performs the measurement Π^t and adds outcome i_t to the set I;

2. It applies V^{q+1} followed by a measurement of Π^{out} of the first qubit and returns "accept" if the outcome was $|1\rangle$, and "reject" if the outcome was $|0\rangle$.

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

We make the final assumption in Definition 6.2.1 for notational simplicity, as querying the same proof index multiple times offers no advantage. The verifier can enforce this restriction by for example selecting the first index (under lexicographic ordering) from those not in the set I, whenever a measurement yields an element already in I.

6.2.2. Remark. We often write V_x to denote the verifier with the input x hard-coded. Moreover, in the case of multiple provers, an index i_t will denote a tuple (j,k) (of the qubit to be queried at step t), where j indicates the corresponding proof and k indicates the index of the qubit in this proof.

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

Some additional explanation is in order. First, we could have used a single symbol for the PVMs Π^t, as they all correspond to the same measurement. However, it is useful to label each PVM by the index indicating the step of the quantum PCP verification in which it is applied. This notation becomes particularly convenient when discussing scenarios such as: "PVM Π^1 returned index i_1, Π^2 returned outcome i_2", and so on. Second, one might wonder why Definition 6.2.1 is not defined entirely in terms of PVMs that absorb the circuits $V^{t'}$. As

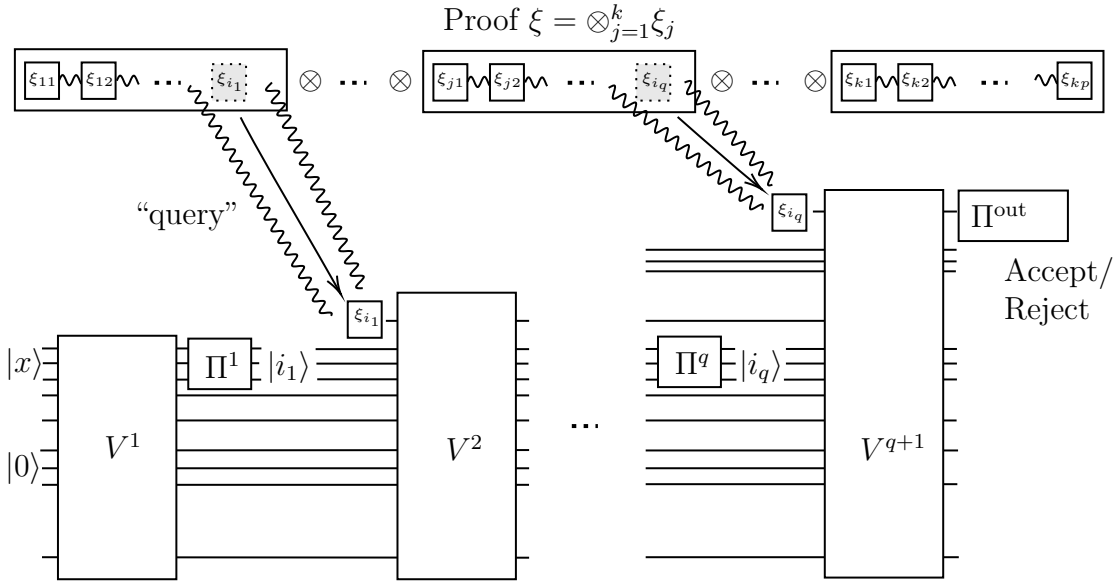


Figure 6.1: Visualisation of a k-prover quantum PCP which allows for q adaptive queries to the provided quantum proof through intermediate measurements, as per Definition 6.2.1. Note that an index measurement i_t denotes a tuple (j, k) where j indicates the corresponding proof and k the index of the qubit in this proof.

we shall see later, this split is necessary to ensure an efficient unitary decomposition whenever the PVMs are at most $\mathcal{O}(\log)$-local. Throughout this chapter, we assume an ordering on the tuples (j, l), where (j, l) refers to the lth qubit of the jth proof, and we use basis state notation, treating strings and their integer representations interchangeably. Since there are only $kp_2(n)$ possible settings, we need to allocate at most $\lceil \log(kp_2(n)) \rceil = \mathcal{O}(\log n)$ qubits to determine the next proof index to be queried, assuming $k, p_2(n) \in \text{poly}(n)$. This ensures that the PVMs remain $\mathcal{O}(\log n)$-local. In the constant-query setting (which is our focus), this property still holds in the non-adaptive case, even when a single PVM represents the parallel application of all individual-index PVMs. As a result, each PVM has at most polynomial circuit complexity, making them efficiently implementable.

We are now ready to define the complexity classes corresponding to our formulations of quantum PCP verifiers.

6.2.4. Definition (QPCP). Let $n \in \mathbb{N}$ be the input size, and let $x \in \{0,1\}^n$ be an input string. A promise problem $A = (A_{\text{YES}}, A_{\text{NO}})$ belongs to $\mathsf{QPCP}_{c,s}[k, q]$, if and only if there exist $c, s : \mathbb{N} \to [0, 1]$ with $c - s > 0$, polynomially bounded functions p_1, p_2, p_3 and a P-uniform family of (k, q, p_1, p_2, p_3)-QPCP verifiers $\mathcal{V} = \{V_n\}$ such that, on input $x \in \{0,1\}^n$:

- If $x \in A_{\text{YES}}$, then there exist quantum states $\xi_j \in \mathrm{D}\left((\mathbb{C}^2)^{\otimes p_2(n)}\right)$ for each

$j \in [k]$, such that V_n accepts $(x, \otimes_{j=1}^{k} \xi_j)$ with probability at least c;

- If $x \in A_{\text{NO}}$, then for all such states $\xi_j \in \mathrm{D}\left((\mathbb{C}^2)^{\otimes p_2(n)}\right)$, the verifier V_n accepts $(x, \otimes_{j=1}^{k} \xi_j)$ with probability at most s.

If $c = 2/3$ and $s = 1/3$, we write $\mathsf{QPCP}[k, q]$ for short, and $\mathsf{QPCP}[q]$ if additionally $k = 1$ holds.

6.2.5. Definition ($\mathsf{QPCP}_{\text{NA}}$). This class is defined in the same way as QPCP, but with the verifier restricted to be a non-adaptive $\mathsf{QPCP}_{\text{NA}}$ verifier, as specified in Definition 3.1a.

In the single-prover setting, weak error reduction via parallel repetition follows straightforwardly by adapting the argument from [AN02].

6.2.6. Lemma (Weak error reduction for the single-prover case). *Let $c - s = \Omega(1)$. Then for every $t \geq 2$, we have*

$$\mathsf{QPCP}_{c,s}[1, q] \subseteq \mathsf{QPCP}_{c',s'}[1, \mathcal{O}(qt)],$$

where $c' = 1 - 2^{-t}$ and $s' = 2^{-t}$.

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

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

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

$$R := \left\lceil \frac{2t \ln 2}{(c-s)^2} \right\rceil. \tag{6.1}$$

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

In the full work [BHW25], we also show that a form of strong error reduction is achievable for non-adaptive quantum PCPs with near-perfect completeness. The question of whether weak error reduction holds in the multi-prover case remains open, but we expect that it can be done using ideas from [HM13].

6.3 Local Hamiltonians from quantum PCPs

The core of this section is Theorem 6.3.6, which shows that, given a QPCP verification circuit (as defined in Definition 6.2.1), one can efficiently construct a local Hamiltonian such that the expectation value of a proof state corresponds to its acceptance probability under the verifier. Our construction follows some of the ideas presented in [Gri18], but extends them to handle the more general case of adaptive verifiers. As a result, the class of local Hamiltonians we obtain is slightly more general than those typically considered in the quantum PCP literature, requiring some additional steps to recover a standard form.

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

6.3.1. Lemma. *Let $n \in \mathbb{N}$ be the input size, k, p_1, p_2, p_3 be polynomially-bounded functions, $q = \mathcal{O}(1)$ and let V_x be a (k, q, p_1, p_2, p_3)-QPCP verifier as in Definition 6.2.1, with hardcoded input $x \in \{0,1\}^n$. Define $M^{t'}_{i_q,x} = \Pi^{t'}_{i_q} V^{t'}_x$ for all $i_q \in [kp_2(n)]$, $t' \in [q]$. The probability that the quantum PCP rejects a proof ξ, conditioned on taking the query path $(i_1, \ldots, i_q)$, is given by*

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

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

$$P_{x,(i_1,\ldots,i_q)} = M^{1\dagger}_{i_1,x}\ldots M^{q\dagger}_{i_q,x} V_x^{q+1\dagger}\Pi_0^{\text{out}} V_x^{q+1} M^q_{i_q,x}\ldots M^1_{i_1,x}. \tag{6.2}$$

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

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

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

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

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

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

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

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

$$\begin{aligned}\Pr[V_x \text{ rejects } \xi \mid (i_1,\ldots,i_q)] &= \text{tr}\left[\Pi_0^{\text{out}} V_x^q \rho^q V_x^{q\dagger}\right] \\ &= \frac{\text{tr}\left[\Pi_0 V_x^q M^q_{i_q,x}\ldots M^1_{i_1,x}\rho^0 M^{1\dagger}_{i_1,x}\ldots M^{q\dagger}_{i_q,x} V_x^{q\dagger}\right]}{\Pr[(i_1,\ldots,i_q)]}.\end{aligned}$$

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

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

with

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

□

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

6.3.2. Lemma. *Let A be a Hermitian operator acting on an n-qubit Hilbert space, where the state is given by a variable q-qubit input ρ tensored with a fixed $(n-q)$-qubit state ρ_{fixed}. Then*

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

for some q-qubit Hermitian operator $B = B(\rho_{\text{fixed}})$ that depends on ρ_{fixed}. Moreover, if $\rho_{\text{fixed}} = |\psi\rangle\langle\psi|$ for some pure state $|\psi\rangle$, then the (α,α')-entry of B in any basis $\{|\alpha\rangle\}$ is given by $b_{\alpha,\alpha'} = \langle\alpha| \langle\psi| A |\alpha'\rangle |\psi\rangle$.

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

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

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

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

where the operator B, given by

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

is indeed q-local.

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

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

□

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

6.3.3. Lemma (Hamiltonians from general quantum PCPs). *Let $n \in \mathbb{N}$ be the input size, k, p_1, p_2, p_3 be polynomially-bounded functions, $q = \mathcal{O}(1)$ and let V_x be a (k, q, p_1, p_2, p_3)-QPCP verifier as in Definition 6.2.1 with hardcoded input $x \in \{0,1\}^n$. Then there exists a Hamiltonian H_x acting on $kp_2(n)$-qubits, consisting of q-local positive semidefinite terms, such that for all ξ, we have*

$$\Pr[V_x \text{ accepts } \xi] = 1 - \operatorname{tr}[H_x\xi]. \tag{6.3}$$

Proof:
Let $\Omega = \binom{[kp_2(n)]}{q}$ be the set of all unordered subsets of size q drawn from $[kp_2(n)]$. By the definition of conditional probability, we have that

$$\Pr[V_x \text{ accesses qubits } (i_1,\ldots,i_q) \text{ from } |\xi\rangle \text{ and rejects }]$$

is given by

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

where $\sigma_{i_1,\ldots,i_q} = \operatorname{tr}_{\bar{C}_{i_1,\ldots,i_q}}\left[\xi\otimes|0\rangle\langle 0|^{\otimes n+p_2(n)}\right]$ with $\bar{C}_{i_1,\ldots,i_q}$ denoting all qubits of ξ except for those with indices $i_1,\ldots,i_q$. Hence, we can write the probability that V_x rejects ξ as

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

For all $\{i_1, \ldots, i_q\} \in \Omega$, we define the $2^q \times 2^q$ matrix $H_{x,i_1,\ldots,i_q}$, where each entry $\langle\alpha| H_{x,i_1,\ldots,i_q} |\beta\rangle$ is given by

$$\sum_{(i_1,\ldots,i_q)\in \mathrm{Perm}(i_1,\ldots,i_q)} \left(\langle 0|^{\otimes p_1(n)+n} \otimes \langle\alpha|\right) P_{x,(i_1,\ldots,i_q)} \left(|0\rangle^{\otimes p_1(n)+n} \otimes |\beta\rangle\right), \tag{6.4}$$

where $\alpha, \beta \in \{0,1\}^q$. By Lemma 6.3.2, we have that for any q-local density matrix ρ we have that the expression

$$\mathrm{tr}\big[H_{x,i_1,\ldots,i_q}\rho\big] = \mathrm{tr}\left[\sum_{(i_1,\ldots,i_q)\in \mathrm{Perm}(i_1,\ldots,i_q)} P_{x,(i_1,\ldots,i_q)} |0\rangle \langle 0|^{\otimes n+p_2(n)} \otimes \rho\right] \tag{6.5}$$

holds. Moreover, since Eq. (6.5) is the sum of all probabilities that a query path $(i_1, \ldots, i_1)$ is taken and the proof is rejected, taken over all possible permutations of the indices, we must have that $H_{x,i_1,\ldots,i_q}$ is positive semidefinite. Now consider the q-local Hamiltonian H defined as

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

We have that

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

since by the linearity of the trace

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

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

6.3.1 Learning the Hamiltonian

We have shown that the acceptance probability of a QPCP verifier on a given proof is equivalent to the expectation value of a corresponding Hamiltonian. What remains is to show how to efficiently approximate the entries of each local term. Before stating the final protocol, we argue that each local term can indeed be learned to arbitrary (inverse polynomial) precision. For this, we employ the Hadamard test, originally introduced in [AJL06].

We require a simple generalisation of the standard Hadamard test, as we need to evaluate inner products between two potentially different input states. A proof of this generalisation can be found in Chapter 2 of [Gri18].

6.3.4. Lemma (Hadamard test [AJL06]). *Let $|\psi\rangle, |\phi\rangle \in \mathbb{C}^{2^n}$ be quantum states prepared by unitaries U_ψ, U_ϕ, i.e., $|\psi\rangle = U_\psi |0^n\rangle$ and $|\phi\rangle = U_\phi |0^n\rangle$. Let $W \in \mathbb{U}(2^n)$ be a unitary operator, to which we have controlled access. Then there exists a polynomial-time quantum algorithm that outputs an estimate $\hat{z}$ such that*

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

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

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

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

Since the Hadamard test applies only to unitary operators, we will rely on the following simple lemma, which shows that every (local) projector onto a basis state can be expressed as a linear combination of unitaries with small circuit depth.

6.3.5. Lemma. *Let $\Pi_z = |z\rangle\langle z|$, where $z \in \{0,1\}^k$ is a computational basis state. Then Π_z can be written as*

$$\Pi_z = \frac{1}{2^k} \sum_{j \in \{0,1\}^k} a_j U_j,$$

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

Proof:
For a single qubit, we have $|0\rangle\langle 0| = \frac{1}{2}(\mathbb{I} + Z)$ and $|1\rangle\langle 1| = \frac{1}{2}(\mathbb{I} - Z)$. Hence, for $z_i \in \{0,1\}$,

$$|z_i\rangle\langle z_i| = \frac{1}{2}\left(Z + (1 - 2z_i)\mathbb{I}\right) = \frac{1}{2}\left(\mathbb{I} + (-1)^{z_i} Z\right).$$

Therefore, expanding the tensor product of the full projector onto the computational basis state $|z\rangle$ yields

$$\Pi_z = \bigotimes_{i=1}^{k} |z_i\rangle\langle z_i| = \frac{1}{2^k} \bigotimes_{i=1}^{k} (\mathbb{I} + (-1)^{z_i} Z) = \frac{1}{2^k} \sum_{j \in \{0,1\}^k} a_j U_j,$$

where each $U_j \in \{\mathbb{I}, Z\}^{\otimes k}$ is of the form $\bigotimes_{i=1}^{k} V_i$ with $V_i \in \{\mathbb{I}, Z\}$, and $a_j \in \{-1, 1\}$ is the product of signs $(-1)^{z_i}$ chosen according to the positions where Z appears in U_j. □

Before we move to prove the existence of the reduction, we need to define some parameters. The operators $H_{x,i_1,\ldots,i_q}$ (whose entries are defined in Eq. (6.4)) are composed of $q+1$ unitaries $\{V^t\}$, a total of q of $\mathcal{O}(\log n)$-local PVM elements (see Section 6.2), and a single 1-local PVM element (which is Π_{out}). If we use $\log(kp_2(n)) + 1$ qubits for each Π^t and decompose each PVM element in $H_{x,i_1,\ldots,i_q}$ into unitaries, as per Lemma 6.3.5, we can write as entry of the sum (that goes over all possible permutations of the indices) that defines $\langle\alpha| H_{x,i_1,\ldots,i_q} |\beta\rangle$ in Eq. (6.4) as

$$\frac{1}{\Gamma} a_j \left(\langle 0|^{\otimes p_1(n)+n} \otimes \langle\alpha| \right) U_{j,x} \left(|0\rangle^{\otimes p_1(n)+n} \otimes |\beta\rangle \right),$$

with

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

where the unitaries $U_{j,x}^l$ are composed of a polynomial number of elementary gates and $a_j \in \{-1, +1\}$. The range of index l can be seen from inspecting Eq. (6.2) of Lemma 6.3.1: each $P_{x,(i_1,\ldots,i_q)}$, which makes up a $H_{x,(i_1,\ldots,i_q)}$, consists of $q+1$ V_t's and their conjugate transposes, two unitaries for each of the first q PVM elements coming from its decomposition and a final single one for the final outcome measurement (which is "sandwiched" in the middle of Eq. (6.2)). Hence, we have a total of $2(q+1) + 2q + 1 = 4q + 3$ unitaries in the product. The total number of unitaries in the linear combination for each $(i_1, \ldots, i_q)$ is given by

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

We define

$$z_{(i_1,\ldots,i_q)}^{\alpha,\beta,j} \coloneqq \left(\langle 0|^{\otimes p_1(n)+n} \otimes \langle\alpha| \right) U_{j,x} \left(|0\rangle^{\otimes p_1(n)+n} \otimes |\beta\rangle \right), \tag{6.6}$$

and

$$h_{i_1,\ldots,i_q}^{\alpha,\beta} = \sum_{(i_1,\ldots,i_q) \in \text{Perm}(i_1,\ldots,i_q)} \sum_{j \in [\Gamma]} a_j z_{(i_1,\ldots,i_q)}^{\alpha,\beta,j}. \tag{6.7}$$

such that

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

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

6.3.6. THEOREM. *Let n be the input size, k, p_1, p_2, p_3 be polynomially-bounded functions, $q = \mathcal{O}(1)$ and let V_x be a (k, q, p_1, p_2, p_3)-QPCP verifier as in Definition 6.2.1 with hardcoded input $x \in \{0,1\}^n$. Then for all $\epsilon > 0$, $\delta > 0$, there exists a quantum reduction to a q-local Hamiltonian $\hat{H}_x = \sum_{i\in[m]} \hat{H}_{x,i}$, with $m = \mathrm{poly}(n)$, such that with probability at least $1-\delta$, each $\hat{H}_{x,i}$ is positive semidefinite and $\left\|\hat{H}_x\right\| \leq 1$, and for any proof ξ,*

$$\left|\Pr[V_x \text{ accepts } \xi] - \left(1 - \mathrm{tr}\left[\hat{H}_x \xi\right]\right)\right| \leq \epsilon. \tag{6.8}$$

The reduction runs in time $\mathrm{poly}(n, 1/\epsilon, \log(1/\delta))$.

Proof:
The reduction is specified in Algorithm 6.3.1 below. We proceed to show its correctness by arguing it has the required precision, success probability, and time complexity.

Precision. Step 1*a* of Algorithm 6.3.1 produces estimates $\tilde{z}^{\alpha,\beta,j}_{(i_1,\ldots,i_q)}$ for the parameters $z^{\alpha,\beta,j}_{(i_1,\ldots,i_q)}$ using Lemma 6.3.4. We have that

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

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

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

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

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

which follows from the bound on the operator norm by the max-norm. Now suppose that $\tilde{H}_{x,i_1,\ldots,i_q}$ is not positive semidefinite. Since $H_{x,i_1,\ldots,i_q}$ is positive semidefinite, we have that $\lambda_0(\tilde{H}_{x,(i_1,\ldots,i_q)}) \geq -2^{q+1}q!\epsilon'$, so we have that adding the

identity term can only double the error, making it at most $2^{q+2}q!\epsilon'$. By another triangle inequality

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

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

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

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

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

as desired.

Success probability. We must count the number of times we run the Hadamard test of Lemma 6.3.4, each of which succeeds with probability at least $1-\delta'$. Recall that Ω denotes the set of all possible index tuples (where order does not matter), given by $\Omega = \binom{[kp_2(n)]}{q}$ for proofs of length p_2. We perform the Hadamard test a total of $|\Omega| \cdot q! \cdot 4^q \cdot \Gamma$ times (see the number of iterations in Algorithm 6.3.1), and so

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

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

Time complexity. By definition of $\mathsf{QPCP}[q]$, we have that V has gate complexity $\text{poly}(n)$. Using Lemma 6.3.5, we have that V, U_ϕ, and U_ψ always have polynomially bounded gate complexities. Filling in our choice of δ' and ϵ', we

have that the total number of (controlled) applications of V, U_ϕ, and U_ψ can be upper bounded as

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

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

Algorithm 6.3.1: Quantum reduction from $\mathsf{QPCP}[k,q]$ verification to a local Hamiltonian

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

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

Algorithm:

1. For $\{i_1,\ldots,i_q\}\in\Omega$
   (a) For $(i_1,\ldots,i_q)\in\text{Perm}(i_1,\ldots,i_q)$
      i. For $\alpha\in\{0,1\}^q$, $\beta\in\{0,1\}^q$:
         - For $j\in[\Gamma]$ compute $a_j,U_{j,x}$ and estimate $z^{\alpha,\beta,j}_{(i_1,\ldots,i_q)}$ as $\tilde{z}^{\alpha,\beta,j}_{(i_1,\ldots,i_q)}$ using Lemma 6.3.4 for both the real and imaginary part with $V=U_{j,x}$, $U_\theta=\hat{V}(\otimes_{i=1}^q(X_i)^{\alpha_i}\otimes\mathbb{I})$ and $U_\phi=\hat{V}(\otimes_{i=1}^q(X_i)^{\beta_i}\otimes\mathbb{I})$, ϵ' and δ'.

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

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

   (c) Compute $\lambda_0(\tilde{H}_{x,i_1,\ldots,i_q})$. If $\lambda_0(\tilde{H}_{x,i_1,\ldots,i_q})<0$, let $\tilde{H}_{x,i_1,\ldots,i_q})\leftarrow\tilde{H}_{x,i_1,\ldots,i_q})-\lambda_0(\tilde{H}_{x,i_1,\ldots,i_q})\mathbb{I}$, else continue.

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

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

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

One can use a simple trick that exploits the freedom we have in the reduction to set the precision. Since every local term is bounded in the operator norm, we must also have that for each of these terms the matrix entries are bounded by 1. Hence, we can convert a bound in terms of approximation in entry-wise error to one in having a least a certain number of bits from a certain bit-wise representation of the value being correct. The advantage of the latter is that it allows us to make the reduction deterministic in the sense that if it succeeds, it always produces *exactly* the same Hamiltonian. This can alternatively be viewed as having a "rounding scheme" which, with high probability, always rounds to the same operator in operator space.[4] Though not strictly needed, this allows one to always reason about the same Hamiltonian when the reduction is used as a subroutine.

6.3.7. Corollary. *For any $\epsilon, \delta > 0$, under the same setup as in Theorem 6.3.6, there exists a quantum reduction which produces a fixed Hamiltonian $\tilde{H}_x$ with probability $1-\delta$ which satisfies Eq.* (6.8) *and runs in time* $\mathrm{poly}(n, 1/\epsilon, \log(1/\delta))$.

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

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

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

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

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

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

To achieve

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

[4] This idea will become important in Chapter 7.

it suffices to set

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

Since the Hadamard test can learn η bits of precision in $\mathcal{O}(2^\eta)$ time, we have that the time complexity of the reduction in Algorithm 6.3.1 still runs in time $\mathrm{poly}(n, 1/\epsilon, \log(1/\delta))$ as $\eta = \mathcal{O}(\log(\mathrm{poly}(n, 1/\epsilon, \log(1/\delta))))$. □

We proceed to show that any Hamiltonian of the form H_x in Eq. (6.3) can be transformed into the form given in Definition 3.1.2. Specifically, we aim to convert a local Hamiltonian with positive semidefinite terms, global operator norm at most 1, and a constant promise gap into another Hamiltonian such that the promise gap becomes $\Omega(m)$, where m is the number of terms in the new Hamiltonian. Each term H_i of this new Hamiltonian remains local and now satisfies $0 \preceq H_i \preceq 1$, such that the global operator norm satisfies $0 \preceq H \preceq m$. In this formulation, at least a constant fraction of the m terms contribute to the energy in the NO-case, which intuitively implies that it suffices to check only a constant number of terms to determine whether the energy is low or high (see Section 6.3.2). This transformation is possible at the cost of increasing the locality by a factor of two and reducing the promise gap by a constant factor when the original Hamiltonian has constant locality.[5]

6.3.8. Lemma (Hamiltonian local smoothing lemma). *Let H be a q-local Hamiltonian on n qubits, such that $H = \sum_{i\in[m]} H_i$, where each H_i is positive semidefinite and $0 \preceq H \preceq 1$. Suppose further that the ground state energy satisfies either $\lambda_0(H) \leq a$ or $\lambda_0(H) \geq b$, with $b, a \in [0,1]$ such that $b - a > 0$. Then there exists a polynomial-time algorithm which, from the descriptions of H constructs a $2q$-local Hamiltonian H' on n qubits such that*

$$H' = \sum_{i\in[m]} H'_i, \quad \text{with each} \quad 0 \preceq H'_i \preceq 1.$$

Moreover, it holds that:

- *if $\lambda_0(H) \leq a$, then $\lambda_0(H') \leq \alpha$;*
- *if $\lambda_0(H) \geq b$, then $\lambda_0(H') \geq \beta$;*

where $\alpha = \left(\frac{a}{2^{q+3}}\right) m$ and $\beta = \left(\frac{b}{2^{q+3}}\right) m$.

[5] We believe such a result may already be known in the literature, but as we could not find a reference, we include a proof for completeness.

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

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

The variational principle tells us that

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

Combining the two, this implies

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

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

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

We have

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

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

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

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

Consider the following procedure:

1. Initialise $L' \coloneqq L$.

2. For each $j \in U$, check if $\alpha_j > 1/m$. If this is not the case, continue the loop (or exit after the last j). If this is the case, set $t_j = \lfloor 2m\alpha_j - 1 \rfloor$. For the first t_j indices $i \in L'$, define

$$Q_i := \hat{H}_i + \frac{1}{2m}\hat{H}_j.$$

These Q_i are at most $2q$-local and have operator norm at most $1/m$. Remove these t_j indices i from L'. Define the q-local leftover term

$$Q_j := \left(1 - \frac{t_j}{2m}\right)\hat{H}_j.$$

3. For each remaining $j \in U$ that was not used in any redistribution, simply set $Q_j := \hat{H}_j$.

4. Let $H'_i := mQ_i$ for each term. Return $H' := \sum H'_i$.

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

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

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

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

□

6.3.2 Kitaev's energy estimation protocol

Now that we have shown how a general quantum PCP can be transformed into a local Hamiltonian with a constant promise gap and a desirable form, we turn to the task of demonstrating that the corresponding promise problem is again contained in $\mathsf{QPCP}[\mathcal{O}(1)]$. To this end, we rely on Kitaev's QMA-protocol for the local Hamiltonian problem [KSV02], with a minor modification that allows the verifier to sample Hamiltonian terms according to a specified probability distribution (Protocol 6.3.1). We present the energy verification protocol in the pure-state setting, noting that the mixed-state case follows directly by interpreting it as a

convex combination of acceptance probabilities of pure states.

Protocol 6.3.1: Kitaev's energy estimation protocol

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

Protocol:

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

$$W_i |\lambda_{i,j}\rangle |b\rangle = |\lambda_{i,j}\rangle \left(\sqrt{\lambda_{i,j}} |b\rangle + \sqrt{1-\lambda_{i,j}} |b\oplus 1\rangle\right). \tag{6.9}$$

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

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

$$1 - \langle\psi| H |\psi\rangle. \tag{6.10}$$

Proof:
This follows from a simple generalisation of Kitaev's original QMA verification protocol, which can be found in Chapter 14 of [KSV02]. Let us now show the correctness of Protocol 6.3.1. For any i, let $|\psi\rangle = \sum_j \alpha_{i,j} |\lambda_{i,j}\rangle$ be the decomposition of $|\psi\rangle$ in the eigenbasis of H_i. The probability that this protocol accepts,

conditioned on picking term i, is given by

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

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

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

□

Kitaev's energy estimation protocol (Protocol 6.3.1) can be viewed as a $(1, q)$-$\mathsf{QPCP}_{\text{NA}}$ verifier, where the completeness and soundness bounds correspond to one minus the promised upper and lower bounds on the ground state energy in the YES- and NO-cases, respectively. When $q = \mathcal{O}(\log n)$, each W_i acts non-trivially only on $q + 1$ qubits and can thus be implemented efficiently. By applying Lemma 6.3.8 in combination with weak error reduction (Lemma 6.2.6), we can correctly decide any local Hamiltonian problem with positive semidefinite terms and a constant promise gap using Kitaev's protocol, taking $p_i = 1/m$ for each $i \in [m]$.

6.3.10. COROLLARY. *For any $q \in \mathcal{O}(1)$ and $\delta > 0$ constant, we have that the local Hamiltonian problem $\mathrm{LH}(q, a, b)$ with $b - a \geq \delta m$, where m is the number of local terms, is contained in $\mathsf{QPCP}_{\text{NA}}[q']$ for some $q' \in \mathcal{O}(1)$.*

6.4 Applications

This section presents several applications of the ideas developed earlier in the chapter, each derived (with varying degrees of overhead) by using the reduction as a subroutine. The section is divided into four subsections, each of which can be read more or less independently.

6.4.1 Reduction to the average particle energy formulation

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

6.4.1. Lemma ([Tro12]). *Consider a finite sequence $\{X_k\}$ of independent, random, Hermitian matrices with dimension d, and let $\{A_k\}$ be a sequence of fixed Hermitian matrices. Assume that each random matrix satisfies*

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

Then for any $t \geq 0$,

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

where

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

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

6.4.2. Proposition. *Consider an instance of $\mathrm{LH}(q, a, b)$ for an n-qubit Hamiltonian $H = \sum_{i\in[m]} H_i$, where $0 \preceq H_i \preceq 1$, and $b - a \geq \gamma m$ for some constant $\gamma > 0$. Then for any $0 < \delta \leq 1$ with $\delta = \Omega(2^{-n})$, there exists a randomized polynomial-time reduction that, with probability at least $1 - \delta$, produces an instance of $\mathrm{LH}(2q, c, d)$ with Hamiltonian $G = \sum_{j\in[l]} G_j$, where $0 \preceq G_j \preceq 1$ and $l = \mathcal{O}(n)$, with $c - d \geq \gamma' m$ for some constant $\gamma' > 0$.*

Proof:
Without loss of generality, we can consider the rescaled instance of $\mathrm{LH}(q, a, b)$ as

$$H = \frac{1}{m}\sum_{i\in[m]} H_i$$

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

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

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

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

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

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

By applying Lemma 6.4.1, we get

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

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

- if $\lambda_0(H) \leq a$, then $\lambda_0(G) \leq a' \coloneqq a + \epsilon$;
- if $\lambda_0(H) \geq b$, then $\lambda_0(G) \geq b' \coloneqq b - \epsilon$,

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

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

which implies a condition on l of

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

For $\gamma = \Omega(1)$ and $\delta = \Omega(2^{-n})$, this yields $l = \Theta(n)$. Now, we must ensure that each G_k satisfies $0 \preceq G_k \preceq 1$. The condition $0 \preceq G_k$ is trivially satisfied, but $G_k \preceq 1$ might not hold if we sample the same term H_i multiple times. To resolve this, we apply the deterministic transformation from Lemma 6.3.8, which maintains the constant relative promise gap $\gamma' = \Omega(1)$ while increasing the locality by a factor of two. Finally, we rescale by a factor l to obtain the required form of Definition 3.1.2. □

A combination of Theorem 6.3.6, Lemma 6.3.8, and Proposition 6.4.2 implies the following corollary:

6.4.3. Corollary. *For any $q = \mathcal{O}(1)$, there exists a $q' = \mathcal{O}(q)$ and a constant $\gamma > 0$, such that $\mathrm{LH}(q', a, b)$, $b - a \geq \gamma n$, with the restriction that the Hamiltonian has only $m = \Theta(n)$ terms, each of which is positive semidefinite, is $\mathsf{QPCP}[q]$-hard under quantum reductions.*

6.4.2 Proof checking versus local Hamiltonian formulations of quantum PCP

The goal of this subsection is to prove Theorem 6.4.4. We will make use of the fact that any QCMA verifier is capable of performing the reduction described in Algorithm 6.3.1, up to polynomial precision. This observation allows us to show that a QCMA upper bound on the q-local Hamiltonian problem with constant promise gap implies a QCMA upper bound on a quantum PCP system with the same locality.

Protocol 6.4.1: QCMA protocol for QPCP$[q]$ assuming that $\mathrm{LH}(2q, a, b) \in$ QCMA for any $b - a \in \Omega(1)$.

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

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

Protocol:

1. The prover sends the witness y.
2. The verifier performs the **variation** of Algorithm 6.3.1 with precision η and maximum error probability δ to obtain a q-local Hamiltonian $\tilde{H}_x = \sum_{i\in[m]} \tilde{H}_{x,i}$ up to η bits of precision.
3. The verifier applies the transformation of Lemma 6.3.8 to obtain a $2q$-local Hamiltonian $\hat{H}_x = \sum_{i\in[m]} \hat{H}_{x,i}$.
4. It accepts if and only if the QCMA protocol, having completeness $\sqrt{\frac{2}{3}}$ and soundness $1 - \sqrt{\frac{2}{3}}$, with witness y for $(\hat{H}, a, b)$ accepts.

6.4.4. Theorem (LH versus proof verification). *Let $q \in \mathbb{Z}_+$ be constant. If for any b, a such that $b - a = \Omega(1)$ we have that* $\mathrm{LH}(2q, a, b)$ *is in* QCMA*, then*

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

Proof:
Let $A = (A_{\text{YES}}, A_{\text{NO}})$ be any promise problem in $\mathsf{QPCP}[q]$, let x be an input, and let V_x denote the (q)-QPCP verifier with x hardcoded into it. Let $\hat{H}_x$ be the corresponding local Hamiltonian obtained via the reductions in Step 2 and Step 3, conditioned on Step 2 succeeding. By Theorem 6.3.7, we have that in this case the same Hamiltonian is always produced with probability $\geq 1 - \delta = \sqrt{2/3}$, so the prover can provide the proof without knowing the outcome of the reduction. Moreover, we note that for our choice of parameters, both Step 2 and Step 3 can be performed in quantum polynomial-time.

By assumption, the $2q$-local Hamiltonian problem with constant promise gap is in QCMA. Hence, there exists a QCMA verifier Q such that:

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

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

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

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

since QCMA allows for strong error reduction. Now consider Protocol 6.4.1. Let Q' be the QCMA verifier that first performs reductions in Steps 2 and 3 to obtain $\hat{H}_x$ to η bits of precision, and then runs $Q(\hat{H}_x, a, b, |y\rangle)$. Since Step 2 succeeds with probability $1 - \delta = \sqrt{\frac{2}{3}}$, and Step 3 is deterministic, we have:

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

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

6.4.3 Adaptive versus non-adaptive quantum PCPs

Since adaptive quantum PCPs generalise non-adaptive quantum PCPs, we have that the inclusion $\mathsf{QPCP}_{\text{A}}[q] \supseteq \mathsf{QPCP}_{\text{NA}}[q]$ is immediate. In this subsection, we will prove that non-adaptive QPCPs can also simulate adaptive QPCPs with only constant overhead, by using our Lemma 6.3.8 and weak error reduction lemma (Lemma 6.2.6) for non-adaptive QPCPs.

Protocol 6.4.2: Non-adaptive simulation of an adaptive quantum PCP

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

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

Protocol:

1. The prover sends a quantum state $|\psi\rangle$.
2. The verifier runs the **variation** of Algorithm 6.3.1, with precision η and maximum error probability δ, to obtain a q-local Hamiltonian $\tilde{H}_x$.
3. The verifier applies the transformation of Lemma 6.3.8 to obtain a $2q$-local Hamiltonian $\hat{H}_x = \sum_i \hat{H}_{x,i}$.
4. The verifier runs Protocol 6.3.1 R times for $\hat{Q}_x = \sum_i \frac{1}{m}\hat{H}_{x,i}$, and accepts if and only if at least a $\frac{(c-s)}{2^{q+4}}$-fraction of the outcomes equal $|1\rangle$.

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

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

with

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

Proof:
We verify the correctness of Protocol 6.4.2, which defines a (q')-$\mathsf{QPCP}_{\mathrm{NA}}$ verifier for some q', which we will show to be $\mathcal{O}(q)$. Let $A = (A_{\mathrm{YES}}, A_{\mathrm{NO}})$ be a promise problem in $\mathsf{QPCP}_{\mathrm{A}}[q]$ for an arbitrary constant q, and let $x \in \{0,1\}^n$ be an input.

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

- If $x \in A_{\mathrm{YES}}$, then $\lambda_0(\tilde{H}_x) \leq s + \epsilon$,

- If $x \in A_{\text{NO}}$, then $\lambda_0(\tilde{H}_x) \geq c - \epsilon$,

with probability at least $1-\delta$. After applying the transformation from Lemma 6.3.8 to obtain $\hat{H}_x$ from $\tilde{H}_x$, we get:

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

Applying Kitaev's energy estimation protocol (Protocol 6.3.1) to $\hat{Q}_x = \sum_i \frac{1}{m} \hat{H}_{x,i}$ yields a $(2q)$-$\mathsf{QPCP}_{\text{NA}}$ verifier with a promise gap of at least

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

since q was constant and $c - s = \Omega(1)$. Since Step 4 performs R parallel runs of a $(2q)$-$\mathsf{QPCP}_{\text{NA}}$ verifier, accepting if and only if a $\frac{c-s}{2^{q+4}}$-fraction of the verifiers accept, it forms an (Rq)-$\mathsf{QPCP}_{\text{NA}}$ verifier with amplified completeness and soundness. By Lemma 6.2.6, and our choice of R (see Eq. (6.1) in the proof of Lemma 6.2.6), it follows that, conditioned on Step 2 succeeding, the acceptance (resp. rejection) probability in Step 4 is at least $\sqrt{2/3}$ in the YES-case (resp. NO-case). Moreover, we have that

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

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

We conclude this subsection with the following observation.

6.4.6. Remark. The proof of Theorem 6.4.5 actually demonstrates that using a uniform distribution (over a subset of all proof qubits) to decide which proof qubits to access suffices. Indeed, Kitaev's energy estimation protocol can be applied with $p_i = 1/m$ for all $i \in [m]$, where m is the number of terms in the Hamiltonian. Since we still apply weak error reduction (Lemma 6.2.6) to boost the completeness and soundness parameters back to their original values, the resulting distribution over queries becomes uniform over a subset of all possible index combinations. This is because parallel repetition restricts us to combinations where each supposed copy is used only once. This resolves an open question posed in [AALV09], and shows that the definition of a quantum PCP given in [Gri18] is in fact fully general.

6.4.4 A quantum PCP for QMA(2) implies QMA(2) = QMA

In this section, we prove that the existence of a multi-prover quantum PCP for QMA(2) implies QMA(2) = QMA. The argument builds on the results of [CS12], which show that the 2-separable local Hamiltonian problem lies in QMA. As we require a slight generalisation of their result, that is, from 2-separable to k-separable Hamiltonians, we include a complete proof for completeness. We begin by formally defining the k-separable local Hamiltonian problem.

6.4.7. Definition (k-separable q-local Hamiltonian problem, LH(k, q, a, b)).

Input: A classical description of a collection of q-local Hermitian operators $\{H_i\}_{i\in[m]}$, with $\|H_i\| \leq 1$ for all $i \in [m]$, which define an n-qubit k-local Hamiltonian $H = \sum_{i=1}^{m} H_i$, where $m = \text{poly}(n)$; a number k; and two efficiently computable real numbers a, b such that $b > a$.

Promise: One of the following two holds:

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

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

Output: YES in case (i); output NO in case (ii).

As part of the protocol we will use the consistency of local density matrices problem (CLDM(q, α, β)) and the protocol which puts it in QMA for $q = \mathcal{O}(\log n)$ and $\beta - \alpha \geq 1/\text{poly}(n)$, which can be found in Section 3.3.3.

We will also make use of the following basic property of QMA, which states that the conjunction (i.e., logical AND) of multiple promise problems in QMA can be decided by a single QMA verifier.

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

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

is in QMA.

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

1. If $x_i \in A_{i,\text{yes}}$, then there exists a proof $|\psi_i\rangle$ such that $\Pr[U_n^i \text{ accepts } |\psi_i\rangle] \geq 1 - 2^{-p_2(n)}$.

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

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

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

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

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

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

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

Next we show that the k-separable q-local Hamiltonian problem (Definition Definition 6.4.7) is in QMA. This relies directly on the QMA containment of the CLDM problem. We first state a QMA protocol.

Protocol 6.4.3: QMA protocol for the k-separable q-local Hamiltonian problem

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

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

1. The prover sends a quantum witness γ and a classical witness $\{\rho_i^j\}$.
2. The verifier performs the following three checks:
   - **Check 1:** It checks that each reduced density matrix ρ_i^j is positive semidefinite and has trace one.
   - **Check 2:** It checks that $\mathrm{tr}\left[H \otimes_{j\in[k]} \rho_i^j\right] \leq a'$.
   - **Check 3:** It splits up the indices of the quantum witness γ in k disjoint sets C_j of equal size, $j \in [k]$. It runs k parallel executions of the QMA protocol given by Protocol 3.3.1, each with completeness $1-\delta$ and soundness δ, on the k instances of the CLDM(q, α, β) with respective input $\{\rho_i^j\}$ and proof $\mathrm{tr}_{\bar{C}_j}[\gamma]$. The check is passed if and only if all CLDM(q, α, β) verifications accept.
3. The verifier accepts if and only if all three checks are passed.

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

Proof:
This follows the same ideas as in [CS12], but now for a k- instead of 2-separable state.

First, we observe that Check 3 defines the logical AND of multiple promise problems in QMA, so by the proof of Lemma 6.4.8, it follows that for our choice of δ, the overall procedure defined by Check 3 has completeness at least $2/3$ and soundness at most $1/3$.

Completeness. By the promise, there exists a state $\xi = \otimes_{j\in[k]}\xi^j$ such that $\mathrm{tr}[H\xi] \leq a$. The prover sends as a quantum witness $\gamma = \bigotimes_{j\in[k]} \gamma^j$, where each γ^j is a witness for the CLDM instance corresponding to ξ^j. For the classical witness, the prover sends descriptions of $\{\tilde{\xi}_i^j\}$ of the q-local reduced density matrices ξ_i^j of ξ, each specified up to trace distance $\epsilon = 2^{-p(n)}$ for some large polynomial $p(n)$,

such that the second and third checks are satisfied (see below). This is possible, as the entries of each density matrix can be described using a polynomial number of bits. In this case, the protocol proceeds as follows:

- The first check of Protocol 6.4.3 is passed directly.
- For the second check, we observe that

$$\sum_{i\in[m]} \mathrm{tr}\Big[H_i \otimes_{j\in[k]} \tilde{\xi}_i^j\Big] \leq a + mk\epsilon \leq a',$$

  provided ϵ is chosen sufficiently small, which means that it also passes.
- For Check 3, by Lemma 6.4.8, this check also succeeds with probability at least 2/3. Hence, the overall acceptance probability is at least 2/3.

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

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

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

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

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

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

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

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

which implies that there must exist a i, j pair such that $\beta \coloneqq \frac{b-a}{2qm} \leq \left\|\xi_i^j - \rho_i^j\right\|_1$, which satisfies the promise of a NO-instance for CLDM(q, α, β). As we have already argued, Step 3 has a success probability of at least $2/3$ for detecting a single NO-instance, by the proof of Lemma 6.4.8. Hence, we have that the overall acceptance probability is at most $1/3$. □

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

Protocol 6.4.4: QMA protocol for $\mathsf{QPCP}[k, q]$.

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

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

Protocol:

1. The prover sends the witness ξ.
2. The verifier runs the **variation** of Algorithm 6.3.1, with precision η and maximum error probability δ, to obtain a q-local Hamiltonian $\tilde{H}_x$.
3. The verifier applies the transformation of Lemma 6.3.8 to obtain a $2q$-local Hamiltonian $\hat{H}_x = \sum_{i\in[m]} \hat{H}_{x,i}$.
4. The verifier runs Protocol 6.4.3 for Hamiltonian $\hat{H}_x$ with completeness $1 - \delta'$ and soundness δ', with $a = \frac{s-\epsilon}{2^{q+3}}m$ and $b = \frac{c-\epsilon}{2^{q+3}}m$, and accepts if and only if Protocol 6.4.3 accepts.

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

Proof:
Suppose that such a k' and q indeed exist. Let $A = (A_{\text{YES}}, A_{\text{NO}})$ be any problem in $\mathsf{QPCP}[k', q]$, with verifier V and input x. We have already verified in the proof of Theorem 6.4.5 that Step 2 of Protocol 6.4.4 can be made to with probability at least $1 - \delta' \geq \sqrt{2/3}$, producing a fixed Hamiltonian when succeeding, and Step 4 can be made to succeed with probability at least $1 - \delta \geq \sqrt{2/3}$ by standard

error reduction for QMA (the arguments holds for any $1 \leq k \leq \text{poly}(n)$). Thus, we have:

- If $x \in A_{\text{YES}}$, then Pr[Protocol 6.4.4 accepts] $\geq 2/3$,
- If $x \in A_{\text{NO}}$, then Pr[Protocol 6.4.4 accepts] $\leq 1/3$,

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

6.5 Open problems

In this chapter, we have studied the quantum PCP conjecture in the proof-checking formulation (as opposed to the more commonly considered local Hamiltonian formulation). Ironically, many of our results rely on reductions to the local Hamiltonian problem. Our contributions include both novel statements and formal proofs of what we consider to be "folklore knowledge"—results that are widely believed to be true but for which we could not find existing proofs in the literature. Given the renewed interest in the quantum PCP conjecture, we believe this was an effort worth undertaking. We conclude by listing several questions on quantum PCPs which we have not yet resolved but believe are interesting directions for future work.

Locality reductions. We do not know whether Theorem 6.4.4 already holds when only the 2-local Hamiltonian problem is in QCMA. In [BH13b], it is claimed that the 2-local Hamiltonian problem is complete for $\mathsf{QPCP}[q]$ for *any* constant $q = \mathcal{O}(1)$ (which would imply the above statement), but it is not clear to us that this is actually the case. The reason is that the gadget constructions of [BDLT08] mentioned in [BH13b], which transform a q-local Hamiltonian into a 2-local Hamiltonian while preserving the relative promise gap, only work when every qubit is involved in only a *constant* number of terms. In fact, the existence of a transformation that maps a q-local Hamiltonian with interaction degree $\Omega(n)$ to a 2-local Hamiltonian with interaction degree $\Omega(n)$ would directly imply that the local Hamiltonian problem with constant promise gap is in NP, thereby disproving the quantum PCP conjecture for any constant q.

The reason is as follows:[6] Take any q-local Hamiltonian H, $q \geq 2$ constant, with some interaction graph of degree $d \geq 2$, and assume without loss of generality that $\|H\| = 1$, with completeness and soundness parameters b and a,

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

respectively, satisfying $b - a = \Omega(1)$. Then the Hamiltonian $H' = H^2$ is $2q$-local, has operator norm still equal to 1, has interaction degree $\Omega(n)$, and completeness and soundness parameters b', a' satisfying $b' - a' \geq (b - a)^2 = \Omega(1)$. If there existed a transformation from H' to a 2-local Hamiltonian H'' with interaction degree $\mathcal{O}(n)$, and completeness and soundness parameters a'', b'' such that $b'' - a'')/\|H''\| = \Omega(1)$, then by [BH13b, Cor. 5] one could decide the correct energy of H in NP.

6.5.1. Open problem. *Is the* 2*-local Hamiltonian problem with constant promise gap hard for* $\mathsf{QPCP}[q]$ *for all* $q = \mathcal{O}(1)$*, with respect to quantum reductions?*

Quantum PCPs with perfect completeness. A major downside of our technique—leveraging the reduction from QPCP to the local Hamiltonian problem—is that it fails to preserve perfect completeness. This limitation arises from the fact that the Hamiltonian is only learned up to some small error. It remains an open question whether alternative techniques could be used to obtain similar results in the perfect completeness setting. For instance, one could ask whether non-adaptive quantum PCPs can still simulate adaptive ones when the number of quantum proof queries is restricted to be constant.

6.5.2. Open problem. *Do our results hold for quantum PCPs with perfect completeness?*

Strong error reduction. It is easy to show that non-adaptive quantum PCPs with near-perfect completeness allow, just like QMA [MW05], for strong error reduction (see the full work [BHW25]). However, it is not known whether this also holds in the general case. The standard approach of taking polynomials of the Hamiltonian to manipulate its spectrum appears to be incompatible with Kitaev's energy estimation protocol, as it generally introduces terms with negative or large coefficients. This conflicts with the requirements of the protocol described in Section 6.3.2, which necessitates that each term be positive semi-definite and have operator norm at most one.

6.5.3. Open problem. *Do arbitrary quantum PCPs with completeness* c *and soundness* s*, where* $c - s = \Omega(1)$*, allow for strong error reduction?*

Chapter 7

Quantum PCPs with classical proofs

7.1 Introduction

In the introduction of Chapter 6, we discussed quantum PCPs in their fully quantum setting: both the verifier and the proof are quantum. A natural question is therefore what happens if the proof is made *classical* (making the verifier classical and the proof quantum is not sensible, unless some special type of access is granted to the classical verifier). With unrestricted access to a single copy of the proof, this leads to the QMA versus QCMA question, for which arguments exist both in favour and against (see Chapter 1). In particular, since quantum proofs suffer from the no-cloning theorem—which is one of the obstructions preventing classical PCP constructions from being applied in the quantum setting—a "QCMA-like" quantum PCP might be more easily attainable than a QMA one.

However, in defining a quantum PCP with a classical proof, one immediately faces a design choice: should the queries to the proof be classical or quantum? Classical queries, where only a single bit can be read out per query, capture the *locality* aspect of a PCP. Quantum queries, in which the whole proof can be "accessed" via superposition in a single query, are more faithful to the standard quantum query model. Moreover, quantum queries are known to offer exponential advantages over classical ones for certain computational tasks [Sim97, CCD$^+$03, AA15]. Hence, the following question will be central to this chapter:

Can classical proofs be more efficiently verified by a quantum verifier?

Of course, in either query model, quantum-classical PCPs *can* verify NP-proofs more efficiently (via the classical PCP theorem) and they capture everything in BQP. So the real question is whether they can do anything interesting outside the union of those two classes.

7.1.1 Results in this chapter

In this chapter, we define quantum-classical PCPs as the complexity classes $\mathsf{QCPCP}_{c,s}[q]$ (for classical queries) and $\mathsf{QCPCP}_{Q,c,s}[q]$ (for quantum queries), where q is the total number of queries made to a classical proof. Here, c and s are the completeness and soundness parameters, respectively, which, when omitted, are taken to be $c = 2/3$ and $s = 1/3$. We will prove the following three results:

- For any $q = \mathcal{O}(1)$ and any c, s such that $c - s \geq 1/\mathrm{poly}(n)$, we have

$$\mathsf{QCPCP}_{Q,c,s}[q] \subseteq \mathsf{QCPCP}_{1-\delta,1/2+\delta}[3] \subseteq \mathsf{BQ} \cdot \mathsf{NP},$$

  for any constant $\delta > 0$;

- For any $q = \mathcal{O}(1)$, there exists an oracle relative to which $\mathsf{QCPCP}_Q[q]$ (and thus $\mathsf{QCPCP}[q]$) does not capture the power of QCMA;

- For any $c \in \mathbb{Z}_+$, there exists an oracle relative to which $\mathsf{QCPCP}[\mathcal{O}(\log^c n)]$ is strictly less powerful than $\mathsf{QCPCP}_Q[\mathcal{O}(\log^c n)]$.

Here, "BQ" denotes a quantum extension of Schöning's BP-operator [Sch89], such that $\mathsf{BQ}\cdot\mathcal{C}$, for a class $\mathcal{C}$, contains all promise problems that admit a quantum reduction to any promise problem B in $\mathcal{C}$. Intuitively, when $\mathcal{C}$ is a classical complexity class, the BQ-operator can be thought of as "pulling the quantumness" out of a problem: it performs quantum pre-processing on the input to transform the problem into something suitable for $\mathcal{C}$. In Section 7.2, we show that the BQ-operator satisfies several properties analogous to those of the BP-operator (Proposition 7.2.9).

The following two conclusions can be drawn from the first bullet point above. First, any constant-query quantum-classical PCP protocol (in either query model) can be simulated by a quantum-classical PCP making only 3 *classical* queries, while maintaining a constant promise gap. Even more surprisingly, this holds even when the original completeness/soundness gap was inverse polynomial, rather than constant, demonstrating that amplification in this regime can be achieved whilst simultaneously decreasing the number of queries (provided the original number of queries is larger than 3). Second, it states that one can pull out the quantumness of quantum-classical PCPs, in terms of its interpretation in the context of the BQ-operator, no matter if the access to the proof is quantum or classical. Since it seems very unlikely that $\mathsf{QCMA} = \mathsf{BQ} \cdot \mathsf{NP}$—as it would imply that the "quantum part" in the computation does not have to use the proof—this provides strong evidence quantum verifiers cannot process classical proofs in quantum computations using only a small number of queries, contrasting the fully classical case.

7.2 Quantum reductions and the BQ-operator

In our first section, we will briefly review Schöning's BP-operator and some of its properties, and introduce a quantum analogue, the "BQ-operator".

7.2.1 Randomised reductions and Schöning's BP-operator

The BP operator is applied to a complexity class $\mathcal{C}$ by allowing the machines defining membership in $\mathcal{C}$ to use a polynomial-length random string during their computation.

7.2.1. Definition. Let $\mathcal{C}$ be any class of promise problems. Then $\mathsf{BP}\cdot\mathcal{C}$ consists of all promise problems $A = (A_{\text{YES}}, A_{\text{NO}})$ for which there exists a promise problem $B = (B_{\text{YES}}, B_{\text{NO}}) \in \mathcal{C}$, and a polynomial p, such that for all $x \in \{0,1\}^*$ the following conditions hold:

- Completeness: If $x \in A_{\text{YES}}$, then $\Pr_{z\in\{0,1\}^{p(|x|)}}[(x,z) \in B_{\text{YES}}] \geq 2/3$,
- Soundness: If $x \in A_{\text{NO}}$, then $\Pr_{z\in\{0,1\}^{p(|x|)}}[(x,z) \in B_{\text{NO}}] \geq 2/3$,

where (x,z) is encoded as a single input string and the probability is over uniformly random $z \in \{0,1\}^{p(|x|)}$.

For our purposes, it will also be useful to consider an alternative definition, which is equivalent to Definition 7.2.1 when we consider classes $\mathcal{C} \supseteq \mathsf{P}$. For this, we recall the definition of randomised reductions.

7.2.2. Definition (Randomised reductions). Let $A = (A_{\text{YES}}, A_{\text{NO}})$ and $B = (B_{\text{YES}}, B_{\text{NO}})$ be promise problems. We say that $A \leq_r B$ if there exists a polynomial-time probabilistic Turing machine M such that:

- Completeness: If $x \in A_{\text{YES}}$, then $\Pr_z[M(x,z) \in B_{\text{YES}}] \geq 2/3$,
- Soundness: If $x \in A_{\text{NO}}$, then $\Pr_z[M(x,z) \in B_{\text{NO}}] \geq 2/3$,

where z are the random bits used by M.

Note that randomised reductions as defined here are not transitive: i.e., $A \leq_r B$ and $B \leq_r C$ do not necessarily imply $A \leq_r C$.[1]

Using Definition 7.2.2, the BP-operator applied to a class $\mathcal{C}$ gives a new class that contains all problems with randomised reductions to a problem $B \in \mathcal{C}$.

[1] Transitivity would hold if the probabilities in the randomised reductions could be amplified.

7.2.3. Definition (BP-operator). Let $\mathcal{C} \supseteq \mathsf{P}$ be a class of (promise) problems. Then $\mathsf{BP} \cdot \mathcal{C}$ consists of all (promise) problems $A = (A_{\text{YES}}, A_{\text{NO}})$ for which there exists a (promise) problem $B = (B_{\text{YES}}, B_{\text{NO}}) \in \mathcal{C}$ such that $A \leq_r B$, i.e.,

$$\mathsf{BP} \cdot \mathcal{C} = \{A : \exists B \in \mathcal{C} \text{ such that } A \leq_r B\}.$$

Clearly, Definition 7.2.6 captures Definition 7.2.1, by letting the randomised reduction consist simply of directly outputting the concatenation of random bits z and the input x. For the other direction, it is crucial that $\mathcal{C} \supseteq \mathsf{P}$: given a random string z and the input x, the machine in the class $\mathcal{C}$ can simply compute $M(x, z)$ from Definition 7.2.2 before its "standard" verification procedure is performed. Note that the Turing machine $M(x, z)$ is deterministic as we have already fixed all the random bits z. The necessity of condition $\mathcal{C} \supseteq \mathsf{P}$ also follows from the fact that Definition 7.2.3 always captures at least the power of BPP, whereas this might not necessarily be true for Definition 7.2.1 when $\mathsf{P} \not\subset \mathcal{C}$. From the above definitions, it is also clear that $\mathsf{BP} \cdot \mathsf{P} = \mathsf{BPP}$.

Let us now state some basic properties known for the BP-operator. For this, we require the notion of majority reducibility. A promise problem A is said to be *majority reducible* to a problem B, denoted $A \leq^p_{\text{maj}} B$, if there exists a polynomial-time computable function f mapping strings to sequences of strings such that for all x, if $f(x) = (y_1, \ldots, y_k)$, then:

- If $x \in A_{\text{YES}}$, then $y_i \in B_{\text{YES}}$ for strictly more than half of the indices i;
- If $x \in A_{\text{NO}}$, then $y_i \in B_{\text{NO}}$ for strictly more than half of the indices i.

It is well-known that if a class $\mathcal{C}$ is closed under majority reducibility, i.e., if $A \leq^p_{\text{maj}} B$ and $B \in \mathcal{C}$ implies that $A \in \mathcal{C}$, then the soundness and completeness parameters of class $\mathsf{BP} \cdot \mathcal{C}$ can be amplified to become inversely exponentially close to 1 and 0, respectively [Sch89].

We now state some well-known properties of the BP-operator:

7.2.4. Proposition (Some properties of the BP-operator [KST12]). *For all classes of promise problems $\mathcal{C}, \mathcal{D}$, the following hold:*

(i) If $\mathcal{C} \subseteq \mathcal{D}$, then $\mathsf{BP} \cdot \mathcal{C} \subseteq \mathsf{BP} \cdot \mathcal{D}$.

(ii) If $\mathcal{C}$ is closed under majority reducibility, then $\mathsf{BP} \cdot \mathsf{BP} \cdot \mathcal{C} = \mathsf{BP} \cdot \mathcal{C}$.

(iii) $\mathsf{BP} \cdot \mathcal{C} \subseteq \mathsf{BPP}^{\mathcal{C}}$.

Properties (i) and (iii) follow directly from the definition of the BP-operator, while (ii) can be shown using probability amplification [KST12].

If we take $\mathcal{C} = \mathsf{NP}$, we can simply choose 3-SAT as a canonical problem to reduce to, since 3-SAT is NP-complete. It is a well-known result that $\mathsf{BP} \cdot \mathsf{NP} = \mathsf{AM}$ [Sch89]. Finally, observe that the BP-operator can indeed be interpreted as "pulling out the randomness" from a complexity class. For example, we have $\mathsf{BP} \cdot \mathsf{MA} = \mathsf{BP} \cdot \mathsf{NP}$, since $\mathsf{BP} \cdot \mathsf{MA} \subseteq \mathsf{BP} \cdot \mathsf{AM} = \mathsf{BP} \cdot \mathsf{BP} \cdot \mathsf{NP} = \mathsf{BP} \cdot \mathsf{NP}$ and $\mathsf{BP} \cdot \mathsf{NP} \subseteq \mathsf{BP} \cdot \mathsf{MA}$, using properties (i) and (ii) from Proposition 7.2.4.

In the next subsection, we define a quantum analogue of this operator, which we call the BQ-operator, and explore how it mirrors the behaviour of BP in a quantum setting.

7.2.2 The BQ-operator

As in the previous subsection for randomised computation, we introduce a notion of reductions, but this time tailored to quantum polynomial-time computation.

7.2.5. Definition (Quantum reductions). Let $A = (A_{\text{YES}}, A_{\text{NO}})$ and $B = (B_{\text{YES}}, B_{\text{NO}})$ be promise problems. We say that $A \leq_q B$ if there exist a polynomials $q, w : \mathbb{N} \to \mathbb{N}$ and a polynomial-time quantum algorithm $\mathcal{A}$ such that, on input $x \in \{0,1\}^*$, $\mathcal{A}$ applies a quantum circuit V to the state $|x\rangle |0\rangle^{\otimes w(|x|)}$, and then measures $q(|x|)$ designated output qubits in the computational basis to obtain an output string $z \in \{0,1\}^{q(|x|)}$, satisfying:

- Completeness: If $x \in A_{\text{YES}}$, then $\Pr_z[z \in B_{\text{YES}}] \geq 2/3$,
- Soundness: If $x \in A_{\text{NO}}$, then $\Pr_z[z \in B_{\text{NO}}] \geq 2/3$,

where the probability is taken over measurement outcomes z.

Similar to randomised reductions, quantum reductions are also not transitive. Having formally defined our notion of quantum reductions, we can define the BQ-operator in a similar way as the BP-operator as per Definition 7.2.3.

7.2.6. Definition (BQ-operator). Let $\mathcal{C}$ be any class of (promise) problems. The class $\mathsf{BQ} \cdot \mathcal{C}$ consists of all (promise) problems $A = (A_{\text{YES}}, A_{\text{NO}})$ for which there exists a (promise) problem $B = (B_{\text{YES}}, B_{\text{NO}}) \in \mathcal{C}$ such that $A \leq_q B$, i.e.,

$$\mathsf{BQ} \cdot \mathcal{C} = \{A : \exists B \in \mathcal{C} \text{ such that } A \leq_q B\}.$$

Similarly, we can also provide an alternative definition to Definition 7.2.6 to mirror Definition 7.2.1, which can be viewed as feeding randomness generated by a quantum algorithm to the verifier in the corresponding class. Crucially, the quantum algorithm's output is allowed to depend on the input x.

7.2.7. Definition. Let $\mathcal{C}$ be any class of (promise) problems. Then, $\mathsf{BQ} \cdot \mathcal{C}$ consists of all (promise) problems $A = (A_{\text{YES}}, A_{\text{NO}})$ for which there exists a (promise) problem $B = (B_{\text{YES}}, B_{\text{NO}}) \in \mathcal{C}$, a polynomial q, and a polynomial-time quantum algorithm $\mathcal{A}$ such that for all inputs $x \in \{0,1\}^*$:

- Completeness: If $x \in A_{\text{YES}}$, then $\Pr_z[(x,z) \in B_{\text{YES}}] \geq 2/3$,
- Soundness: If $x \in A_{\text{NO}}$, then $\Pr_z[(x,z) \in B_{\text{NO}}] \geq 2/3$,

where $z \in \{0,1\}^{q(|x|)}$ is the measurement outcome of $\mathcal{A}$ on input x.

Since the quantum algorithm generating the random bits z is allowed to use the input x, we automatically have that Definitions 7.2.6 and 7.2.7 are equivalent. From the above definitions, it is clear that $\mathsf{BQ} \cdot \mathsf{P} = \mathsf{BQP}$, just as $\mathsf{BP} \cdot \mathsf{P} = \mathsf{BPP}$, which makes it at least from that point of view a reasonable definition.

Let us now show that the same properties of the BP-operator (Proposition 7.2.4) also hold for the BQ-operator. We will need the following amplification lemma, which can be proven in the same way as it was for the BP-operator.

7.2.8. Lemma (Probability amplification for the BQ-operator). *Let $\mathcal{C}$ be a class of (promise) problems closed under majority reducibility. Then for every (promise) problem $A = (A_{\text{YES}}, A_{\text{NO}}) \in \mathsf{BQ} \cdot \mathcal{C}$ and every polynomial $p : \mathbb{N} \to \mathbb{N}$, there exists a (promise) problem $C = (C_{\text{YES}}, C_{\text{NO}}) \in \mathcal{C}$ and a polynomial-time quantum algorithm $\mathcal{B}$ such that for all inputs $x \in \{0,1\}^*$:*

- *If $x \in A_{\text{YES}}$, then $\Pr_z[(x,z) \in C_{\text{YES}}] \geq 1 - 2^{-p(|x|)}$,*
- *If $x \in A_{\text{NO}}$, then $\Pr_z[(x,z) \in C_{\text{NO}}] \geq 1 - 2^{-p(|x|)}$,*

where $z \in \{0,1\}^{q(|x|)}$ is the measurement outcome of $\mathcal{B}$ on input x.

Proof:
By Definition 7.2.7, there exists a problem $B = (B_{\text{YES}}, B_{\text{NO}}) \in \mathcal{C}$ and a polynomial-time quantum algorithm $\mathcal{A}$, such that

- if $x \in A_{\text{YES}}$, then $\Pr_y[(x,y) \in B_{\text{YES}}] \geq 2/3$;
- if $x \in A_{\text{NO}}$, then $\Pr_y[(x,y) \in B_{\text{NO}}] \geq 2/3$,

where y is the measurement outcome of running $\mathcal{A}$ on input x. Let $n = |x|$ be the input size. Now define $z = (z_1, \ldots, z_k)$, where each z_i is a measurement outcome from an independent run of $\mathcal{A}$, with $k = cp(n)^2$ for some sufficiently large constant $c > 0$. By a standard Chernoff bound, we have:

- If $x \in A_{\text{YES}}$, then $\Pr_z[\text{the majority of } z_i \text{ satisfy } (x, z_i) \in B_{\text{YES}}] \geq 1 - 2^{-p(n)}$,
- If $x \in A_{\text{NO}}$, then $\Pr_z[\text{the majority of } z_i \text{ satisfy } (x, z_i) \in B_{\text{NO}}] \geq 1 - 2^{-p(n)}$.

Define a promise problem $C = (C_{\text{YES}}, C_{\text{NO}})$ as follows:

- $(x, z) \in C_{\text{YES}}$ if the majority of z_i satisfy $(x, z_i) \in B_{\text{YES}}$,
- $(x, z) \in C_{\text{NO}}$ if the majority of z_i satisfy $(x, z_i) \in B_{\text{NO}}$.

Since $\mathcal{C}$ is closed under majority reducibility, we have $C \in \mathcal{C}$, because $C \leq^p_{\text{maj}} B$ and the input (x, z) encodes k separate instances $(x, z_1), \ldots, (x, z_k)$ of the problem $B \in \mathcal{C}$. Thus, for all x, we obtain

- If $x \in A_{\text{YES}}$, then $\Pr_z[(x, z) \in C_{\text{YES}}] \geq 1 - 2^{-p(n)}$,
- If $x \in A_{\text{NO}}$, then $\Pr_z[(x, z) \in C_{\text{NO}}] \geq 1 - 2^{-p(n)}$.

The quantum algorithm $\mathcal{B}$ runs k independent executions of $\mathcal{A}$ and measures all designated output qubits to generate the instance (x, z). □

7.2.9. Proposition. *For all classes of (promise) problems $\mathcal{C}, \mathcal{D}$, the following hold:*

(i) $\mathcal{C} \subseteq \mathcal{D}$ *implies* $\mathsf{BQ} \cdot \mathcal{C} \subseteq \mathsf{BQ} \cdot \mathcal{D}$.

(ii) $\mathsf{BP} \cdot \mathcal{C} \subseteq \mathsf{BQ} \cdot \mathcal{C}$.

(iii) $\mathsf{BQ} \cdot \mathsf{BQ} \cdot \mathcal{C} = \mathsf{BQ} \cdot \mathcal{C}$ *if the class* $\mathcal{C}$ *is closed under majority reducibility.*

(iv) $\mathsf{BQ} \cdot \mathcal{C} \subseteq \mathsf{BQP}^{\mathcal{C}}$.

Proof:
Points (i), (ii), and (iv) follow directly from the definition of the BQ-operator. For (iii), assume $\mathcal{C}$ is closed under majority reducibility. Then so are $\mathsf{BQ} \cdot \mathcal{C}$ and $\mathsf{BQ} \cdot \mathsf{BQ} \cdot \mathcal{C}$ by Lemma 7.2.8, since $\mathsf{BQ} \cdot \mathcal{C}$ trivially contains P. We give the argument for the case $x \in A_{\text{YES}}$, as the NO-case is analogous (just ensure that soundness is amplified such that it is at most $1 - \sqrt{\frac{2}{3}}$ in this setup). Let $A \in \mathsf{BQ} \cdot \mathsf{BQ} \cdot \mathcal{C}$, so there exists a problem $B \in \mathsf{BQ} \cdot \mathcal{C}$ and a quantum algorithm $\mathcal{A}$ such that if $x \in A_{\text{YES}}$, then

$$\Pr_y[(x, y) \in B_{\text{YES}}] \geq \sqrt{\frac{2}{3}},$$

where y is the output of measuring $\mathcal{A}(x)$. Since $B \in \mathsf{BQ} \cdot \mathcal{C}$, and $\mathsf{BQ} \cdot \mathcal{C}$ is closed under majority reducibility, by Lemma 7.2.8 we can amplify this problem to obtain a problem $C \in \mathcal{C}$ and a quantum algorithm $\mathcal{A}'$ such that, if $(x, y) \in B_{\text{YES}}$, then

$$\Pr_z[(x, y, z) \in C_{\text{YES}}] \geq \sqrt{\frac{2}{3}},$$

where z is the output of $\mathcal{A}'(x, y)$. Define $\mathcal{A}''$ to run $\mathcal{A}$, then $\mathcal{A}'$ on the output. Then

$$\Pr_{y,z}[(x, y, z) \in C_{\text{YES}}] \geq \left(\sqrt{\frac{2}{3}}\right)^2 = \frac{2}{3}.$$

Since $C \in \mathcal{C}$, this shows $A \in \mathsf{BQ} \cdot \mathcal{C}$. For the other inclusion, note that $\mathcal{C} \subseteq \mathsf{BQ} \cdot \mathcal{C}$ for any $\mathcal{C}$. □

Points (i) and (iii) of Proposition 7.2.9 together imply that $\mathsf{BQ} \cdot \mathsf{BQP} = \mathsf{BQP}$, since $\mathsf{BQP} = \mathsf{BQ} \cdot \mathsf{P} = \mathsf{BQ} \cdot \mathsf{BQ} \cdot \mathsf{P} = \mathsf{BQ} \cdot \mathsf{BQP}$. Furthermore, this also means that $\mathsf{BQ} \cdot \mathsf{P} = \mathsf{BQ} \cdot \mathsf{BPP}$, since

$$\mathsf{BQ} \cdot \mathsf{BPP} = \mathsf{BQ} \cdot \mathsf{BP} \cdot \mathsf{P} \subseteq \mathsf{BQ} \cdot \mathsf{BQ} \cdot \mathsf{P} = \mathsf{BQ} \cdot \mathsf{BQP} = \mathsf{BQP} = \mathsf{BQ} \cdot \mathsf{P},$$

and $\mathsf{BQ} \cdot \mathsf{P} \subseteq \mathsf{BQ} \cdot \mathsf{BPP}$ since $\mathsf{P} \subseteq \mathsf{BPP}$. This is expected, as quantum algorithms can generate randomness on their own.

As before, when we take $\mathcal{C} = \mathsf{NP}$ and choose $B = \text{3-SAT}$ as the problem we reduce to, the class $\mathsf{BQ} \cdot \mathsf{NP}$ contains all problems that admit a quantum reduction to 3-SAT. We have the known inclusions $\mathsf{PCP}[\mathcal{O}(\log(n)), \mathcal{O}(1)] = \mathsf{NP} \subseteq \mathsf{MA} \subseteq \mathsf{AM} = \mathsf{BP} \cdot \mathsf{NP} \subseteq \mathsf{BQ} \cdot \mathsf{NP}$ and $\mathsf{BQP} \subseteq \mathsf{BQ} \cdot \mathsf{NP}$, but the precise relation of $\mathsf{BQ} \cdot \mathsf{NP}$ to QCMA or QMA remains unclear in either inclusion direction.

It is also important to note a key difference between the classes defined via the BP- and BQ-operators in the context of oracles. For instance, it is known that for the BP-operator we have $\mathsf{BPP}^{\mathcal{O}} = \mathsf{BP} \cdot \mathsf{P}^{\mathcal{O}}$ for all oracle sets $\mathcal{O}$ [RR95]. In contrast, there exist oracles $\mathcal{O}$ for which $\mathsf{BQP}^{\mathcal{O}} \neq \mathsf{BQ} \cdot \mathsf{P}^{\mathcal{O}}$; for example, the oracle that encodes Simon's problem. Hence, whenever we discuss oracle separations involving the BQ-operator, we will always interpret $(\mathsf{BQ} \cdot \mathcal{C})^{\mathcal{O}}$ as granting the quantum reduction unitary access to the oracle.

7.3 Quantum-classical PCPs

In this section, we formally define quantum-classical PCPs. We begin by defining quantum-classical PCPs where the verifier makes non-adaptive classical queries to a classical proof.

7.3.1. Definition (Quantum-classical PCPs with classical queries). Let $\mathcal{G}$ be the universal gate set consisting of the generators of the Clifford group and the T-gate. Let $n \in \mathbb{N}$ be the input size, and let $p, q, a, g : \mathbb{N} \to \mathbb{N}$. A promise problem $A = (A_{\text{YES}}, A_{\text{NO}})$ has a (p, q, a, g, c, s)-QCPCP verifier if and only if there exist $c, s : \mathbb{N} \to [0, 1]$ with $c - s > 0$, and a polynomial-time quantum algorithm V over gate set $\mathcal{G}$, using at most $g(n)$ gates, such that:

- on input $x \in \{0, 1\}^n$, and $a(n)$ ancilla qubits initialised to $|0\rangle^{\otimes a(n)}$, the algorithm applies at most $g(n)$ gates from $\mathcal{G}$ to $|x\rangle\,|0\rangle^{\otimes a(n)}$;

- the algorithm is allowed to read[2] at most $q(n)$ bits of y;
- the output is determined by measuring the first qubit and accepting if and only if the outcome is $|1\rangle$;

such that:

- if $x \in A_{\text{YES}}$, there exists a proof $y \in \{0,1\}^{p(n)}$ such that V accepts with probability at least c;
- if $x \in A_{\text{NO}}$, then for all proofs $y \in \{0,1\}^{p(n)}$, V accepts with probability at most s.

If a, g are polynomially bounded, we simply write (p,q,c,s)-QCPCP verifier, further simplified to (p,q)-QCPCP verifier if additionally $c = 2/3$ and $s = 1/3$ hold. (Note: this definition is robust to the choice of universal gate set whenever $s \leq 1 - 1/\exp(n)$, as standard techniques allow simulation with polylogarithmic overhead).

7.3.2. Definition. A promise problem $A = (A_{\text{YES}}, A_{\text{NO}})$ belongs to $\mathsf{QCPCP}_{c,s}[p,q]$ if and only if there exist polynomially bounded functions a, g, such that A has a P-uniform family of (p,q,a,g,c,s)-QCPCP verifiers $\mathcal{V} = \{V_n : n \in \mathbb{N}\}$. If additionally p is polynomially bounded, we simply write $\mathsf{QCPCP}_{c,s}[q]$, and further simplify to $\mathsf{QCPCP}[q]$ if also $c = 2/3$ and $s = 1/3$ hold.

In many settings, one actually allows for a more powerful quantum access model to classical strings: instead of making queries to a single location of the proof, one can also define a slightly more general version of quantum-classical probabilistically checkable proofs by allowing the entries of the proof string to be read *coherently*.

7.3.3. Definition. Consider the same setup as in Definition 7.3.2. A (p,q,a,g,c,s)-QCPCP_Q verifier is defined identically to a (p,q,a,g,c,s)-QCPCP verifier, except that the verifier is allowed to perform quantum queries to the proof $y \in \{0,1\}^{p(n)}$. That is, for a single query, it may apply the unitary oracle U_y defined as

$$U_y : |i\rangle\,|a\rangle \mapsto |i\rangle\,|a \oplus y_i\rangle\,,$$

where i is a bit string denoting an index, $a \in \{0,1\}$, and y_i denotes the ith bit of the classical proof string y. If a, g are polynomially bounded, we simply write (p,q,c,s)-QCPCP_Q verifier, further simplified to (p,q)-QCPCP_Q verifier if additionally $c = 2/3$ and $s = 1/3$ hold.

[2]More precisely, we define "read" as follows: the classical proof y is stored in an additional register W as $|y\rangle$, and the number of classical queries corresponds to the number of qubits in W on which V acts non-trivially. This is similar in spirit to Definition 6.2.1 for quantum proofs.

7.3.4. Definition (Quantum-classical PCPs with quantum queries). A promise problem $A = (A_{\text{YES}}, A_{\text{NO}})$ belongs to $\mathsf{QCPCP}_{Q,c,s}[p,q]$ if and only if there exist polynomially bounded functions a, g, such that A has a P-uniform family of (p,q,a,g,c,s)-QCPCP_Q verifiers $\mathcal{V} = \{V_n : n \in \mathbb{N}\}$. If additionally p is polynomially bounded, we simply write $\mathsf{QCPCP}_{Q,c,s}[q]$, and further simplify to $\mathsf{QCPCP}_Q[q]$ if also $c = 2/3$ and $s = 1/3$ hold.

A quantum-classical PCP conjecture via proof verification could then be formulated as stating that there exists a constant $q = \mathcal{O}(1)$ such that $\mathsf{QCPCP}_Q[q]$, or even $\mathsf{QCPCP}[q]$, captures the full power of QCMA. It is easy to see that when the proof is allowed to be exponentially long, this is immediately seen to be true via the "string-hiding"-interpretation of the Bernstein–Vazirani problem (see also Lemma 7.5.6). Hence, $\mathsf{QCMA} \subseteq \mathsf{QCPCP}_Q[\exp(n), 1]$.

There is another observation one can make simply by considering Definition 7.3.2. Recall from Chapter 6 that the two formulations of the quantum PCP conjecture—the proof verification and the local Hamiltonian problem formulations—are known to be equivalent under *quantum* reductions. It is a longstanding open question whether the $\mathsf{QPCP}[\mathcal{O}(1)]$-to-local-Hamiltonian reduction can be made classical, or whether a quantum reduction is inherently necessary [AAV13]. However, from our definition of $\mathsf{QCPCP}[q]$, it follows that it is unlikely that a reduction with exactly the same properties as the known quantum reduction exists, as illustrated by the following proposition.

7.3.5. Proposition (No-go for classical polynomial-time reductions). *For any $q = \mathcal{O}(1)$ and any $0 \leq \epsilon < 1/6$, there does not exist a classical polynomial-time reduction from the (q)-QPCP circuit verification problem defined by a P-uniform family of verifiers $\{V_x : x \in \{0,1\}^*\}$ (where V_x denotes the verifier V with input x hardwired) to a family of $\mathcal{O}(q)$-local Hamiltonians $\{H_x : x \in \{0,1\}^*\}$ such that for all $x \in \{0,1\}^*$ and all quantum proofs $|\psi\rangle \in (\mathbb{C}^2)^{\text{poly}(|x|)}$:*

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

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

Proof:
The claim follows from the observation that a (q)-QPCP verifier V_x can simulate a (q)-QCPCP verifier Q_x by using the same verification procedure, while restricting the prover to sending classical basis states as quantum proofs. This restriction can be enforced by measuring the q proof qubits in the computational basis before any quantum operation is applied. Now, suppose there exists a classical polynomial-time reduction that maps V_x to a local Hamiltonian H_x such that for all quantum proofs $|\psi\rangle \in (\mathbb{C}^2)^{\text{poly}(|x|)}$,

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

Then for every classical basis state $|y\rangle \in \{0,1\}^{\mathrm{poly}(|x|)}$, the quantity $\langle y| H_x |y\rangle$ can be efficiently computed classically, since H_x is $\mathcal{O}(q)$-local and $|y\rangle$ is a product state. A classical verifier can thus simulate the acceptance probability of the QCPCP verifier Q_x on input y up to ϵ precision. For any $\epsilon < 1/6$, this is sufficient to preserve completeness and soundness, implying that the corresponding QCPCP$[q]$ problem lies in NP. Hence, unless QCPCP$[q] \subseteq$ NP (which would imply BQP $\subseteq$ NP), no such reduction can exist. □

7.4 Pulling the quantumness out of quantum-classical PCPs

In this section, we will study the power of quantum-classical PCPs in both types of query models and connect it to the previously introduced BQ-operator. The key idea will be that quantum-classical PCP verifications naturally induce a decision problem related to multi-linear polynomials, analogous to how quantum PCPs induce a local Hamiltonian problem (see Chapter 6).

7.4.1 A threshold problem for multi-linear polynomials

A *multi-linear polynomial* of degree d in N binary variables with real coefficients $\beta_S \in \mathbb{R}$ is a function $P : \{0,1\}^N \to \mathbb{R}$ of the form

$$P(y) = \sum_{S \subseteq [N], |S| \leq d} \beta_S \prod_{i \in S} y_i, \tag{7.1}$$

where $y = (y_1, y_2, \ldots, y_N)$ is a string of N variables with $y_i \in \{0,1\}$ for all $i \in [N]$, and the sum is taken over all subsets S (including the empty subset $\emptyset$) of the index set $[N] = \{1, 2, \ldots, N\}$ where $|S| \leq d$. For example, any multi-linear polynomial of degree 2 can be written as

$$P(y) = \alpha_\emptyset + \sum_i \alpha_{\{i\}} y_i + \sum_{i<j} \alpha_{\{i,j\}} y_i y_j.$$

We introduce the following decision problem, which we call the *multi-linear polynomial threshold problem.*

7.4.1. Definition (Multi-linear polynomial threshold problem). Let $P(y)$ be a multi-linear polynomial of degree d in N binary variables with real coefficients $\{\beta_S\}$. Suppose that for all $y \in \{0,1\}^N$, we have $P(y) \in D \subset [-2, 2]$, where D is a finite, evenly spaced set satisfying $\log(|D|) \leq \mathrm{poly}(N)$. Given some $a \in D$, decide whether

(i) there exists a $y \in \{0,1\}^N$ such that $P(y) \geq a$, or

(ii) for all $y \in \{0,1\}^N$, we have $P(y) < a$.

Clearly, the above is a valid decision problem and is contained in NP: this is because P takes values in a finite, discrete set D, which ensures that the promise condition "$< a$" can equivalently be expressed as "$\leq b$" for some $b = a - \delta_D$, where $\delta_D > 0$ is the spacing between successive values in D and is at most exponentially small (so it can be exactly represented using a polynomial number of bits).

We will be interested in multi-linear polynomials of constant degree and bounded range, i.e., $\|P(y)\|_\infty \leq 1$, where $\|P(y)\|_\infty = \sup_y |P(y)|$ denotes the uniform norm. Since we will only have indirect access to the polynomial, there is no a priori restriction on what the coefficients look like, nor even whether they can be accurately approximated using an efficient bit representation. Nevertheless, it can be shown that if the polynomial itself is bounded, then all of its coefficients must also be bounded. To show this, we will make use of the following lemma.

7.4.2. Lemma. *Let $f(l)$ be a recurrence given by*

$$f(l) = x + \sum_{i=0}^{l-1} a_i f(i),$$

where $\{a_i\}$ are real coefficients and the initial value is $f(0) = x$. Then,

$$f(l) = \left(\sum_{S \subseteq \{0,\ldots,l-1\}} \prod_{i \in S} a_i \right) x.$$

Proof:
We prove the result by induction on l.

Base Case. For $l = 0$, we have

$$f(0) = x = \left(\sum_{S \subseteq \emptyset} \prod_{i \in S} a_i \right) x,$$

using that the empty product is defined as 1.

Induction Step. Assume the formula holds for $l = k$, i.e.,

$$f(k) = \left(\sum_{S \subseteq \{0,\ldots,k-1\}} \prod_{i \in S} a_i \right) x.$$

We show it holds for $l = k + 1$:

$$\begin{aligned} f(k+1) &= x + \sum_{i=0}^{k} a_i f(i) \\ &= x + \sum_{i=0}^{k} a_i \left(\sum_{S \subseteq \{0,\ldots,i-1\}} \prod_{j \in S} a_j \right) x \\ &= x \left(1 + \sum_{i=0}^{k} \sum_{S \subseteq \{0,\ldots,i-1\}} a_i \prod_{j \in S} a_j \right). \end{aligned}$$

Now observe that each term in the double sum corresponds uniquely to a subset $S' = S \cup \{i\} \subseteq \{0, \ldots, k\}$, with $i \notin S$. Together with the empty set (corresponding to the standalone x), this enumerates all subsets of $\{0, \ldots, k\}$. Hence,

$$f(k+1) = \left(\sum_{S \subseteq \{0,\ldots,k\}} \prod_{i \in S} a_i \right) x,$$

which completes the induction. □

7.4.3. Lemma. *Let $P : \{0,1\}^N \to \mathbb{R}$ be a multi-linear polynomial of degree d in N binary variables with real coefficients $\{\beta_S\}$. Then for each S, it holds:*

$$|\beta_S| \leq \left(1 + 2^d\right)^d \|P\|_\infty.$$

Proof:
Take a subset $S \subseteq [N]$, $|S| \leq l$, such that $S = \arg\max_{S' : |S'| \leq l} |\beta_{S'}|$. We will keep l as a variable, and give an upper bound on β_S that is monotonically increasing in l. This way, we can simply set $l = d$ at the end of the proof to obtain our upper bound. Let $y^S = (y_1, \ldots, y_N)$, $y_i \in \{0, 1\}$ for all $i \in [N]$, with $y_i = 1$ if and only if $i \in S$. It must hold that for any $S' \subseteq [N]$,

$$\prod_{i \in S'} y_i^S = 1 \text{ if and only if } S' \subseteq S.$$

We also have that

$$P(y^S) = \sum_{S' \subseteq [N], |S'| \leq d} \beta_{S'} \prod_{i \in S'} y_i^S \leq \|P\|_\infty$$

and thus

$$\sum_{S' \subseteq S} \beta_{S'} \leq \|P\|_\infty,$$

keeping only the non-zero terms. Isolating the β_S term and taking the absolute value gives

$$|\beta_S| \leq \left| \|P\|_\infty - \sum_{S' \subset S} \beta_{S'} \right| \leq \|P\|_\infty + \sum_{S' \subset S} |\beta_{S'}| \leq \|P\|_\infty + \sum_{i=0}^{l-1} \binom{l}{i} \max_{S' \subseteq S, |S'|=i} |\beta_{S'}|. \tag{7.2}$$

The idea is now that, for any l, we can upper bound Eq. (7.2), through a recursive formula in l:

$$B(l) = \|P\|_\infty + \sum_{i=0}^{l-1} \binom{l}{i} B(i),$$

with $B(0) = \|P\|_\infty$. Let us show that this is indeed a valid upper bound, in the sense that $B(l) \geq \max_{S' \subseteq S, |S'| \leq l} |\beta_{S'}|$, by using induction. For $l = 0$, we have $B(0) = \|P\|_\infty \geq |\beta_{S'}|$ all $S' \subseteq S$ with $|S'| = 0$, which is correct (this corresponds to the only coefficient that has no variable). For $l = k - 1$, making the induction hypothesis that for all $i \leq k - 1$ that $B(i) \geq \max_{S' \subseteq S : |S'| \leq i|} |\beta_{S'}|$, we have

$$B(k) \geq \|P\|_\infty + \sum_{i=0}^{k-1} \binom{k}{i} B(i) = \|P\|_\infty + \sum_{i=0}^{k-1} \binom{k}{i} \max_{S' \subseteq S, |S'| \leq i} |\beta_{S'}| \geq \max_{S' \subseteq S, |S'| \leq k} |\beta_{S'}|$$

by Eq. (7.2) and the way we defined $|\beta_S|$ at the start of the proof, so it is indeed a valid upper bound on β_S (note that the maximum is performed over an even larger set). According to Lemma 7.4.2, we have that

$$B(l) = \left(\sum_{S' \subseteq \{0,\ldots,l-1\}} \prod_{i \in S'} a_i \right) \|P\|_\infty,$$

where $a_i = \binom{l}{i}$. Since for any i it holds that $a_i \leq 2^l$, we can upper bound $B(l)$ as

$$B(l) \leq \left(\sum_{S \subseteq \{0,\ldots,l-1\}} \prod_{i \in S} 2^l \right) \|P\|_\infty.$$

The sum can be evaluated as

$$\sum_{S \subseteq \{0,\ldots,l-1\}} \prod_{i \in S} 2^l = \sum_{k=0}^{l} \binom{l}{k} 2^{lk} = (1 + 2^l)^l,$$

by the binomial theorem. Since this expression is monotonically increasing in l, it is maximised for $l = d$. Putting everything together, we find an upper bound of

$$|\beta_S| \leq \left(1 + 2^d\right)^d \|P\|_\infty,$$

as was to be shown. □

Hence, when $\|P(y)\|_\infty \leq 1$ and d is constant, the coefficients are also bounded in absolute value by some constant. This implies that all coefficients can be specified up to inverse exponential precision using a polynomial number of bits, and that given such a specification, $P(y)$ can also be exactly represented using a finite number of bits.

7.4.4. Lemma. *Let $P(y)$ be a multi-linear polynomial of degree d in N binary variables with real coefficients $\{\beta_S\}$, where each β_S is specified using k bits of precision. Then, each possible function value of $P(y)$ can be represented exactly using at most $\log\left(\binom{N}{d}\right) + k$ bits.*

Proof:
Since each β_S is represented using k bits, it can take up at most 2^k different values. Since $P(y)$ is always a sum over different choices of coefficients as the individual monomials which make up the polynomial can only be 0 or 1, and there are at most $\binom{N}{d}$ of them, we have that the number of values it can take is upper bounded by $\binom{N}{d}2^k$. Therefore, $\log\left(\binom{N}{d}\right) + k$ bits suffice to exactly describe all possible values for $P(y)$. □

Thus, whenever d is constant and the β_S's are specified using only a polynomial number of bits, we have that the set D is at most exponentially big.

7.4.2 The polynomial method

We will now briefly review the ideas behind the polynomial method of [BBC+01] and show how it connects to quantum-classical PCPs with quantum queries. Let $|0^m\rangle$, where $m = \text{poly}(N)$, be a fixed initial state. The output state of a q-query quantum algorithm with query access to an input $y \in \{0,1\}^N$ in the form of a unitary $U_y : |i\rangle|a\rangle \to |i\rangle|a \oplus y_i\rangle$, with $a \in \{0,1\}$, can be written as[3] (implicitly tensoring the O_y query operations with identity matrices)

$$|\psi_q(y)\rangle = U_q O_y U_{q-1} O_y \ldots O_y U_1 O_y U_0 |0^m\rangle . \tag{7.3}$$

Since this state only depends on the input y through the q query operations, it is shown in [BBC+01] that $|\psi_q(y)\rangle$ can be written as

$$|\psi_q(y)\rangle = \sum_{z \in \{0,1\}^m} \alpha_z(y) |z\rangle ,$$

[3] This also holds when the queries to U_y are controlled [BBC+01].

where each $\alpha_z(y)$ is a multi-linear complex-valued polynomial in y of degree at most q. For the output probability of measuring a designated output qubit in state $|1\rangle$, denoted $P(y)$, we then have

$$P(y) = \|(|1\rangle\langle 1| \otimes \mathbb{I}) |\psi_q(y)\rangle\|^2 = \sum_{z \in \{1\} \times \{0,1\}^{m-1}} |\alpha_z(y)|^2 = \sum_{S \subseteq [N], |S| \leq 2q} \beta_S \prod_{i \in S} y_i, \tag{7.4}$$

where the right-hand side is a multi-linear polynomial of degree at most $2q$ with *real* coefficients β_S.

The original application of the polynomial method is to prove lower bounds for quantum query algorithms, as the existence of a q-query quantum algorithm to compute a function f implies the existence of a degree-$2q$ polynomial that approximates f. In the context of a $\mathsf{QCPCP}_Q[q]$ verifier, the only difference is that the proof $y \in \{0,1\}^N$ is now the string being queried, and the input $x \in \{0,1\}^n$ is used as part of the initial state of the quantum algorithm, i.e., $|x\rangle |0^{m-n}\rangle$.

The polynomial method directly implies the following lemma, since for a fixed input x, the state $|x\rangle$ can be hardcoded into the unitary U_0 from Eq. (7.3).

7.4.5. Lemma. *Let $A = (A_{\text{YES}}, A_{\text{NO}})$ be a promise problem in $\mathsf{QCPCP}_Q[p,q]$ with a P-uniform family of (p,q)-QCPCP_Q-verifiers $\{V_n :\in \mathbb{N}\}$. For a fixed input x with $|x| = n$, let V_x be the corresponding $\mathsf{QCPCP}_Q[p,q]$ verifier, with x hardcoded into the circuit and quantum query access to a classical proof $y \in \{0,1\}^{p(n)}$. Then the acceptance probability of V_x is a multi-linear polynomial in the bits of y, of degree at most $2q$.*

A crucial point is that for quantum-classical PCPs with polynomially-sized classical proofs, the string length N is only *polynomial* in the input size n (whereas for n-bit Boolean functions typically considered in query complexity, we have $N = 2^n$). Therefore, the polynomial that describes the acceptance probability of the $\mathsf{QCPCP}_Q[q]$ verifier, given an input x, has an efficient classical description whenever q is constant (provided the coefficients are specified using a polynomial number of bits, giving exponential precision). We will in the next section see that this allows us to *learn* an approximation of the polynomial that characterises the proof input/output behaviour of the (q)-QCPCP_Q verifier corresponding to any promise problem in $\mathsf{QCPCP}_Q[q]$, given only access to a description of the input x and the verifier.

7.4.3 The quantum reduction

We will show that for a constant number of queries, it is possible to apply a quantum reduction to transform a $\mathsf{QCPCP}_Q[q]$ circuit verification problem into a multi-linear polynomial threshold problem, as per Definition 7.4.1. The key idea is to employ an algorithm that approximately learns all coefficients of the

polynomial (up to degree $2q$) "from the ground up." Specifically, we show that each coefficient β_S can be expressed in terms of a simple estimation procedure, based on previously estimated coefficients $\beta_{S'}$ for $S' \subset S$. This algorithm never has to query a U_y corresponding to actual proof (as this would not be possible since no proof has been given); instead, it runs the (q)-QCPCP_Q verifier on a large number of predetermined settings of "fake" proofs y^S. For any $y^S \in \{0,1\}^{\mathrm{poly}(n)}$, the unitary U_{y^S} can be efficiently implemented given the full description of y^S. The algorithm used in this reduction is given in Algorithm 7.4.1.

Algorithm 7.4.1: Quantum reduction from a problem in QCPCP_Q with verifier V_x to a fixed multi-linear polynomial threshold problem.

Input: A classical description of a (p, q, c, s)-QCPCP verifier verifier V_x with input x hardcoded into it, a maximum failure probability δ.

Set:

$$\epsilon := \frac{(c-s)}{16p(n)^{2q}\left(1+p(n)^{2q}\right)^{2q}}, \quad T := \left\lceil \frac{1}{\epsilon^2} \log\left(\frac{(2q+1)p(n)^{2q}}{\delta}\right)\right\rceil, \quad l := \lceil \log(2/\epsilon + 1)\rceil.$$

Algorithm:

1. For all $S \subseteq [p(n)]$ with $|S| \leq 2q$:

   Suppose that we have already stored $\hat{\beta}_{S'}$ for all $S' \subset S$:

   (a) Define $y \in \{0,1\}^{p(n)}$ such that $y_i = 1$ if and only if $i \in S$.

   (b) For $t \in [T]$, prepare $|\psi_q(y)\rangle$ using V_x, and measure the first qubit in the computational basis. Let $X_t = 1$ if the outcome is $|1\rangle$, and $X_t = 0$ otherwise.

   (c) Define the empirical estimate $\hat{X} := \frac{1}{T}\sum_{t=1}^{T} X_t$, and truncate $\hat{X}$ to its first l bits.

   (d) Compute the coefficient estimate $\hat{\beta}_S := \hat{X} - \sum_{S' \subset S} \hat{\beta}_{S'}$, and store $\hat{\beta}_S$.

2. Output:

$$\hat{P}(y) := \sum_{S \subseteq [p(n)], |S| \leq 2q} \hat{\beta}_S \prod_{i \in S} y_i, \quad a := \arg\min_{d \in D} \left| d - \frac{c+s}{2} \right|,$$
$$D := \left\{ -2^{4+4q} + i \cdot 2^{-l} \cdot n^{-2q} \;\middle|\; i \in \left\{0, 1, \ldots, n^{2q} \cdot 2^{l+5+4q}\right\}\right\}$$

7.4.6. THEOREM. *Let $q = \mathcal{O}(1)$ be constant, and suppose $c - s \geq 1/\mathrm{poly}(n)$. Let $A = (A_{\text{YES}}, A_{\text{NO}})$ be a promise problem in $\mathsf{QCPCP}_{Q,c,s}[p,q]$ with a P-uniform family of (p,q)-QCPCP_Q-verifiers $\{V_n :\in \mathbb{N}\}$. Then, for any $\delta \geq 1/\exp(n)$, there exists a quantum polynomial-time reduction to the multi-linear polynomial threshold problem with degree $d = 2q$ and $\log(|D|) = \mathrm{poly}(n)$, which succeeds with probability at least $1 - \delta$. Moreover, conditioned on the reduction succeeding, this produces a fixed instance.*

Proof:
By definition, for any promise problem $A = (A_{\text{YES}}, A_{\text{NO}})$ in $\mathsf{QCPCP}_{Q,c,s}[q]$, there exist polynomially bounded functions $p, a, g : \mathbb{N} \to \mathbb{N}$ such that A has a P-uniform family of (p,q,a,g,c,s)-QCPCP_Q verifiers $\mathcal{V} = \{V_n\}$. Fix some input size n, and denote V_x for the corresponding verifier with the input x hardcoded into it.

We prove the reduction by showing the correctness of Algorithm 7.4.1. To make the presentation of the proof more structured, we proceed in several steps: first, we verify that Algorithm 7.4.1 works in principle and establish the minimum required error allowed in the learned polynomial (**The reduction**). Next, we analyse to what precision each coefficient must be estimated (**Estimating $\left\|\Pi_1 \left|\psi_q(y^S)\right\rangle\right\|^2$**, **Estimation precision**), ensure the overall estimation succeeds with high probability (**Success probability**), and determine how many bits of each estimate are correct (**Bits of precision and D**). Finally, we show that all steps can be implemented in quantum polynomial time (**Complexity and combining relevant parameters**).

The reduction. We proceed by proving that the reduction given by Algorithm 7.4.1 works in principle. By Lemma 7.4.5, we have that the probability that the $\mathsf{QCPCP}_Q[q]$ verifier V_x, with input x hardcoded into it, accepts a proof y is given by

$$P(y) = \sum_{S \subseteq [p(n)], |S| \leq 2q} \beta_S \prod_{i \in S} y_i.$$

With each S, we associate a string $y^S \in \{0,1\}^{p(n)}$ such that $y_i^S = 1$ if and only if $i \in S$. For this string y^S, we can efficiently construct a "fake" proof unitary U_{y^S}, and consider the action of the QCPCP_Q verifier V_x when given query access to U_{y^S}. For any y^S, the corresponding U_{y^S} can be efficiently implemented as it only operates on $\lceil \log p(n) \rceil + 1 = \mathcal{O}(\log n)$ qubits. We can express its acceptance probability as

$$\begin{aligned} P(y^S) &= \sum_{S' \subseteq [p(n)], |S'| \leq 2q} \beta_{S'} \prod_{i \in S'} y_i^S \\ &= \sum_{S' \subset S} \beta_{S'} + \beta_S, \end{aligned} \tag{7.5}$$

keeping only the non-zero terms. Write $\Pi_1 = |1\rangle\langle 1| \otimes \mathbb{I}$ for the projection onto the output qubit being in state $|1\rangle$. Since we have that $P(y^S) = \left\|\Pi_1 \left|\psi_q(y^S)\right\rangle\right\|^2$ by Eq. (7.4), we can rewrite Eq. (7.5) as

$$\beta_S = \left\|\Pi_1 \left|\psi_q(y^S)\right\rangle\right\|^2 - \sum_{S' \subset S} \beta_{S'}.$$

Each coefficient β_S can thus be expressed in terms of other coefficients $\beta_{S'}$, where $S' \subset S$, and the probability that V_x accepts the input y^S. Hence, if no errors occurred in the estimation of each of the β_S, Algorithm 7.4.1 would give a perfect description of the polynomial P. However, errors will be introduced, since $\|\Pi_1 |\psi_q(y)\rangle\|^2$ can only be estimated. To ease the presentation, we will introduce several error parameters, which will eventually be combined into the single error parameter ϵ as specified in Algorithm 7.4.1.

Assume that (i) $\left|\hat{\beta}_S - \beta_S\right| \leq \epsilon_1$ for all $S \subseteq [p(n)], |S| \leq 2q$, (ii) the overall estimation succeeds with probability at least $1-\delta$, and (iii) we have already obtained a value of a that satisfies the above criteria. Then the estimated polynomial $\hat{P}$ satisfies

$$\begin{aligned} \left\|\hat{P}(y) - P(y)\right\|_\infty &\leq \sum_{S \subseteq [p(n)], |S| \leq 2q} \left|\hat{\beta}_S - \beta_S\right| \\ &\leq p(n)^{2q} \epsilon_1. \end{aligned} \tag{7.6}$$

To distinguish both the completeness and soundness case, we require that this error is at most $(c-s)/4$, which is guaranteed when $\epsilon_1 := \frac{1}{4}(c-s)p(n)^{-2q}$. For this choice of ϵ_1, we have:

- If $x \in A_{\text{YES}}$, then $\Pr\left[\exists y : \hat{P}(y) \geq a\right] \geq 1 - \delta$,
- If $x \in A_{\text{NO}}$, then $\Pr\left[\forall y : \hat{P}(y) < a\right] \geq 1 - \delta$,

where the probability is taken over the outcome of the reduction. We now show that, for our parameter choices, all three conditions above are satisfied, and that we indeed produce a fixed polynomial every time the reduction succeeds.

Estimating $\left\|\Pi_1 \left|\psi_q(y^S)\right\rangle\right\|^2$. In the reduction, we must estimate $\left\|\Pi_1 \left|\psi_q(y^S)\right\rangle\right\|^2 \in [0,1]$ for several values of y^S, which is done in Algorithm 7.4.1 by preparing the state $\left|\psi_q(y^S)\right\rangle$ and measuring the first qubit in the computational basis. Suppose for now we want to estimate $\left\|\Pi_1 \left|\psi_q(y^S)\right\rangle\right\|^2$ up to precision ϵ, using the standard mean estimation procedure. Let X_t be the random variable such that $X_t = 1$ if the outcome is $|1\rangle$ and $X_t = 0$ if the outcome is $|0\rangle$. Step 1c produces the random variable

$$\hat{X} = \frac{1}{T} \sum_{t \in [T]} X_t,$$

i.e., the empirical average of T independent outcomes. Since $0 \leq X_t \leq 1$, Hoeffding's inequality gives

$$\Pr\left[\left|\hat{X} - \mathbb{E}[X]\right| \leq \epsilon\right] \geq 1 - \exp\left(-2T\epsilon^2\right).$$

To achieve success probability $1 - \delta'$, it suffices to choose

$$T \geq \frac{\log\left(\frac{1}{\delta'}\right)}{2\epsilon^2}. \tag{7.7}$$

Estimation precision. We now determine a minimum required precision ϵ_2 for estimating each $\left\|\Pi_1 \left|\psi_q(y^S)\right\rangle\right\|^2$ in order to guarantee that $\left|\hat{\beta}_S - \beta_S\right| \leq \epsilon_1$ for all $S \subseteq [p(n)]$ with $|S| \leq 2q$. Let $X_S := \left\|\Pi_1 \left|\psi_q(y^S)\right\rangle\right\|^2$ and $\hat{X}_S$ be its estimator. For a given S, we have

$$\begin{aligned}
\left|\beta_S - \hat{\beta}_S\right| &= \left|X_S - \sum_{S' \subset S} \beta_{S'} - \left(\hat{X}_S - \sum_{S' \subset S} \hat{\beta}_{S'}\right)\right| \\
&\leq \left|X_S - \hat{X}_S\right| + \left|\sum_{S' \subset S} \left(\hat{\beta}_{S'} - \beta_{S'}\right)\right| \\
&\leq \left|X_S - \hat{X}_S\right| + \sum_{S' \subset S} \left|\hat{\beta}_{S'} - \beta_{S'}\right|.
\end{aligned}$$

Since the error depends only on the cardinality of the S's, we group terms by this cardinality and define β_l to represent a generic coefficient of cardinality l. Since the error only increases with increasing values l, we only have to check $l = 2q$. Using $\left|X_S - \hat{X}_S\right| \leq \epsilon_2$ for all S, we get

$$\left|\beta_{2q} - \hat{\beta}_{2q}\right| \leq \epsilon_2 + \sum_{i=0}^{2q-1} \binom{p(n)}{i} \left|\hat{\beta}_i - \beta_i\right|.$$

By a similar argument we made in the proof of Lemma 7.4.3, we can upper bound $\left|\beta_{2q} - \hat{\beta}_{2q}\right|$ by a function $B(2q)$ that is defined via a recursion:

$$B(l) = \epsilon_2 + \sum_{i=0}^{l-1} \binom{p(n)}{i} B(i), \tag{7.8}$$

with $B(0) = \epsilon_2$. Eq. (7.8) matches the structure required by Lemma 7.4.2 with $a_i = \binom{p(n)}{i}$. Since $i \leq 2q$, we can bound each term using

$$\binom{p(n)}{i} \leq p(n)^{2q}.$$

By applying Lemma 7.4.2, we find

$$\left|\beta_{2^q} - \hat{\beta}_{2^q}\right| \leq B(2q) \leq \epsilon_2 \left(\sum_{S \subseteq \{0,\ldots,2q-1\}} \prod_{i \in S} p(n)^{2q} \right) = \epsilon_2 \left(1 + p(n)^{2q}\right)^{2q} \leq \epsilon_1,$$

provided we choose

$$\epsilon_2 = \frac{\epsilon_1}{\left(1 + p(n)^{2q}\right)^{2q}}.$$

Success probability. To ensure that the overall reduction succeeds with probability at least $1-\delta$, we must guarantee that the estimation of $\left\|\Pi_1 \left|\psi_q(y^S)\right\rangle\right\|^2$ up to accuracy ϵ succeeds for all relevant subsets S. Recall that we set $1-\delta'$ to be the success probability of each individual estimation. The total number of such subsets S (including the empty subset) is upper bounded as

$$1 + \sum_{l=0}^{2q} \binom{p(n)}{l} \leq (2q+1)p(n)^{2q}.$$

It follows that we require

$$(1-\delta')^{(2q+1)p(n)^{2q}} \geq 1 - \delta,$$

which is equivalent to

$$\delta' \leq 1 - (1-\delta)^{\left((2q+1)p(n)^{2q}\right)^{-1}} \leq \frac{\delta}{(2q+1)p(n)^{2q}},$$

applying the inequality $(1+x)^r \leq 1 + rx$ for $x \geq -1$, $r \in [0,1]$. Thus, it suffices to set

$$\delta' := \frac{\delta}{(2q+1)p(n)^{2q}}.$$

Bits of precision and D. We now translate our estimation guarantee into a bit-precision guarantee. Specifically, we aim to pick an $\epsilon \leq \epsilon_2$ such that taking the first l bits of our estimate $\hat{X}$ guarantees, with high probability, that those bits are correct and that the additive error remains at most ϵ_2. Recall that for any S, the random variable $X_i \in \{0,1\}$, and that our estimate $\hat{X}$ is the average over T trials. Therefore, the set of all possible outcomes of $\hat{X}$ is given by

$$G_T := \left\{0, \frac{1}{T}, \frac{2}{T}, \ldots, 1 - \frac{1}{T}, 1\right\}.$$

Assume T satisfies $T+1 = 2^k$ for some $k \in \mathbb{N}$, so that G_T can be exactly represented using k bits. If we want the first $l \leq k$ bits of $\hat{X}$ to be correct, this is equivalent to requiring

$$\epsilon_2 \leq \frac{1}{2(2^l - 1)}.$$

We now pick an $\epsilon \leq \epsilon_2$ such that this inequality holds tightly. We can do so by setting

$$\frac{1}{2\left(2^{\lceil \log_2(1/(2\epsilon_2)+1)\rceil} - 1\right)} \geq \frac{\epsilon_2}{2(1+\epsilon_2)} \geq \epsilon_2/4 =: \epsilon,$$

so that at least

$$l := \left\lceil \log\left(\frac{1}{2\epsilon_2} + 1\right)\right\rceil = \left\lceil \log\left(\frac{2}{\epsilon} + 1\right)\right\rceil$$

bits of $\hat{X}$ are guaranteed to be correct. Since we build our coefficients β_S only of out of adding different estimates of $\hat{X}$ that have an exact bit representation using l bits, we can simply extend our bit representation of $G_T \subset [0,1]$ with more bits to capture larger possible function values. By Lemma 7.4.3, we know that every coefficient β_S lies in the interval

$$\beta_S \in \left[-\left((1+2^{2q})^{2q}\right),\ (1+2^{2q})^{2q}\right],$$

which means that, conditioning on the reduction succeeding, all $\hat{\beta}_S$ are in the interval

$$\hat{\beta}_S \in \left[-\left((1+2^{2q})^{2q}\right) - \epsilon_1,\ (1+2^{2q})^{2q} + \epsilon_1\right] \subset I,$$

with $I = [-2^{4+4q}, 2^{4+4q}]$. Hence, we can extend our bit representation to now represent I with $k = l+5+4q$ bits, whilst still ensuring that our bit representation exactly contains G_T. By Lemma 7.4.4, we then have that our function values D can be exactly represented using at most $\log\left(\binom{N}{2q}\right) + k$ bits. Since $N = p(n)$, if we take $k' = 2q\log p(n) + k$ and set

$$D := \left\{-2^{4+4q} + i \cdot 2^{5+4q-k'} \;\middle|\; i \in \{0,1,\ldots,2^{k'}\}\right\},$$

we have that $G_T \subset D$ and we can exactly represent all possible function values of our polynomial $\hat{P}$. It remains to choose $a \in D$ such that $a \in (s,c)$ and the reduction is valid. Since the spacing δ_D between successive elements in D satisfies $\delta_D \ll (c-s)/2$, choosing

$$a := \arg\min_{d \in D} \left| d - \frac{c+s}{2} \right|$$

guarantees that $a \in D$ and $a \in (s,c)$. Lastly, note that $\log|D| = \text{poly}(n)$, as desired.

Complexity and combining relevant parameters. Finally, we combine all previously introduced parameters to show that the reduction runs in polynomial time when q is constant. Recall that in Eq. (7.7), we derived the required number of samples T in terms of the final estimation precision ϵ. Using that $\epsilon := \epsilon_2/4$, setting

$$\epsilon := \frac{(c-s)}{16p(n)^{2q}\left(1+p(n)^{2q}\right)^{2q}} = \frac{1}{\mathrm{poly}(n)},$$

we obtain the number of samples as

$$T := \left\lceil \frac{1}{\epsilon^2} \log\left(\frac{(2q+1)p(n)^{2q}}{\delta}\right)\right\rceil = \mathrm{poly}(n).$$

The total runtime of the reduction is upper bounded by

$$\mathcal{O}(p(n)^{2q} \cdot T) = \mathrm{poly}(n),$$

which confirms that the entire procedure is efficient for constant q, $p(n) = \mathrm{poly}(n)$, $c - s \geq 1/\mathrm{poly}(n)$ and $\delta \geq 1/\exp(n)$. □

7.4.4 Implications

As in Chapter 6, quantum reductions turn out to be a versatile tool for proving properties of quantum PCPs. In this section, we apply the quantum reduction from Theorem 7.4.6 to establish several results about quantum-classical PCPs.

A new upper bound on quantum-classical PCPs. Theorem 7.4.6 directly implies the following corollary, as we give a quantum reduction to a problem in NP (note that this is *not* the promise version).

7.4.7. Corollary. *For any constant $q \in \mathbb{N}$, we have*

$$\mathsf{QCPCP}_Q[q] \subseteq \mathsf{BQ} \cdot \mathsf{NP}.$$

The inclusion $\mathsf{BQ} \cdot \mathsf{NP} \subseteq \mathsf{QCPCP}[\mathcal{O}(1)]$ does not necessarily hold, as the quantum reduction might produce different NP problems, each requiring different witnesses. Thus, the prover may not know which proof to provide.

Constant query hierarchy collapse, query equivalence and amplification. We will now show that any constant-query quantum-classical PCP protocol can be simulated by a quantum-classical PCP making only 3 classical queries, while maintaining a constant promise gap. Even more surprisingly, this holds even when the original completeness/soundness gap was inverse polynomial, rather than constant. This shows that we can, in fact, perform gap amplification for quantum-classical PCPs, but that—combined with Corollary 7.4.7—this actually provides evidence *against* the existence of a quantum-classical PCP for QCMA, as it is unlikely that $\mathsf{QCMA} \subseteq \mathsf{BQ} \cdot \mathsf{NP}$.

We will use Håstad's 3-query PCP for NP.

7.4.8. Lemma ([Hå01]). *For every constant $\delta > 0$ and every decision problem $D \in \mathsf{NP}$, there exists a PCP verifier V for D that makes 3 queries, has completeness at least $1 - \delta$, and soundness at most $\frac{1}{2} + \delta$.*

7.4.9. Proposition. *For any $q \in \mathbb{N}$ constant and for any $c, s \geq 1/\mathrm{poly}(n)$, we have*

$$\mathsf{QCPCP}_{Q,c,s}[q] \subseteq \mathsf{QCPCP}_{1-\delta,1/2+\delta}[3],$$

for any constant $0 < \delta < \frac{1}{2}$.

Proof:
From Theorem 7.4.6, we know that for any promise problem $A = (A_{\text{YES}}, A_{\text{NO}})$ in $\mathsf{QCPCP}_{Q,c,s}[q]$, with input x and completeness and soundness parameters c, s satisfying $c - s \geq 1/\mathrm{poly}(n)$, there exists a quantum reduction to a multi-linear polynomial threshold problem, which we denote as D. Conditioned on the quantum reduction succeeding, this reduction is deterministic: since the multi-linear polynomial is learned up to a fixed number of bits of precision, it will always output the same polynomial. Since deciding whether a multi-linear polynomial has a setting for which it evaluates to a value larger than a, or evaluates to a value smaller than a, is in NP, it can be correctly decided by a PCP verifier V from Lemma 7.4.8 with completeness $1 - \delta_1$ and soundness $1/2 + \delta_1$. A $\mathsf{QCPCP}_{1-\delta,1/2+\delta}[3]$ verification protocol Q naturally follows. Let $\delta_1 = \delta_2 = \delta/2$.

1. The prover sends a proof y corresponding to the case where the reduction succeeds.

2. The verifier performs the quantum reduction (with maximum error probability δ_2) to obtain a multi-linear threshold problem D, and then uses the PCP verifier V to decide whether $x \in A_{\text{YES}}$ or $x \in A_{\text{NO}}$.

We have:

- If $x \in A_{\text{YES}}$, then $\Pr[Q \text{ accepts } (x, y)] \geq (1 - \delta_1)(1 - \delta_2) \geq 1 - \delta$,

- If $x \in A_{\text{NO}}$, then $\Pr[Q \text{ accepts } (x, y)] \leq (1 - \delta_1)(1/2 + \delta_2) + \delta_2 \leq 1/2 + \delta$.

Since the reduction from Theorem 7.4.6 runs in time polynomial in $\log(1/\delta_2)$, it will clearly run in polynomial time when δ_2 is constant. □

We give two other remarks and one observation that follow from Proposition 7.4.9.

7.4.10. Remark. The proof ideas behind Proposition 7.4.9 also show that, without loss of generality, a quantum-classical PCP can be assumed to use a uniformly random distribution over (a subset of) the indices of the proof to be queried, followed by a single quantum circuit that determines the outcome. This is because there exist constant-query non-adaptive PCPs that use a uniform random distribution, for example, via the NP-hard gapped version of 3-SAT [HÅ01].

7.4.11. Remark. Even though Proposition 7.4.9 makes use of a classical PCP construction, which is nonrelativising, the theorem itself *does* relativise. This is because the quantum reduction allows the $\mathsf{QCPCP}_Q[q]$ verifier to make queries to an additional oracle, meaning that the learned polynomial will implicitly encode oracle calls. Crucially, the degree of the polynomial depends only on the number of quantum queries to the proof, and not on the presence of any oracle.

Finally, we observe that Proposition 7.4.9 also implies that any verification performed by a quantum-classical PCP with a constant number of adaptive classical queries can be simulated by a quantum-classical PCP verification with a constant number of non-adaptive queries. This also follows from the results in Chapter 6, when the reduction from an adaptive to a non-adaptive quantum PCP presented in that chapter is applied to the setting of quantum-classical PCPs with classical proofs.

7.5 Oracle separations for quantum-classical PCPs

In this section, we will look into quantum-classical PCPs in the presence of oracles. We will see that our result that $\mathsf{QCPCP}[\mathcal{O}(1)] \subseteq \mathsf{BQ} \cdot \mathsf{NP}$ (which holds relative to all oracles, see Remark 7.4.11) implies the existence of a classical oracle relative to which the quantum-classical PCP conjecture is false. We will also study the setting where the number of queries is (poly)logarithmic instead of constant. We show that, relative to an oracle, quantum-classical PCPs with quantum queries are more powerful than those with classical queries.

7.5.1 The OR ∘ Forrelation oracle

We will make use of an oracle from [AIK22], which crucially relies on the following result by Raz and Tal [RT19].

7.5.1. Lemma (From [RT19], Theorem 1.2). *For all sufficiently large N, there exists an explicit distribution $\mathcal{F}_N$ that we call the Forrelation distribution over $\{0,1\}^N$ such that:*

1. *There exists a quantum algorithm $\mathcal{A}$ that makes* polylog(N) *queries and runs in time* polylog(N) *such that*

$$\left| \Pr_{x \in \mathcal{F}_N}[\mathcal{A}(x) = 1] - \Pr_{y\{0,1\}^N}[\mathcal{A}(y) = 1] \right| \geq 1 - \frac{1}{N^2}.$$

2. *For any $C \in \mathsf{AC}^0[\text{quasipoly}(N), \mathcal{O}(1)]$:*

$$\left| \Pr_{x \in \mathcal{F}_N}[C(x) = 1] - \Pr_{y\{0,1\}^N}[C(y) = 1] \right| \leq \frac{\text{polylog}(N)}{\sqrt{N}}$$

The precise definition of the Forrelation distribution is not relevant for our purposes, but can be found in [RT19]. In [AIK22], this result is used to provide oracle separations between $\mathsf{BQP}^{\mathsf{PH}}$ and $\mathsf{PH}^{\mathsf{BQP}}$. We will make use of the following specific oracle from [AIK22].

7.5.2. Definition (OR $\circ$ Forrelation oracle, adapted from [AIK22]). We define $\boldsymbol{\mathcal{O}}$ as the oracle such that for each $n \in \mathbb{N}$, we add into $\boldsymbol{\mathcal{O}}$ a region $\mathcal{R}$ consisting of a function $f_n : \{0,1\}^{n^2} \times \{0,1\}^{n^2} \to \{0,1\}$, defined as follows:

(i) If $0^n \notin L^{\boldsymbol{\mathcal{O}}}$ (i.e., $L^{\boldsymbol{\mathcal{O}}}(0^n) = 0$), then $f_n(x,y)$ is drawn uniformly at random for all $(x,y) \in \{0,1\}^{n^2}$.

(ii) If $0^n \in L^{\boldsymbol{\mathcal{O}}}$ (i.e., $L^{\boldsymbol{\mathcal{O}}}(0^n) = 1$), then there exists a single $\hat{x}$, drawn uniformly at random from $\{0,1\}^{n^2}$, such that for all $y \in \{0,1\}^{n^2}$, the values $f_n(\hat{x}, y)$ are drawn from the Forrelation distribution $\mathcal{F}_{2^{n^2}}$. For all $x \in \{0,1\}^{n^2} \setminus \{\hat{x}\}$ and $y \in \{0,1\}^{n^2}$, the values $f_n(x,y)$ are drawn uniformly at random.

Outside the region $\mathcal{R}$, the oracle $\boldsymbol{\mathcal{O}}$ always outputs 0.

Let us show that the same oracle implies an oracle separation between QCMA and $\mathsf{BQP}^{\mathsf{NP}}$. To establish this separation, it suffices to show that the oracle problem is contained in $\mathsf{QCMA}^{\boldsymbol{\mathcal{O}}}$, as the non-containment in $\mathsf{BQP}^{\mathsf{NP}}$ has already been proven in [AIK22].

7.5.3. Proposition. *There exists an oracle $\boldsymbol{\mathcal{O}}$ relative to which*

$$\mathsf{QCMA}^{\boldsymbol{\mathcal{O}}} \not\subset \mathsf{BQP}^{\mathsf{NP}^{\boldsymbol{\mathcal{O}}}}.$$

Proof:
The proof follows the construction used in the proofs of Corollary 48 and Claim 49 in [AIK22]. Let L be a uniformly random unary language, and let $\mathcal{D}$ be the resulting distribution over oracles $\boldsymbol{\mathcal{O}}$. We now show that $L^{\boldsymbol{\mathcal{O}}} \in \mathsf{QCMA}^{\boldsymbol{\mathcal{O}}}$ with probability 1 over $\boldsymbol{\mathcal{O}} \sim \mathcal{D}$.

Let $M^{\boldsymbol{\mathcal{O}}}(0^n, z)$ be a $\mathsf{QCMA}^{\boldsymbol{\mathcal{O}}}$ verifier which takes input 0^n and a classical witness $z \in \{0,1\}^{n^2}$, and runs the quantum algorithm $\mathcal{A}$ of Lemma 7.5.1 on the function $f_n(z, y)$. If $0^n \in L$, the witness is $z = \hat{x}$, while if $0^n \notin L$, it can be any z. From Lemma 7.5.1, with $N = 2^n$, we have

$$\Pr_{\boldsymbol{\mathcal{O}} \sim \mathcal{D}}[M^{\boldsymbol{\mathcal{O}}}(0^n, z) \neq L^{\boldsymbol{\mathcal{O}}}(0^n)] \leq 2^{-2n^2}.$$

By Markov's inequality,

$$\Pr_{\boldsymbol{\mathcal{O}} \sim \mathcal{D}}\left[\Pr\left[M^{\boldsymbol{\mathcal{O}}}(0^n, z) \neq L^{\boldsymbol{\mathcal{O}}}(0^n)\right] \geq \frac{1}{3}\right] \leq 3 \cdot 2^{-2n^2}.$$

By evoking the Borel-Cantelli Lemma, we can show that with probability 1 over $\boldsymbol{\mathcal{O}}$ we have that $M^{\boldsymbol{\mathcal{O}}}$ correctly decides $L^{\boldsymbol{\mathcal{O}}}(0^n)$ for all but finitely many $n \in \mathbb{N}$. We have

$$\Pr_{\mathcal{O} \sim \mathcal{D}}[M^{\boldsymbol{\mathcal{O}}} \text{ does not decide } L^{\boldsymbol{\mathcal{O}}}(0^n)] \leq \sum_{n=1}^{\infty} \leq 3 \cdot 2^{-2n^2} < \infty.$$

Hence, the probability that $M^{\boldsymbol{\mathcal{O}}}$ fails on infinitely many inputs n is zero. Consequently, we have that $M^{\boldsymbol{\mathcal{O}}}$ can be modified in a $\mathsf{QCMA}^{\boldsymbol{\mathcal{O}}}$ algorithm that correctly decides $L^{\boldsymbol{\mathcal{O}}}(0^n)$ for all $n \in \mathbb{N}$, with probability 1 over $\mathcal{O} \sim \mathcal{D}$. The proof that $L^{\boldsymbol{\mathcal{O}}} \notin \mathsf{BQP}^{\mathsf{PH}^{\boldsymbol{\mathcal{O}}}}$ (and therefore $L^{\boldsymbol{\mathcal{O}}} \notin \mathsf{BQP}^{\mathsf{NP}^{\boldsymbol{\mathcal{O}}}}$) is given in [AIK22], Corollary 48. □

7.5.2 The quantum-classical PCP conjecture is false relative to an oracle

We will first argue that some established inclusions, which are used to derive our desired oracle separation, relativise. This allows us to directly apply the $\mathsf{OR}\circ$ Forrelation oracle as defined in Section 7.5.1.

7.5.4. Lemma. *For any $q = \mathcal{O}(1)$ and all oracles $\boldsymbol{\mathcal{O}}$, we have*

$$\mathsf{QCPCP}[q]^{\boldsymbol{\mathcal{O}}} \subseteq \mathsf{QCPCP}_Q[q]^{\boldsymbol{\mathcal{O}}} \subseteq (\mathsf{BQ} \cdot \mathsf{NP})^{\boldsymbol{\mathcal{O}}} \subseteq \mathsf{BQP}^{\mathsf{NP}^{\boldsymbol{\mathcal{O}}}}.$$

Proof:
Let $\mathcal{O}$ be any oracle set. The inclusion $\mathsf{QCPCP}[q]^{\mathcal{O}} \subseteq \mathsf{QCPCP}_Q[q]^{\mathcal{O}}$ follows trivially by a simulation argument.

The inclusion $\mathsf{QCPCP}_Q[q]^{\mathcal{O}} \subseteq (\mathsf{BQ} \cdot \mathsf{NP})^{\mathcal{O}}$ follows from the fact that in the proof of Theorem 7.4.6, all oracle calls (viewed as quantum queries U_f to $\mathcal{O}$) can be absorbed into the unitaries in Eq. (7.3) that do not correspond to the proof queries. This shows that the proof of Theorem 7.4.6 and thus $\mathsf{QCPCP}_Q[q]^{\mathcal{O}} \subseteq (\mathsf{BQ} \cdot \mathsf{NP})^{\mathcal{O}}$ relativizes (see also Remark 7.4.11).

To show that $(\mathsf{BQ} \cdot \mathsf{NP})^{\mathcal{O}} \subseteq \mathsf{BQP}^{\mathsf{NP}^{\mathcal{O}}}$, consider that $(\mathsf{BQ} \cdot \mathsf{NP})^{\mathcal{O}}$ consists of all promise problems $A = (A_{\text{YES}}, A_{\text{NO}})$ such that there exists a polynomial-time deterministic Turing machine M with access to $\mathcal{O}$, and a polynomial-time quantum algorithm $\mathcal{A}^{\mathcal{O}}$ (with unitary access to $\mathcal{O}$), such that for all x with $|x| = n$:

- (Completeness) If $x \in A_{\text{YES}}$, then $\Pr_z[\exists y : M^{\mathcal{O}}(x, y, z) \text{ accepts}] \geq 2/3$;
- (Soundness) If $x \in A_{\text{NO}}$, then $\Pr_z[\forall y : M^{\mathcal{O}}(x, y, z) \text{ rejects}] \geq 2/3$,

where $z \in \{0,1\}^{p(n)}$ is the (measured) output of the quantum algorithm $\mathcal{A}^{\mathcal{O}}(x)$. Given such an x and the string z produced by $\mathcal{A}^{\mathcal{O}}$, the question of whether there exists a y such that $M^{\mathcal{O}}(x, y, z)$ accepts can be decided by an $\mathsf{NP}^{\mathcal{O}}$ oracle. Since $\mathsf{BQP}^{\mathsf{NP}^{\mathcal{O}}}$ has access to both quantum computation and the oracle $\mathsf{NP}^{\mathcal{O}}$ (and thus also to $\mathcal{O}$), it can run the quantum reduction $\mathcal{A}^{\mathcal{O}}$, obtain z, and use its $\mathsf{NP}^{\mathcal{O}}$ oracle to verify the existential condition. Hence, $\mathsf{BQP}^{\mathsf{NP}^{\mathcal{O}}}$ can simulate the $(\mathsf{BQ} \cdot \mathsf{NP})^{\mathcal{O}}$ reduction, and overall succeeds with probability at least $2/3$. □

These relativising inclusions directly imply our desired oracle separation.

7.5.5. Theorem. *For any $q = \mathcal{O}(1)$, there exists an oracle $\mathcal{O}$ relative to which*

$$\mathsf{QCPCP}[q]^{\mathcal{O}} \subseteq \mathsf{QCPCP}_Q[q]^{\mathcal{O}} \subsetneq \mathsf{QCMA}^{\mathcal{O}}.$$

Proof:
This follows from Proposition 7.5.3 and Lemma 7.5.4. The latter shows that for all constant q and all oracles $\mathcal{O}$, we have $\mathsf{QCPCP}[q]^{\mathcal{O}} \subseteq \mathsf{QCPCP}_Q[q]^{\mathcal{O}} \subseteq \mathsf{BQP}^{\mathsf{NP}^{\mathcal{O}}}$, while Proposition 7.5.3 guarantees the existence of an oracle $\mathcal{O}$ such that $\mathsf{QCMA}^{\mathcal{O}} \not\subseteq \mathsf{BQP}^{\mathsf{NP}^{\mathcal{O}}}$. Finally, the inclusion $\mathsf{QCPCP}_Q[q]^{\mathcal{O}} \subseteq \mathsf{QCMA}^{\mathcal{O}}$ holds trivially, since the verifier in $\mathsf{QCMA}^{\mathcal{O}}$ has access to the full classical proof $y \in \{0,1\}^{\text{poly}(n)}$, and can efficiently implement the unitary U_y required by the QCPCP_Q verifier. □

7.5.3 Oracle separations for logarithmic queries

Whilst our results in Section 7.4 indicate that in the constant query regime, quantum queries offer no advantage over classical queries (even in the presence of external oracles, see Remark 7.4.11), we will now prove that this changes when the number of queries is (poly-)logarithmic relative to an oracle.

We will rely on the following lemma, which shows that the Bernstein–Vazirani algorithm [BV93] can be used to decode $\mathcal{O}(\log n)$ bits from a polynomially-sized classical string using only a single quantum query.

7.5.6. Lemma. *Given any $\mathcal{O}(\log n)$-bit string x, there exists a polynomially-sized classical proof y such that a quantum algorithm can recover x with certainty using only a single quantum query to y.*

Proof:
This lemma follows from the Bernstein-Vazirani algorithm [BV93]. The algorithm can learn a secret bit string x of length $l = \mathcal{O}(\log n)$, provided it has quantum oracle access to a function $f : \{0,1\}^l \to \{0,1\}$ defined by $f(z) = z \cdot x \mod 2$. For any such $x \in \{0,1\}^l$, the prover constructs a function f satisfying this property, and sends a proof $y = (y_1, \ldots, y_{2^l})$, where $y_{\bar{z}} = f(z)$ and $\bar{z}$ is the integer representation of the string z. Since $l \in \mathcal{O}(\log n)$, we have $|y| = 2^l = \text{poly}(n)$. Since a phase query can be implemented using one standard oracle query, the verifier can extract x by following the Bernstein–Vazirani algorithm, making only a single quantum query to y. □

Next, we state the lower bound lemma that lower bounds the query complexity of the OR-function, given access to some additional classical bits to assist in the verification.

7.5.7. Lemma. *Suppose we are given oracle access to an n-qubit phase oracle O_f, and want to decide which of the following holds:*

(i) There exists an n-bit string x^* such that $O_f |x^*\rangle = -|x^*\rangle$,

(ii) $O_f |x\rangle = |x\rangle$ for all x.

Then, even given q-bit bit string that helps identify the marked input x^* in Case (i), any quantum algorithm still requires

$$\Omega\left(\sqrt{\frac{2^n}{2^q}}\right)$$

queries to decide between Case (i) and (ii) with bounded probability of error.

Proof:
Let $C_1 > 0$ be a constant to be determined later. Suppose that there exists a T-query quantum algorithm with $T < C_1\sqrt{2^n/2^q}$ that uses a q-qubit computational basis state $|y\rangle$ as extra input (this can be viewed as a proof) and $m = \text{poly}(n)$ ancilla qubits initialized to $|0^m\rangle$, and that can distinguish between the two cases with success probability at least $2/3$. We will derive a contradiction to the known $\Omega(\sqrt{2^n})$ quantum lower bound for the OR-function [BBC$^+$01, Amb02]. Without loss of generality, let the final state of the algorithm be:

$$|\Psi\rangle = U_T O_c U_{T-1} O_c \dots O_c U_1 O_c U_0 |y\rangle |0^m\rangle ,$$

where O_c denotes controlled applications of O_f (tensored with identities), and the U_i are unitaries independent of O_f.

Let $\left|\Psi_{(\text{i})}\right\rangle$ and $\left|\Psi_{(\text{ii})}\right\rangle$ denote the final states in cases (i) and (ii), respectively. To achieve success probability $\geq 2/3$, the trace distance must satisfy:

$$\| \left|\Psi_{(\text{i})}\right\rangle\!\left\langle\Psi_{(\text{i})}\right| - \left|\Psi_{(\text{ii})}\right\rangle\!\left\langle\Psi_{(\text{ii})}\right| \|_1 \geq \frac{1}{3},$$

which implies $|\left\langle\Psi_{(\text{i})}\middle|\Psi_{(\text{ii})}\right\rangle|^2 \leq 8/9$. Now consider uncomputing the unitaries: define

$$|\Psi'\rangle = U_0^\dagger O_c^\dagger \cdots O_c^\dagger U_T^\dagger |\Psi\rangle .$$

In Case (ii), where $O_f = \mathbb{I}$, the oracle calls do nothing, so we have $\left|\Psi'_{(\text{ii})}\right\rangle = |y\rangle |0^m\rangle$. Since inner products are preserved under unitaries,

$$|\left\langle\Psi'_{(\text{i})}\middle|\Psi'_{(\text{ii})}\right\rangle|^2 \leq 8/9.$$

Measuring $|\Psi'\rangle$ in the computational basis using the projector $M = \{|y\rangle\!\langle y| \otimes |0^m\rangle\!\langle 0^m| , \mathbb{I} - |y\rangle\!\langle y| \otimes |0^m\rangle\!\langle 0^m|\}$ yields outcome $|y\rangle |0^m\rangle$ with probability 1 in Case (ii), and at most $8/9$ in Case (i). Now, suppose we guess a $y \in \{0,1\}^q$ uniformly at random instead. In Case (ii), we always observe $|y\rangle |0^m\rangle$. In Case (i), there's a $1/2^q$ chance our guess hits the "helpful" basis state, in which case we observe $|y\rangle |0^m\rangle$ with probability at most $8/9$. Using amplitude amplification [BHMT02], we can amplify this difference to obtain success probability $\geq 2/3$, at the cost of a factor of $C_2\sqrt{2^q}$ more queries, for some constant $C_2 > 0$. Thus, we would have a quantum algorithm computing the OR-function using $C_1 C_2 \sqrt{2^n}$ queries, contradicting the lower bound for small enough C_1. Hence, any such quantum algorithm must use $\Omega(\sqrt{2^n/2^q})$ queries. □

The intuition behind why the lower bound from Lemma 7.5.7 can be used in a PCP setting is that it allows us to "fix" the dishonest prover's strategy in Case (ii) to the same witness that would be optimal in Case (i). That is, the strategy used in Case (i), which leads to acceptance, can be reused in Case (ii), where it may no longer be optimal. However, this does not affect the validity

of the lower bound, since in the soundness case the prover is free to provide any proof, and a suboptimal strategy is still a valid one. Consequently, the lower bound applies even in the PCP setting, where the prover attempts to maximise the verifier's acceptance probability. Using this idea, we can prove our desired oracle separation using the lower bound of Lemma 7.5.7.

7.5.8. THEOREM. *For any $c \in \mathbb{N}$, there exists an oracle $\boldsymbol{\mathcal{O}} = \{O_n : n \in \mathbb{Z}_+\}$ such that*

$$\mathsf{QCPCP}[\mathcal{O}(\log^c n)]^{\boldsymbol{\mathcal{O}}} \subsetneq \mathsf{QCPCP}_Q[\mathcal{O}(\log^c n)]^{\boldsymbol{\mathcal{O}}}.$$

Proof:
Let L be a unary language, and let $p(n)$ be some polynomial. Define the oracle $\boldsymbol{\mathcal{O}} = \{O_n : n \in \mathbb{Z}_+\}$ as follows:

- If $0^n \in L$, then O_n contains a single string x of length $\log^{c+1} n$.
- If $0^n \notin L$, then O_n contains no string x of length $\log^{c+1} n$.

To show that $L \in \mathsf{QCPCP}_Q[\mathcal{O}(\log^c n)]^{\boldsymbol{\mathcal{O}}}$, consider the following protocol: the prover sends a classical proof $y = y^{(1)} \dots y^{(k)}$ composed of $k = \Theta(\log^c n)$ substrings $y^{(j)}$, each of size $\mathrm{poly}(n)$. By Lemma 7.5.6, there exists a y such that each $y^{(j)}$ encodes the function values of a function $f_{x^{(j)}}$ for some string $x^{(j)}$ of length $\Theta(\log n)$, such that $x = x^{(1)}x^{(2)} \dots x^{(k)}$. Using the algorithm from Lemma 7.5.6, the verifier can learn each $x^{(j)}$ with a single quantum query to $y^{(j)}$, reconstruct x, and then make a single query to the oracle to verify whether $x \in O_n$ (i.e., whether $0^n \in L$). This uses only $\mathcal{O}(\log^c n)$ quantum queries.

We will show that $L \notin \mathsf{QCPCP}[\mathcal{O}(\log^c n)]^{\boldsymbol{\mathcal{O}}}$ using the lower bound of Lemma 7.5.7. Fix some n, and let $O_\emptyset$ denote the empty oracle, and O_x the oracle containing the hidden string x. Let $C > 0$ be any constant. From Section 7.4.4, it is known that the verifiers of $\mathsf{QCPCP}[C \log^c n]^{\boldsymbol{\mathcal{O}}}$ can be assumed to be non-adaptive and sample indices from a distribution independent of the oracle (queries to O_n are only made after accessing the proof). Hence, we have

$$\Pr\left[V^{O_\emptyset} \text{ queries } y_{i_1}, \dots, y_{i_q}\right] = \Pr\left[V^{O_x} \text{ queries } y_{i_1}, \dots, y_{i_q}\right]$$

for all index tuples $(i_1, \dots, i_q)$. The verifier's acceptance probability in both cases is given by

$$\mathbb{E}_{i_1,\dots,i_q}\left[\Pr\left[V^O \text{ accepts querying } y_{i_1}, \dots, y_{i_q}\right]\right],$$

where $O \in \{O_\emptyset, O_x\}$. Denote D for a distribution over index tuples $(i_1, \dots, i_q)$ used by the verifier to choose which proof bits to query. Suppose for contradiction that $L \in \mathsf{QCPCP}[C \log^c n]^{\boldsymbol{\mathcal{O}}}$, with completeness c and soundness s, where $c - s = \Omega(1)$. Then there must exist a verifier and a distribution D such that:

$$\max_y \mathbb{E}_{(i_1,\dots,i_q)\sim D}\left[\Pr\left[V^{O_x} \text{ accepts querying } y_{i_1}, \dots, y_{i_q}\right]\right] \geq c,$$

and

$$\max_z \mathbb{E}_{(i_1,\ldots,i_q)\sim D}\left[\Pr\left[V^{O_\emptyset} \text{ accepts querying } z_{i_1},\ldots,z_{i_q}\right]\right] \leq s.$$

Let y be the proof string that achieves the maximum in the first expression. Then,

$$\mathbb{E}_{(i_1,\ldots,i_q)\sim D}\left[\begin{array}{l}\Pr\left[V^{O_x} \text{ accepts querying } y_{i_1},\ldots,y_{i_q}\right] \\ \quad -\Pr\left[V^{O_\emptyset} \text{ accepts querying } y_{i_1},\ldots,y_{i_q}\right]\end{array}\right] \geq c-s=\Omega(1).$$

This is upper bounded by

$$\max_{(i_1,\ldots,i_q)}\left(\Pr\left[V^{O_x} \text{ accepts } \mid y_{i_1},\ldots,y_{i_q}\right]-\Pr\left[V^{O_\emptyset} \text{ accepts } \mid y_{i_1},\ldots,y_{i_q}\right]\right).$$

This implies that there exists a q-bit string $y^*_{i_1},\ldots,y^*_{i_q}$ that detects if the oracle O contains a string in $C\log^c n$ queries. However, for any such $C>0$,

$$C\log^c n = o\left(\sqrt{\frac{2^{\log^{c+1} n}}{2^{\log^c n}}}\right),$$

so by Lemma 7.5.7 the maximum distinguishing bias is $o(1)$, which contradicts $c-s=\Omega(1)$. Therefore, $L\notin \mathsf{QCPCP}[\log^c n]^{\mathcal{O}}$, as claimed.

□

Part III.

Unitary query and sample-to-sample complexity

Chapter 8

Lower bounds for unitary query complexity

8.1 Introduction

Quantum query complexity is the study of how many queries a quantum algorithm has to make to some black-box input X to decide whether X satisfies some property $\mathcal{P}$. While quantum query complexity conventionally focuses on X being inherently classical (i.e., a classical bit string), it is also possible to consider the setting where X is a black-box unitary. The former—known as *property testing*—is very useful for obtaining insights into the differences in computational power between various computational classes, classical or quantum. The latter—*unitary property testing*—provides another way to compare inherently quantum classes. These problems, first studied by Wang [Wan11], have gained considerable attention recently [SY23, CNY03, WZ23].

Query complexities can vastly differ among different classes of computational models. For example, the search problem, which is to decide whether a string of length N is either the all-zeros string or has at least one entry with a "1", is known to have classical query complexity of $\Theta(N)$ and quantum query complexity of $\Theta(\sqrt{N})$ [Gro96, BBBV97]. However, given a string by an untrusted prover, the query complexity of the search problem is just 1 in both cases. A similar result holds for a unitary property testing analogue of search as introduced by Aaronson and Kuperberg [AK07], where one has to decide whether a given black-box unitary U applies either the identity operation $\mathbb{I}$ or the reflection $\mathbb{I} - 2\left|\psi\right\rangle\!\left\langle\psi\right|$ for some unknown N-dimensional quantum state $\left|\psi\right\rangle$. This problem generally has a quantum query complexity of $\Theta(\sqrt{N})$, but can again be solved by just a single query if a *quantum state* is provided by an untrusted prover as an extra input. For many other unitary property testing problems, it is unclear whether quantum proofs and/or trusted advice states might help in solving these tasks.

8.1.1 Results of this chapter

In this chapter, we reduce unitary property testing (and related unitary problems) to the problem of *unitary channel discrimination* from quantum information theory. In particular, our lower bound technique adopts the following strategy:

1. Given a unitary property $\mathcal{P} = (\mathcal{P}_{\text{yes}}, \mathcal{P}_{\text{no}})$, find two unitaries U_1, U_2 such that $U_1 \in \mathcal{P}_{\text{yes}}$ and $U_2 \in \mathcal{P}_{\text{no}}$.

2. Show that a unitary property tester for $\mathcal{P}$ implies a distinguisher for U_1 and U_2.

3. Prove a lower bound on the channel discrimination complexity of U_1 and U_2, which implies a lower bound on the query complexity of the unitary property tester.

Since this is an information-theoretic lower bound technique, it is easy to show that the technique also applies in the setting where one is given access to quantum proofs and/or advice states.

As we will show throughout the chapter, this strategy turns out to be a powerful procedure for obtaining simple, yet often tight, lower bounds for a wide range of unitary property testing and related problems. Table 8.1 summarises the obtained lower bounds for the examples we consider.

Problem	Query lower bound
Quantum phase estimation	$\Omega\left(1/\epsilon\right)$
Entanglement entropy	$\Omega\left(1/\sqrt{\Delta}\right)$
Subset support verification	$\Omega\left(\sqrt{\lvert S\rvert}\right)$
Quantum amplitude estimation	$\Omega\left(1/\epsilon\right)$
Thermal state preparation	$\Omega(\beta)$
Hamiltonian Simulation	$\Omega(t)$
Hamiltonian learning	$\Omega(1/\epsilon)$
GSP of gapped Hamiltonians	$\Omega(1/\Delta)$

Table 8.1: Obtained bounds for the query complexity of unitary property testing and related problems. All bounds hold for any C-tester with $\mathsf{C} \subseteq \mathsf{QMA}(\text{poly}(n))/\mathsf{qpoly}$ and are, with the exception of the entanglement entropy problem, shown to be tight (up to logarithmic factors). All bounds, except for the subset support verification and entanglement entropy problem, where known in previous works. However, as far as the author is aware, none of these were also shown to hold in the presence of quantum proof and advice.

To highlight its simplicity, consider the lower bound for quantum phase estimation. The first lower bound proof, as proven by Bessen in [Bes05], relies on

frequency analysis and spans a 7-*page* double-column paper. Using our technique, an optimal lower bound in a stronger setting (including proofs and advice) can be shown with a proof of just 7 *lines*, once the technique has been set up.

One of the main takeaways from the results in Table 8.1 is that whenever high precision is required in a black-box setting, neither quantum proofs nor advice seems to help in reducing the required query complexity. Essentially, we will see that this establishes in a unitary query setting what is known as *Heisenberg-limited scaling* (or the *Heisenberg limit*) in quantum sensing and metrology [HB93, Ou96]. This limit implies that achieving a $1/N$-factor improvement in accuracy requires $\sim N$ additional "resources", which, in our case, correspond to queries. Since the Heisenberg limit is an information-theoretic notion, it should apply universally, including to computational settings that allow (quantum) proofs and advice.

Since our bounds hold in such a strong setting, we can also use them to prove quantum oracle separations. The power of the complexity class $\mathsf{QMA}(2)$, which is a generalisation of QMA where there are two non-interacting provers, is one of the major open questions in quantum complexity theory. The best upper bound we currently have is $\mathsf{QMA}(2) \subseteq \mathsf{NEXP}$, which follows by simply guessing exponential-size classical descriptions of the two quantum proofs. Our results can be lifted to yield a quantum oracle relative to which $\mathsf{QMA}(2)$ does not even contain SBQP, a variant of BQP with an exponentially small promise gap. Hence, if $\mathsf{QMA}(2)$ were able to solve highly precise problems—such as the local Hamiltonian problem at exponentially precise precision, which is PSPACE-complete [FL16]—it would need to do so in a way that does not work in a quantum black-box setting. Similarly, we can use the same idea to show a quantum oracle separation relative to which $\mathsf{QMA/qpoly}$ does not contain SBQP.

8.2 Classes of unitary property testers

We follow the conventions from the work of She and Yuen [SY23]. For a fixed number of qubits n, let $\mathcal{P}_{\text{YES}}$ and $\mathcal{P}_{\text{NO}}$ (called the yes and no instances, respectively) be *disjoint* subsets of all n-qubit unitary operators.[1] A tester for a unitary problem $\mathcal{P} = (\mathcal{P}_{\text{YES}}, \mathcal{P}_{\text{NO}})$ is a quantum algorithm that, given query access to $U \in \mathcal{P}_{\text{YES}}$ or $U \in \mathcal{P}_{\text{NO}}$, accepts with high (resp. low) probability when $U \in \mathcal{P}_{\text{YES}}$ (resp. $U \in \mathcal{P}_{\text{NO}}$). The queries are also allowed to be controlled or to access the inverse. This is more general than some other definitions used in the literature, where the NO-instances are defined to be ϵ-far from the set of yes instances in terms of some distance measure (see, for example, [MdW13]). However, for our purposes, we only require the sets $\mathcal{P}_{\text{YES}}$ and $\mathcal{P}_{\text{NO}}$ to be disjoint.

[1] More generally, one would consider all d-dimensional unitaries for an arbitrary dimension d, but this would not affect any of our results.

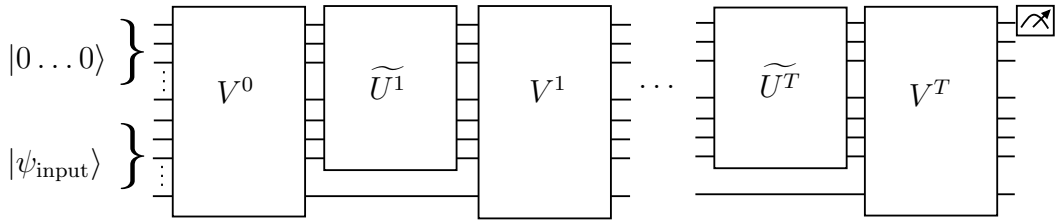


Figure 8.1: Query complexity model for a unitary property tester making T queries to a unitary of interest U. The initial state $|\psi_{\text{init}}\rangle = |0\ldots0\rangle\,|\psi_{\text{input}}\rangle$ is allowed to consist of an input-independent part and an input-dependent part, depending on the class C of the C-tester. The unitary of interest U can be accessed directly, through its inverse, controlled or controlled inverse, i.e. we have that $\tilde{U}^t \in \{U_i, U_i^\dagger, cU_i, cU_i^\dagger\}$ for all $t \in [T]$, $i \in \{1, 2\}$.

We can now construct classes of testers similar to those in [SY23]. First, let us start with the simplest class of testers, which do not use proofs or advice. Without loss of generality, one can always consider the input states to be pure, since the acceptance probability of any mixed state is always upper bounded by that of a pure state via a convexity argument. The initial state of any quantum algorithm is assumed to have some number of ancilla qubits, all initialised in $|0\rangle$, and may sometimes be supplied with an additional input state, depending on the class of tester considered.

8.2.1. Definition (BQP-tester, from [SY23]). Let $\mathcal{P} = (\mathcal{P}_{\text{YES}}, \mathcal{P}_{\text{NO}})$ be a unitary property on n qubits. We say $\mathcal{P}$ has a BQP-tester if there exists a quantum algorithm such that the following holds:

- If $U \in \mathcal{P}_{\text{YES}}$, the quantum algorithm makes queries to U and accepts with probability $\geq 2/3$.
- If $U \in \mathcal{P}_{\text{NO}}$, the quantum algorithm makes queries to U and accepts with probability $\leq 1/3$.

Note that there is no restriction on the number of queries the tester makes to U, as this is the quantity being characterised. Nor is there any restriction on the amount of space or time used by the BQP-tester in this definition, which makes the "P" in "BQP-tester" somewhat awkward. Nonetheless, we adopt this notation to follow the convention in [SY23] and to facilitate a direct connection to separations between actual complexity classes (see for example, Section 8.5).

Let us now add, possibly unentangled, quantum (resp. classical) proofs to define QMA (resp. QCMA) testers. Again, by the same argument made for the BQP-tester, we do not have to bound the allowed size of either the proof or the advice states.

8.2.2. Definition (QMA(k)-tester). Let $\mathcal{P} = (\mathcal{P}_{\text{YES}}, \mathcal{P}_{\text{NO}})$ be a unitary property on n qubits. We say $\mathcal{P}$ has a QMA(k)-tester if the following holds:

- If $U \in \mathcal{P}_{\text{YES}}$, then there exists k quantum states $\{|\xi_i\rangle\}_{i\in[k]}$ such that on input $|\xi_1\rangle \dots |\xi_k\rangle$ the algorithm makes queries to U and accepts with probability $\geq 2/3$.

- If $U \in \mathcal{P}_{\text{NO}}$, then for all k quantum states $\{|\xi_i\rangle\}_{i\in[k]}$, the algorithm acting on $|\xi_1\rangle \dots |\xi_k\rangle$ makes queries to U and accepts with probability $\leq 1/3$.

If $k = 1$, we abbreviate to a QMA-tester.

Since we do not assume any restriction on the proof sizes, we cannot use the result of Harrow and Montanaro [HM13] to let QMA(2)-testers cover all QMA(poly(n))-testers.

The definitions of testers of unitary properties in [SY23] do not include classes that allow for advice. A technical difficulty arises when one wants to include advice states, as it becomes necessary to specify a notion of input length on which the advice may depend. In [SY23], no such restrictions were necessary, and all classes of testers were defined in a way that did not depend on any notion of input size n. We will make the choice that n is given by the number of qubits the unitary acts on.

Note that, in cases where the property is also parametrised by some parameter, this is not restrictive in the advice setting, as each parameter setting defines a different unitary property and hence is allowed a different advice string. To clarify this point, take the example of a unitary property testing formulation of quantum phase estimation with a variable number of qubits n but a fixed sequence of known eigenstates $|\psi_n\rangle$ (e.g., $|0^n\rangle$). Here, the goal is to determine whether an unknown n-qubit unitary U belongs to $\mathcal{P}_{\text{YES}} = \{U : U|0^n\rangle = |0^n\rangle\}$ or $\mathcal{P}_{\text{NO}} = \{U : U|0^n\rangle = e^{2\pi i\theta}|0^n\rangle, \epsilon \leq \theta \leq 1/2\}$ for some fixed $\epsilon > 0$. In other words, given the promise that U has an eigenstate $|\psi\rangle$, the task is to determine if its eigenphase is 0 or $\geq \epsilon$.

In this case, ϵ could also be parametrised as a function of some m, i.e., $\epsilon = \epsilon(m)$. For a fixed choice of ϵ, the advice should be identical for each fixed n. However, even for fixed values of n, the advice can vary for different values of ϵ, as each constitutes a new property testing problem.

Having set our convention of what the input size n represents, we will now state our definitions of the testers with advice (again with the "poly"-part being redundant).

8.2.3. Definition (BQP/qpoly-tester). Let $\mathcal{P} = (\mathcal{P}_{\text{YES}}, \mathcal{P}_{\text{NO}})$ be a unitary property on n qubits. We say $\mathcal{P}$ has a BQP/qpoly-tester if there exists a quantum advice state $|\psi_n\rangle$, and a quantum algorithm such that the following holds:

- If $U \in \mathcal{P}_{\text{YES}}$, on input $|\psi_n\rangle$ the quantum algorithm makes queries to U and accepts with probability $\geq 2/3$.
- If $U \in \mathcal{P}_{\text{NO}}$, on input $|\psi_n\rangle$ the quantum algorithm makes queries to U and accepts with probability $\leq 1/3$.

We can also combine proofs and advice to arrive at even more powerful classes of testers.

8.2.4. Definition (QMA(k)/qpoly-tester). Let $\mathcal{P} = (\mathcal{P}_{\text{YES}}, \mathcal{P}_{\text{NO}})$ be a unitary property on n qubits. We say $\mathcal{P}$ has a QMA(k)/qpoly-tester if there exists a quantum advice state $|\psi_n\rangle$, and a quantum algorithm such that the following holds:

- If $U \in \mathcal{P}_{\text{YES}}$, then there exist k quantum states $\{|\xi_i\rangle\}_{i\in[k]}$ such that, on input $|\psi_n\rangle |\xi_1\rangle \dots |\xi_k\rangle$, the algorithm makes queries to U and accepts with probability $\geq 2/3$.
- If $U \in \mathcal{P}_{\text{NO}}$, then for all k quantum states $\{|\xi_i\rangle\}_{i\in[k]}$, the algorithm acting on $|\psi_n\rangle |\xi_1\rangle \dots |\xi_k\rangle$ makes queries to U and accepts with probability $\leq 1/3$.

There is one more class we would like to introduce, which is fundamentally different from all the classes of testers discussed so far in the sense that it allows for an exponentially small gap between the completeness and soundness parameters.

8.2.5. Definition (SBQP-tester). Let $\mathcal{P} = (\mathcal{P}_{\text{YES}}, \mathcal{P}_{\text{NO}})$ be a unitary property on n qubits. We say $\mathcal{P}$ has an SBQP-tester if there exists a quantum algorithm and a polynomial $p(n)$ such that the following holds:

- If $U \in \mathcal{P}_{\text{YES}}$, the quantum algorithm makes queries to U and accepts with probability $\geq 2^{-p(n)}$.
- If $U \in \mathcal{P}_{\text{NO}}$, the quantum algorithm makes queries to U and accepts with probability $\leq 2^{-p(n)-1}$.

8.3 Lower bounds by unitary channel discrimination

In this section, we will prove Theorem 8.3.7. First, we will set up some more preliminaries regarding unitary channel discrimination. For unitary channels, the definition of the diamond norm allows for a more easily computable expression, as shown in the following lemma.

8.3.1. Lemma (Adapted from [ZS10]). *Let $U_1, U_2 \in \mathbb{U}(d)$, with the property that $0 \notin \mathrm{conv}(\mathrm{eig}(U_1^\dagger U_2))$. Then*

$$\frac{1}{2}\left\|\mathcal{U}(U_1) - \mathcal{U}(U_2)\right\|_\diamond = \sqrt{1 - D^2}, \tag{8.1}$$

with

$$D = \frac{1}{2}\min_{k,l}\left|e^{i\theta_k} + e^{i\theta_l}\right| \tag{8.2}$$

where $e^{i\theta_k}, e^{i\theta_l} \in \mathrm{eig}(U_1^\dagger U_2)$.

The quantity D has a nice geometrical interpretation: if $0 \notin \mathrm{conv}(\mathrm{eig}(U_1^\dagger U_2))$, then it is precisely the distance between the convex hull of the eigenvalues of $U_1^\dagger U_2$ and the origin. To determine whether $0 \notin \mathrm{conv}(\mathrm{eig}(U_1^\dagger U_2))$, it is useful to consider the *spectral arc-length* of U, given by the following definition.

8.3.2. Definition (Spectral arc-length). The spectral arc-length $\Theta(U)$ of a unitary $U \in \mathbb{U}(d)$ is defined as the length of the shortest interval $I \subseteq [0, 2\pi)$ such that the corresponding arc e^{iI} contains the spectrum of U.

It is easy to show that $0 \notin \mathrm{conv}(\mathrm{eig}(U)$ if and only if $\Theta(U) < \pi$ [Wol23] (see also Fig. 8.2).

There are more ways to characterise the diamond distance for unitaries, which might be more convenient for some choices of unitaries as we try to prove lower bounds later down the road. Since we are interested in the asymptotic scaling of our lower bounds, it will sometimes be convenient to use the (global-phase shifted) operator norm difference between the two unitaries, as it is equivalent to the diamond norm up to a (inverse) factor of (at most) two.

8.3.3. Lemma (Adapted from [HKOT23]). *Let $U_1, U_2 \in \mathbb{U}(d)$ be unitaries. Then we have that*

$$\frac{1}{2}\left\|\mathcal{U}(U_1) - \mathcal{U}(U_2)\right\|_\diamond \leq \min_\phi \left\|e^{i\phi}U_1 - U_2\right\| \leq \left\|\mathcal{U}(U_1) - \mathcal{U}(U_2)\right\|_\diamond.$$

As Lemma 8.3.3 involves a minimisation over a global phase factor $e^{i\phi}$, using the vanilla operator norm distance between two unitaries U_1 and U_2 would also yield an upper bound to the left side of the inequality. Moreover, one can also show that a sufficient upper bound in operator norm implies the condition that $0 \notin \mathrm{conv}(\mathrm{eig}(U)$ is immediately satisfied, as illustrated by the following lemma.

8.3.4. Lemma. *Let $U_1, U_2 \in \mathbb{U}(d)$ be unitaries for which $\|U_1 - U_2\| \leq 2\sin\left(\frac{\pi-\delta}{4}\right)$ for some small $0 < \delta < \pi$. Then it holds that $0 \notin \mathrm{conv}(\mathrm{eig}(U_1^\dagger U_2))$. For $\delta = 0.04$, we have $2\sin\left(\frac{\pi-\delta}{4}\right) = 1.4$.*

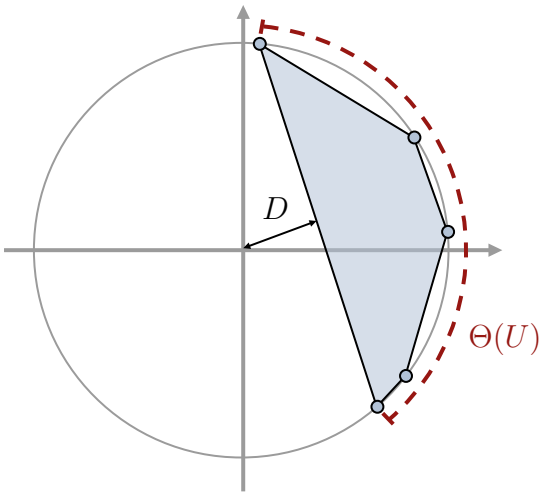


Figure 8.2: Geometrical interpretation in the complex plane of the spectrum of some $U \in \mathbb{U}(5)$. Since U is unitary, all its eigenvalues (blue dots) lie on the unit circle in the complex plane. Since the spectral arc-length (indicated by the red dashed line) satisfies $\Theta(U) < \pi$, we have that 0 is not in the convex hull of the eigenvalues of U (blue shaded area). The shortest distance from the origin to the convex hull of the spectrum of U is D.

Proof:
By unitary invariance of the operator norm, it must hold that

$$\|U_1 - U_2\| = \left\|\mathbb{I} - U_1^\dagger U_2\right\| = \max_{\theta_j : e^{i\theta_j} \in \mathrm{eig}(U_1^\dagger U_2)} \left|1 - e^{i\theta_j}\right|.$$

Since all $e^{i\theta_j}$ are points on the unit circle in the complex plane, we can assume without loss of generality that $\theta_j \in [0, 2\pi]$ for all $j \in [d]$. Thus, we get the condition

$$\max_{\theta_j : e^{i\theta_j} \in \mathrm{eig}(U_1^\dagger U_2)} 2\left|\sin\left(\frac{\theta_j}{2}\right)\right| \leq 2\sin\left(\frac{\pi - \delta}{4}\right),$$

which means that $0 \leq \theta_j \leq (\pi - \delta)/2$ or $2\pi - (\pi - \delta)/2 \leq \theta_j \leq 2\pi$ for all $j \in [d]$. Hence, all eigenvalues of $U_1^\dagger U_2$ lie within the arc e^{iI} on the unit circle, where $I = [-(\pi - \delta)/2, (\pi - \delta)/2]$ is a closed interval of length $\pi - \delta$. Since e^{iI} is a closed arc of the unit circle with length strictly less than π, it follows that $0 \notin \mathrm{conv}(\mathrm{eig}(U_1^\dagger U_2))$. Therefore, Definition 8.3.2 combined with the assumption $0 < \delta < \pi$ implies that $\Theta(U_1^\dagger U_2) < \pi$, which means that $0 \notin \mathrm{conv}(\mathrm{eig}(U_1^\dagger U_2))$. □

Now that we have established ways to evaluate the diamond distance between two unitaries, we will proceed by showing that adding a well-chosen global phase to one of the unitaries ensures that access to the inverse, controlled access, or a

combination of both, does not increase the ability to discriminate between the two. To show this, first observe that the condition of 0 not being in the convex hull of the eigenvalues of $U_1^\dagger U_2$ is invariant to adding a global phase to one of the unitaries.

8.3.5. Lemma. *Let $U_1, U_2 \in \mathbb{U}(d)$ such that $0 \notin \mathrm{conv}(\mathrm{eig}(U_1^\dagger U_2))$. Let $U_2^\theta = e^{i\theta}U_2$. Then for all $\theta \in [0, 2\pi)$ we have $0 \notin \mathrm{conv}(\mathrm{eig}(U_1^\dagger U_2^\theta))$.*

Proof:
This follows directly from the invariance of D under rotations, since $\mathrm{eig}(U_1^\dagger U_2^\theta) = \{e^{i\theta}e^{i\theta_l} | e^{i\theta_l} \in \mathrm{eig}(U_1^\dagger U_2)\}$ and

$$\frac{1}{2}\min_{k,l}\left|e^{i\theta}e^{i\theta_k} + e^{i\theta}e^{i\theta_l}\right| = \frac{1}{2}\min_{k,l}\left|e^{i\theta}\right|\left|e^{i\theta_k} + e^{i\theta_l}\right| = \frac{1}{2}\min_{k,l}\left|e^{i\theta_k} + e^{i\theta_l}\right|$$

for any $\theta \in [0, 2\pi)$. □

The next lemma proves the aforementioned claim that given some U_1, U_2 we can always pick a global phase such that the diamond distance between U_1 and U_2 precisely characterises the distance in the controlled, inverse, or controlled inverse setting.

8.3.6. Lemma (Diamond distance for different query types). *Let $U_1, U_2 \in \mathbb{U}(d)$ such that $0 \notin \mathrm{conv}(\mathrm{eig}(U_1^\dagger U_2))$, and let $U_2^\theta = e^{i\theta}U_2$ for some $\theta \in [0, 2\pi)$. Then there exists a choice of θ, such that for any combination of*

$$(\tilde{U}_1, \tilde{U}_2^\theta) \in \{(U_1, U_2^\theta), (U_1^\dagger, \left(U_2^\theta\right)^\dagger), (cU_1, cU_2^\theta), (cU_1^\dagger, \left(cU_2^\theta\right)^\dagger)\}$$

we have

$$\frac{1}{2}\left\|\mathcal{U}(\tilde{U}_1) - \mathcal{U}(\tilde{U}_2^\theta)\right\|_\diamond = \frac{1}{2}\left\|\mathcal{U}(U_1) - \mathcal{U}(U_2^\theta)\right\|_\diamond.$$

Proof:
By Lemma 8.3.5, we must have that $0 \notin \mathrm{conv}(\mathrm{eig}(U_1^\dagger U_2^\theta))$ for any $\theta \in [0, 2\pi)$. We consider the separate cases using Lemma 8.3.1. The (U_1, U_2^θ)-case is trivially true.

Inverse queries. By Theorem 1.3.22 in [HJ12] we have $\mathrm{eig}((U_1^\dagger)^\dagger \left(U_2^\theta\right)^\dagger) = \mathrm{eig}(\left(U_2^\theta\right)^\dagger (U_1^\dagger)^\dagger) = \mathrm{eig}((U_1^\dagger U_2^\theta)^\dagger)$. Therefore, for all $\theta \in [0, 2\pi)$ it holds that

$$\begin{aligned}
\frac{1}{2}\left\|\mathcal{U}(U_1^\dagger) - \mathcal{U}(\left(U_2^\theta\right)^\dagger)\right\|_\diamond &= \sqrt{1 - \left(\frac{1}{2}\min_{k,l}|e^{-i(\theta+\theta_k)} + e^{-i(\theta+\theta_l)}|\right)^2} \\
&= \sqrt{1 - \left(\frac{1}{2}\min_{k,l}|e^{i(\theta+\theta_k)} + e^{i(\theta+\theta_l)}|\right)^2} \\
&= \frac{1}{2}\left\|\mathcal{U}(U_1) - \mathcal{U}(U_2^\theta)\right\|_\diamond.
\end{aligned}$$

Controlled queries. Note that we have that

$$\mathrm{eig}(cU_1^\dagger cU_2^\theta) = \mathrm{eig}(|0\rangle\langle 0| \otimes \mathbb{I} + |1\rangle\langle 1| \otimes U_1^\dagger U_2^\theta) = \mathrm{eig}(U_1^\dagger U_2^\theta) \cup \{1\}.$$

For all U_1, U_2 there exists a θ such that $1 \in \mathrm{eig}(U_1^\dagger U_2^\theta)$, simply by taking any eigenvalue $e^{i\theta_l} \in \mathrm{eig}(U_1^\dagger U_2)$ and letting $\theta = 2\pi - \theta_l$. Hence, for this choice of θ we have

$$\frac{1}{2}\left\|\mathcal{U}(cU_1) - \mathcal{U}(cU_2^\theta)\right\|_\diamond = \frac{1}{2}\left\|\mathcal{U}(U_1) - \mathcal{U}(U_2^\theta)\right\|_\diamond.$$

Controlled queries to the inverse. This follows directly by combining the two arguments above, and holds for the same value of θ. □

We can now state the main theorem behind our lower bound technique and give the proof.

8.3.7. Theorem (Lower bound for unitary discrimination). *Let $\theta \in [0, 2\pi)$ and let $U_1, U_2 \in \mathbb{U}(d)$ such that $0 \notin \mathrm{conv}(\mathrm{eig}(U_1^\dagger U_2))$. Now let $U \in \{U_1, U_2^\theta\}$ with $U_2^\theta = e^{i\theta}U_2$ be a unitary to which one has black-box access, including controlled operations, applications of the inverse, and a combination of both. Suppose one has to decide whether (i) $U = U_1$ or (ii) $U = U_2^\theta$ holds, promised that either one of them is the case, and suppose $\frac{1}{2}\left\|\mathcal{U}(U_1) - \mathcal{U}(U_2)\right\|_\diamond \leq \epsilon$. Then there exists a $\theta \in [0, 2\pi)$ such that to decide with success probability $\geq 2/3$ whether (i) or (ii) holds, any C-tester with $\mathsf{C} \subseteq \mathsf{QMA}(\mathrm{poly}(n))/\mathsf{qpoly}$ needs to make at least*

$$T \geq \Omega\left(\frac{1}{\epsilon}\right)$$

queries to U.

Proof:
We omit the dependence of θ in subsequent notation and assume it is chosen according to Lemma 8.3.6. Without loss of generality, we can write the circuit $V^{U_i,(T)}$ which makes T queries to $\tilde{U}_i^{\,t} \in \{U_i, U_i^\dagger, cU_i, cU_i^\dagger, \}$, $i \in \{1, 2\}$, as

$$V^{U_i,(T)} = V_T \tilde{U}_i^T V_{T-1} \tilde{U}_i^{T-1} \ldots V_1 \tilde{U}_i^1 V_0.$$

We want to measure a designated output qubit of $V^{U_i,(T)}$ in the computational basis, which should output 1 if we have U_1 and output 0 if we have U_2.

By the operational interpretation of trace distance, we then have that for any input state $|\psi_{\mathrm{init}}\rangle$:

$$\left|\Pr\left[V^{U_1,(T)}(|\psi_{\mathrm{init}}\rangle) \text{ outputs } 1\right] - \Pr\left[V^{U_2,(T)}(|\psi_{\mathrm{init}}\rangle) \text{ outputs } 1\right]\right| \leq \ldots$$
$$\ldots \frac{1}{2}\left\|\mathcal{U}(V^{U_1,(T)}) - \mathcal{U}(V^{U_2,(T)})\right\|_\diamond.$$

Let us show by induction on the number of queries that

$$\left\|\mathcal{U}(V^{U_1,(T)}) - \mathcal{U}(V^{U_2,(T)})\right\|_\diamond \leq 2T\epsilon.$$

For a single query, we have

$$\begin{aligned}
\left\|\mathcal{U}(V^{U_1,(1)}) - \mathcal{U}(V^{U_2,(1)})\right\|_\diamond &= \left\|\mathcal{U}(V_1\tilde{U}_1^1 V_0) - \mathcal{U}(V_1\tilde{U}_2^1 V_0)\right\|_\diamond && (8.3)\\
&= \left\|\mathcal{U}(\tilde{U}_1^1) - \mathcal{U}(\tilde{U}_2^1)\right\|_\diamond && (8.4)\\
&= \left\|\mathcal{U}(U_1) - \mathcal{U}(U_2)\right\|_\diamond && (8.5)\\
&\leq 2\epsilon, && (8.6)
\end{aligned}$$

using the unitary invariance of the diamond distance (Proposition 2.3.7) in going from line Eq. (8.3) to line Eq. (8.4) and Lemma 8.3.6 from going to line Eq. (8.4) to Eq. (8.5). For T queries, assuming the induction hypothesis on $T-1$ queries, we find

$$\begin{aligned}
\left\|\mathcal{U}(V^{U_1,(T)}) - \mathcal{U}(V^{U_2,(T)})\right\|_\diamond &= \left\|\mathcal{U}(V_T\tilde{U}_1^T V^{U_1,(T-1)}) - \mathcal{U}(V_T\tilde{U}_2^T V^{U_2,(T-1)})\right\|_\diamond && (8.7)\\
&\leq \left\|\mathcal{U}(V_T\tilde{U}_1^T) - \mathcal{U}(V_T\tilde{U}_2^T)\right\|_\diamond && (8.8)\\
&\qquad + \left\|\mathcal{U}(V^{U_1,(T-1)}) - \mathcal{U}(V^{U_2,(T-1)})\right\|_\diamond &&\\
&\leq \left\|\mathcal{U}(V_T\tilde{U}_1^T) - \mathcal{U}(V_T\tilde{U}_2^T)\right\|_\diamond + 2(T-1)\epsilon && (8.9)\\
&= \left\|\mathcal{U}(\tilde{U}_1^T) - \mathcal{U}(\tilde{U}_2^T)\right\|_\diamond + 2(T-1)\epsilon && (8.10)\\
&= \left\|\mathcal{U}(U_1) - \mathcal{U}(U_2)\right\|_\diamond + 2(T-1)\epsilon && (8.11)\\
&\leq 2T\epsilon. && (8.12)
\end{aligned}$$

Here we used the hybrid argument from Proposition 2.3.7 in going from Eq. (8.7) to Eq. (8.8), the induction hypothesis in going from Eq. (8.8) to Eq. (8.9), the unitary invariance of the diamond norm from Eq. (8.9) to Eq. (8.10) and Lemma 8.3.6 from Eq. (8.10) to Eq. (8.11). Therefore, in order to have

$$\left|\Pr\left[V^{U_1,(T)}(|\psi_{\text{init}}\rangle) \text{ outputs } 1\right] - \Pr\left[V^{U_2,(T)}(|\psi_{\text{init}}\rangle) \text{ outputs } 1\right]\right| \geq \frac{2}{3} - \frac{1}{3} = \frac{1}{3},$$

we require

$$\frac{1}{3} \leq \frac{1}{3}\left\|\mathcal{U}(V^{U_1,(T)}) - \mathcal{U}(V^{U_2,(T)})\right\|_\diamond \leq T\epsilon,$$

which implies

$$T \geq \frac{1}{3\epsilon}. \tag{8.13}$$

We now have to show that Eq. (8.13) holds for $\mathsf{QMA}(\mathrm{poly}(n))/\mathsf{qpoly}$-testers. Let $\mathcal{P} = (\mathcal{P}_{\text{yes}}, \mathcal{P}_{\text{yes}})$, where $\mathcal{P}_{\text{yes}} = \{U_1^1, U_1^2, \dots\}$ and $\mathcal{P}_{\text{no}} = \{U_2^1, U_2^2, \dots\}$, such that for each $n \in \mathbb{Z}_+$ we have $\frac{1}{2}\|\mathcal{U}(U_1^n) - \mathcal{U}(U_2^n)\|_\diamond \leq \frac{1}{3T}$. Now suppose the above corollary is false, then there must exist an input quantum state $|\psi_{\text{input}}\rangle = |\psi_n\rangle |\xi_1\rangle \dots |\xi_k\rangle$ such that a T'-query quantum algorithm, $T' < T$, starting in the initial state $|\psi_{\text{init}}\rangle = |0 \dots 0\rangle |\psi_{\text{input}}\rangle$, which makes queries to U can decide with probability $> 2/3$ whether $U \in \mathcal{P}_{\text{yes}}$ or $U \in \mathcal{P}_{\text{no}}$ (in the no-case the soundness property holds for all states, so also the one that is used in the yes-case). This contradicts the bound of Eq. (8.13), which holds for any $|\psi_{\text{init}}\rangle$. □

8.3.8. Remark. Theorem 8.3.7 also holds when we replace the diamond norm condition on U_1 and U_2 by

$$\min_\phi \left\| e^{i\phi} U_1 - U_2 \right\| \leq \|U_1 - U_2\| \leq \epsilon, \tag{8.14}$$

since we show that

$$\left\| \mathcal{U}(V^{U_1}) - \mathcal{U}(V^{U_2}) \right\|_\diamond \leq T \|\mathcal{U}(U_1) - \mathcal{U}(U_2)\|_\diamond,$$

where

$$\frac{1}{2}\|\mathcal{U}(U_1) - \mathcal{U}(U_2)\|_\diamond \leq \min_\phi \left\| e^{i\phi} U_1 - U_2 \right\|,$$

by Lemma 8.3.3.

With Theorem 8.3.7 in hand, our proposed lower bound technique is now straightforward: given a unitary property $\mathcal{P} = (\mathcal{P}_{\text{yes}}, \mathcal{P}_{\text{no}})$, one picks two unitaries U_1, U_2 such that $U_1 \in \mathcal{P}_{\text{yes}}$ and $U_2 \in \mathcal{P}_{\text{no}}$; this implies that any unitary property tester for $\mathcal{P}$ implies a distinguisher for U_1 and U_2. One evaluates the diamond distance (or an upper bound thereof) between U_1 and U_2 using Lemmas 8.3.1 and 8.3.3 or any other suitable method, and verifies the property of $0 \notin \mathrm{conv}(\mathrm{eig}(U_1^\dagger U_2))$, possibly with the help of Lemma 8.3.5. The lower bound on the channel discrimination query complexity of U_1, U_2 follows then from Theorem 8.3.7, which directly implies a lower bound on the query complexity of the unitary property tester.

8.4 Applications

In this section, we will apply the lower bound method of Section 8.3 to some problems in unitary property testing as well as some other unitary problems, showcasing the fact that the technique is very simple to use and (for all but one problem) leads to optimal lower bounds.

8.4.1 Unitary property testing

Quantum phase estimation. In quantum phase estimation, one is given a unitary U and an eigenstate $|\psi\rangle$, and the task is to determine the eigenphase of U corresponding to $|\psi\rangle$ up to some precision ϵ with probability $\geq 2/3$. A lower bound can be obtained by reducing the quantum counting problem to the amplitude estimation problem, which is then reduced to the phase estimation problem. From this, the lower bound of $\Omega(1/\epsilon)$ follows from the lower bound for quantum counting given in [NW99]. However, our method allows us to prove the same lower bound in only a couple of lines, and shows that it holds even in the presence of proofs and advice.

8.4.1. Proposition (Lower bound for quantum phase estimation). *For $d \in \mathbb{Z}_+$ fixed, let $U \in \mathbb{U}(d)$ and let $|\psi\rangle$ be a quantum state such that $U|\psi\rangle = e^{2\pi i\theta}|\psi\rangle$ for some $\theta \in [0,1)$. Suppose that either (i) $\theta \geq b$ or (ii) $\theta \leq a$, with $b - a = \epsilon > 0$. Then any C-tester, where $\mathsf{C} \subseteq \mathsf{QMA}(\mathrm{poly}(n))/\mathsf{qpoly}$, that decides whether (i) or (ii) holds with success probability $\geq 2/3$ must make at least*

$$\Omega(1/\epsilon)$$

(controlled) queries to U (or its inverse).

Proof:
Let $U_1 = \mathbb{I}$ and $U_2 = e^{2i\pi\epsilon}|0\rangle\langle 0| + |1\rangle\langle 1|$, with $\epsilon > 0$. Note $|\psi\rangle = |0\rangle$ is an eigenstate of both U_1 and U_2 with eigenphases $\theta_1 = 0$ and $\theta_2 = \epsilon$, respectively. Hence, any algorithm that decides whether some U with eigenstate $|\psi\rangle$ has an eigenphase $\theta \leq a := 0$ or $\geq b := \epsilon$ can discriminate U_1 from U_2. We have $\mathrm{eig}(U_1^\dagger U_2) = \{e^{2i\pi\epsilon}, 1\}$, which means $0 \notin \mathrm{conv}(\mathrm{eig}(U_1^\dagger U_2)$. Using $D = \frac{1}{2}|1 + e^{2i\pi\epsilon}|$ we find

$$\frac{1}{2}\left\|\mathcal{U}(U_1) - \mathcal{U}(U_2)\right\|_\diamond = \sqrt{1 - \frac{1}{4}|1 + e^{2i\pi\epsilon}|^2} = |\sin(\pi\epsilon)| \leq \pi\epsilon$$

for $\epsilon > 0$. Hence, by Theorem 8.3.7 we find $T \geq \Omega\left(1/\epsilon\right)$. □

This matches the well-known upper bound by Kitaev [Kit95].

Entanglement entropy. In the entanglement entropy problem, one wants to decide whether a certain bipartite state $|\psi\rangle_{AB}$ has low or small entanglement entropy between the subsystems A and B. We consider the the (2-Rényi) entanglement entropy $S_2(\cdot)$, which for a mixed state ρ_A is defined as

$$S_2(\rho_A) = -\ln[\mathrm{tr}\left[\rho_A^2\right]].$$

8.4.2. Proposition (Lower bound for entanglement entropy). *For $d \in \mathbb{Z}_+$ fixed, let $U = \mathbb{I} - 2\,|\psi\rangle\langle\psi|$ be a unitary for some bipartite quantum state $|\psi\rangle = |\psi\rangle_{AB} \in \mathbb{C}^d \otimes \mathbb{C}^d$. Suppose that either (i) $S_2\big(\operatorname{tr}_B(|\psi\rangle\langle\psi|)\big) \leq a$ or (ii) $S_2\big(\operatorname{tr}_B(|\psi\rangle\langle\psi|)\big) \geq b$, with $b - a \geq \Delta > 0$. Then any C-tester, where $\mathsf{C} \subseteq \mathsf{QMA}(\mathrm{poly}(n))/\mathsf{qpoly}$, that decides whether (i) or (ii) holds with success probability $\geq 2/3$ must make at least*

$$T \geq \Omega\left(\frac{1}{\sqrt{\Delta}}\right)$$

(controlled) queries to U (or its inverse).

Proof:
We reduce from the following unitary channel discrimination problem: $U_i = \mathbb{I} - 2\,|\psi_i\rangle\langle\psi_i|$, $i \in [1,2]$, with

$$|\psi_1\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)\,, \quad |\psi_2\rangle = \sqrt{\frac{1+\sqrt{\Delta}}{2}}\,|00\rangle + \sqrt{\frac{1-\sqrt{\Delta}}{2}}\,|11\rangle \tag{8.15}$$

We have

$$\rho_{1,A} = \operatorname{tr}_B[|\psi_1\rangle\,\langle\psi_1|] = I_2$$

which has $S_2(\rho_{1,A}) = \ln(2)$ and

$$\rho_{2,A} = \operatorname{tr}_B[|\psi_2\rangle\,\langle\psi_2|] = \frac{1+\sqrt{\Delta}}{2}\,|0\rangle\,\langle 0| + \frac{1-\sqrt{\Delta}}{2}\,|1\rangle\,\langle 1|$$

which has $S_2(\rho_{2,A}) = -\ln((1+\Delta)/2) \leq \ln(2) - \Delta/2$, so $b - a \geq \Delta/2$. Hence, if we could compute $S_2(\rho_i)$ up to precision $< \Delta/4$ for $i \in \{1,2\}$ given access to U_i we could distinguish U_1 from U_2. We have that

$$\operatorname{eig}(U_1^\dagger U_2) = \{1, -i\sqrt{\Delta} + \sqrt{1-\Delta}, i\sqrt{\Delta} + \sqrt{1-\Delta}\},$$

which means that $0 \notin \operatorname{conv}(\operatorname{eig}(U_1^\dagger U_2)$ for $\Delta < 1$. We can brute force over all combinations of eigenvalues to find $D = \sqrt{1-\Delta}$ and thus

$$\frac{1}{2}\,\|\mathcal{U}(U_1) - \mathcal{U}(U_2)\|_\diamond = \sqrt{1-(1-\Delta)} = \sqrt{\Delta}$$

for $\Delta > 0$. By Theorem 8.3.7, we find a lower bound of $T \geq \Omega(1/\sqrt{\Delta})$. □

This removes the logarithmic factor of $\tilde{\Omega}(1/\sqrt{\Delta})$ in [WZ23], and also resolves their open question of whether their bound could be made to hold for a QMA-tester (we show, in fact, that it is robust against even stronger classes of testers).

A standard quantum in the literature to estimate the entanglement entropy, usually considered in the setting where the input to the state is given in a sample setting, is through the use of the SWAP-test [FKS21]. This gives an upper bound on the query complexity in terms of the precision Δ, the dimension d, and an upper bound on the entanglement entropy S_{upper}.

8.4.3. PROPOSITION. *Let $d = 2^n$ for some $n \in \mathbb{Z}_+$, and let $a, b \in [0, \ln d]$ with $b - a = \Delta > 0$. Suppose we are given a number $S_{\text{upper}} \in [0, \ln d]$, and access to a bipartite state $|\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ via a unitary $U = \mathbb{I} - 2|\psi\rangle\langle\psi|$, for which $S_2(|\psi\rangle) \leq S_{\text{upper}}$ holds. Then there exists a quantum algorithm that solves the entanglement entropy problem using*

$$\mathcal{O}\left(\frac{\sqrt{d}\, e^{S_{\text{upper}}}}{\Delta}\right)$$

queries to U, with success probability at least $2/3$.

Proof:
Let $\mathcal{A}$ be the quantum circuit which creates the maximally entangled state from the all-zeros state, i.e.,

$$\mathcal{A}|0\ldots0\rangle = \frac{1}{\sqrt{d}}\sum_{j=0}^{d-1}|j\rangle|j\rangle = \frac{1}{\sqrt{d}}\sum_{j=0}^{d-1}|\psi_j\rangle\left|\bar{\psi}_j\right\rangle,$$

using the fact that the maximally entangled state can be written in any orthonormal basis. By using exact amplitude amplification, we can prepare $|\psi\rangle$ exactly with probability 1 using $\mathcal{O}(\sqrt{d})$ queries to U and $\mathcal{A}$ [Hø00]. Let $\mathcal{B}$ be the algorithm which does this. If the SWAP-test [BCWdW01] is applied to two copies of a mixed state ρ, the probability of measuring 0 on the first qubit is given by $\frac{1}{2} + \frac{1}{2}\operatorname{tr}[\rho^2]$ (a proof of this can be found in [KMY03]). Using the SWAP test in conjunction with quantum amplitude estimation [BHMT02], we can estimate $\operatorname{tr}[\rho_A^2]$ up to additive precision ϵ using $\mathcal{O}(1/\epsilon)$ copies of $|\psi\rangle$ (i.e., calls to $\mathcal{B}$) with high probability. To make the estimation, when it succeeds, biased towards overestimating, we can apply the simple trick of providing an estimate $\bar{x}$ up to $\epsilon/2$ additive precision and making our new estimate $\hat{x} := \bar{x} + \epsilon/2$. Let $x := \operatorname{tr}[\rho_A^2]$. If we are guaranteed an upper bound on the entanglement entropy of S_{upper}, then we have that $x \in [e^{-S_{\text{upper}}}, 1]$. Choosing $\epsilon = \Delta e^{-S_{\text{upper}}}$, we have

$$\begin{aligned}
|-\ln x - (-\ln \hat{x})| &= |-\ln(x+\epsilon) - (-\ln(x))| \\
&= \left|\ln\left(\frac{x}{x+\epsilon}\right)\right| \\
&= \ln(1+\epsilon/x) \\
&\leq \ln\left(1+\epsilon e^{S_{\text{upper}}}\right) \\
&\leq \epsilon e^{S_{\text{upper}}} \\
&\leq \Delta/4,
\end{aligned}$$

which is a sufficient precision to distinguish between the two cases. Hence, we need to make a total of

$$T = \mathcal{O}\left(\sqrt{d}\cdot\frac{1}{\epsilon}\right) = \mathcal{O}\left(\frac{\sqrt{d}e^{S_{\text{upper}}}}{\Delta}\right)$$

queries to U. □

Our lower bound does not incorporate any dependence on the dimension, nor does it account for the fact that the estimation might become harder as the entanglement entropy approaches its maximum value. The lack of dimension-dependence might be explained by the fact that our lower bound does not take into account the difficulty of preparing $|\psi\rangle$ given access to U_i.

However, even when S_{upper} and d are constant, this bound is still quadratically worse in the precision Δ than the lower bounds from Proposition 8.4.2 and [WZ23]. We leave it as an open question whether these bounds (upper or lower) can be improved, especially in terms of achieving a better dependency on the precision Δ.

Subset support verification and amplitude estimation. Here we consider a variant to the amplitude estimation problem, which we will call the *subset support verification* problem. In this problem, one is given access to a unitary U which prepares a subset state for some subset $S \subseteq \{0,1\}^n$ when applied to the all-zeros state, as well as a bit string $j \in \{0,1\}^n$. The task is to decide whether $j \in S$ or $j \notin S$.

8.4.4. Proposition (Lower bound for subset support verification). *Let U be a unitary that applies a transformation of the form*

$$U|0^n\rangle = \frac{1}{\sqrt{|S|}} \sum_{i \in S} |i\rangle ,$$

for some subset $S \subseteq \{0,1\}^n$ and let $j \in \{0,1\}^n$. Then any C-tester, where $\mathsf{C} \subseteq \mathsf{QMA}(\text{poly}(n))/\mathsf{qpoly}$, that decides whether $j \in S$ or $j \notin S$ with success probability $\geq 2/3$ must make at least

$$T \geq \Omega\left(\sqrt{|S|}\right)$$

(controlled) queries to U (or its inverse).

Proof:
Let $S \subseteq \{0,1\}^n$ be any subset that contains 0^n, and let $j = 0^n$ (the proof can easily be modified to work for any choice of j). We can construct an explicit unitary U by setting $U = 2|v\rangle\langle v| - \mathbb{I}$, where

$$(2|v\rangle\langle v| - \mathbb{I})|0^n\rangle = |S\rangle .$$

Solving for $|v\rangle$, using that $\langle 0 \ldots 0|S\rangle$ is real we obtain

$$|v\rangle = \frac{|0^n\rangle + |S\rangle}{\sqrt{2(1 + \langle 0 \ldots 0|S\rangle)}}.$$

Let $S_1 = S$ and $S_2 = S_1 \setminus \{0^n\}$. We define

$$|v_1\rangle = \frac{|0^n\rangle + |S_1\rangle}{\sqrt{2(1+\sqrt{1/|S|})}}, \quad |v_2\rangle = \frac{|0^n\rangle + |S_2\rangle}{\sqrt{2}},$$

and

$$U_1 = 2\,|v_1\rangle\langle v_1| - \mathbb{I}, \quad U_2 = 2\,|v_2\rangle\langle v_2| - \mathbb{I}.$$

We can write $|v_2\rangle = \sqrt{1-\alpha}\,|v_1\rangle + \sqrt{\alpha}|v_1^\perp\rangle$ for some $\alpha \in [0,1]$. We have

$$U_2 = 2\left[(1-\alpha)\,|v_1\rangle\langle v_1| + \sqrt{1-\alpha}\sqrt{\alpha}\,|v_1\rangle\,\langle v_1^\perp| + \sqrt{1-\alpha}\sqrt{\alpha}\,|v_1\rangle\langle v_1^\perp| + \alpha|v_1^\perp\rangle\langle v_1^\perp|\right] - \mathbb{I}.$$

Writing $U_1^\dagger U_2$ in the $\{|v_1\rangle, |v_1^\perp\rangle, \dots\}$ basis we obtain

$$U_1^\dagger U_2 = \begin{bmatrix} B & 0 \\ 0 & I_{2^{n-1}} \end{bmatrix},$$

where B is a 2×2-matrix given by

$$B = \begin{bmatrix} 1-2\alpha & 2\sqrt{1-\alpha}\sqrt{\alpha} \\ -2\sqrt{1-\alpha}\sqrt{\alpha} & 1-2\alpha \end{bmatrix}.$$

Since $U_1^\dagger U_2$ is a block-diagonal matrix, its eigenvalues are given by

$$\mathrm{eig}(U_1^\dagger U_2) = \mathrm{eig}(B) \cup \mathrm{eig}(\mathbb{I}_{2^{n-1}}) = \{1, 1-2\alpha-2\sqrt{\alpha^2-\alpha}, 1-2\alpha+2\sqrt{\alpha^2-\alpha}\}.$$

Again, it is easy to see that for $\alpha < 1$ we have $0 \notin \mathrm{conv}(\mathrm{eig}(U_1^\dagger U_2)$. Therefore, $D = 1-2\alpha$ and

$$\frac{1}{2}\left\|(\mathcal{U}(U_1) - \mathcal{U}(U_2))\right\|_\diamond = 2\sqrt{\alpha-\alpha^2} \leq 2\sqrt{\alpha} \leq 2\sqrt{\frac{1}{|S|}}$$

since,

$$\begin{aligned}
\sqrt{\alpha} &= \sqrt{1 - |\langle v_1|v_2\rangle|^2} \\
&= \sqrt{1 - \left|\frac{1 + \langle 0^n|S_2\rangle + \langle S_1|0^n\rangle + \langle S_1|S_2\rangle}{2\sqrt{1+\sqrt{1/|S|}}}\right|^2} \\
&= \sqrt{1 - \left|\frac{1 + \sqrt{1/|S|} + \sqrt{(|S|-1)/|S|}}{2\sqrt{1+\sqrt{1/|S|}}}\right|^2} \\
&= \sqrt{\frac{1}{2} - \frac{1}{2}\sqrt{\frac{|S|-1}{|S|}}} \\
&\leq \sqrt{\frac{1}{2} - \frac{1}{2}\frac{|S|-1}{|S|}} \\
&\leq \sqrt{\frac{1}{|S|}}.
\end{aligned}$$

Hence, by Theorem 8.3.7 we have that $T \geq \mathcal{O}\left(\sqrt{|S|}\right)$ to distinguish U_1 and U_2 with probability $\geq 2/3$. □

The upper bound of $\mathcal{O}(\sqrt{|S|})$ follows directly from quantum amplitude estimation [BHMT02], for which the optimality is also a corollary of Proposition 8.4.4, as it proves a lower bound in a more restricted setting.

8.4.5. Corollary (Quantum amplitude estimation lower bound). *Given a unitary U which acts as*

$$U|0^n\rangle = \sqrt{\alpha}\,|\psi\rangle + \sqrt{1-\alpha}\,|\psi^\perp\rangle,$$

suppose that either (i) $\sqrt{\alpha} \leq a$ or (ii) $\sqrt{\alpha} \geq b$, with $b - a = \epsilon > 0$. Then any C*-tester, where* $\mathsf{C} \subseteq \mathsf{QMA}(\mathrm{poly}(n))/\mathsf{qpoly}$*, that decides whether (i) or (ii) holds with success probability $\geq 2/3$ must make at least*

$$T \geq \Omega\left(\frac{1}{\epsilon}\right)$$

(controlled) queries to U (or its inverse).

Proof:
Let $|S\rangle = \frac{1}{\sqrt{|S|}}\sum_{i\in S}|i\rangle$ with $|S| = 1/\epsilon^2$, where ϵ is chosen such that $|S|$ is integer

(this assumption does not change the bound qualitatively). Let $|\psi\rangle = |0^n\rangle$. Note that

$$\sqrt{\alpha_1} = \langle 0^n | U_1 | 0^n \rangle = 1/\sqrt{|S|}, \quad \sqrt{\alpha_2} = \langle 0^n | U_2 | 0^n \rangle = 0,$$

so deciding whether $\sqrt{\alpha_i} = \frac{1}{\sqrt{|S|}} = \epsilon$ or $\sqrt{\alpha_i} = 0$ is sufficient to distinguish U_1 from U_2. By the proof of Proposition 8.4.4, we then find $T \geq \Omega(1/\epsilon)$, completing the proof. □

8.4.2 Other unitary problems

The next two examples are not unitary property testing cases, but quantum algorithmic primitives involving the implementation of some unitary given access to some building block in the form of a block-encoding [LC17, LC19].

8.4.6. Definition (Block-encoding). Let M be $2^n \times 2^n$-dimensional matrix, $\alpha, \epsilon > 0$ and $a \in \mathbb{Z}_+$. An $(n+a)$-qubit unitary operator U is an (α, a, ϵ)-*block-encoding* of M if

$$\left\| \alpha(\langle 0|^{\otimes a} \otimes \mathbb{I}) U (|0\rangle^{\otimes a} \otimes \mathbb{I}) - M \right\| \leq \epsilon.$$

If $\alpha = a = 1$ and $\epsilon = 0$, we simply say that U is a block-encoding of M.

Given some n-qubit operator M, one can construct a $n+1$-qubit unitary operator U provided all singular values of M are upper bounded by 1 in the following way. Let $M = R\Sigma V^\dagger$ be the singular value decomposition (SVD) of M. Then

$$U = \begin{pmatrix} M & R\sqrt{\mathbb{I} - \Sigma^2}V^\dagger \\ R\sqrt{\mathbb{I} - \Sigma^2}V^\dagger & -M \end{pmatrix} \tag{8.16}$$

is a $(1, 1, 0)$-block-encoding of M, since

$$(\langle 0| \otimes \mathbb{I}) U (|0\rangle \otimes \mathbb{I}) = M.$$

Thermal state preparation (quantum Gibbs sampling). For some Hamiltonian H, the *Gibbs state* (or thermal state) ρ_β at inverse temperature β is defined as

$$\rho_\beta = \frac{e^{-\beta H}}{\mathrm{tr}[e^{-\beta H}]}. \tag{8.17}$$

We say a quantum algorithm is an *approximate Gibbs sampler* if it prepares the Gibbs state ρ_β up to some trace distance ϵ. In [CKBG23, Appendix G], an optimal lower bound in β is proven. We demonstrate that this bound can also be derived using our framework. The main distinction is that we prove the result by directly examining the diamond distance between the block-encodings of the Hamiltonians, rather than using reflections about a purified Gibbs state.

8.4.7. Proposition (Lower bound for quantum Gibbs sampling). *Let H be a Hamiltonian to which we have access through a block-encoding U_H. Suppose $\beta \geq \sqrt{\frac{14}{3}} \approx 2.16$. Then it takes at least*

$$T \geq \Omega(\beta)$$

queries to U_H to prepare the thermal state at inverse temperature β up to trace distance at most $1/24$.

Proof:
Let $H_1 = \left(\frac{1}{2} + \frac{1}{\beta}\right)|0\rangle\langle 0| + \left(\frac{1}{2} - \frac{1}{\beta}\right)|1\rangle\langle 1|$, and $H_2 = \left(\frac{1}{2} - \frac{1}{\beta}\right)|0\rangle\langle 0| + \left(\frac{1}{2} + \frac{1}{\beta}\right)|1\rangle\langle 1|$. We have that by Eq. (8.17) the thermal states are given by

$$\rho_{1,\beta} = \frac{1}{1+e^2}\begin{bmatrix} 1 & 0 \\ 0 & e^2 \end{bmatrix}$$

and

$$\rho_{2,\beta} = \frac{1}{1+e^2}\begin{bmatrix} e^2 & 0 \\ 0 & 1. \end{bmatrix}$$

Therefore,

$$\frac{1}{2}\left\|\rho_{1,\beta} - \rho_{2,\beta}\right\|_1 = 1 - \frac{2}{1+e^2} \geq \frac{3}{4}.$$

Now suppose that we can only prepare some $\tilde{\rho}_{i,\beta}$ such that $\frac{1}{2}\left\|\tilde{\rho}_{i,\beta} - \rho_{i,\beta}\right\|_1 \leq \epsilon$ for $i \in \{1,2\}$. By applying the reverse triangle inequality twice, we find that

$$\begin{aligned}
\frac{1}{2}\left\|\tilde{\rho}_{1,\beta} - \tilde{\rho}_{2,\beta}\right\|_1 &= \frac{1}{2}\left\|\rho_{1,\beta} - \rho_{2,\beta} - (\rho_{1,\beta} - \tilde{\rho}_{1,\beta}) - (\tilde{\rho}_{2,\beta} - \rho_{2,\beta})\right\|_1 \\
&\geq \frac{1}{2}\Big|\left\|\rho_{1,\beta} - \rho_{2,\beta} - (\rho_{1,\beta} - \tilde{\rho}_{1,\beta})\right\|_1 - \left\|\tilde{\rho}_{2,\beta} - \rho_{2,\beta}\right\|_1\Big| \\
&\geq \frac{1}{2}\Big|\Big|\left\|\rho_{1,\beta} - \rho_{2,\beta}\right\|_1 - \left\|\rho_{1,\beta} - \tilde{\rho}_{1,\beta}\right\|_1\Big| - \left\|\tilde{\rho}_{2,\beta} - \rho_{2,\beta}\right\|_1\Big| \\
&\geq \frac{3}{4} - 2\epsilon \geq \frac{2}{3}
\end{aligned}$$

when $\epsilon \leq \frac{1}{24}$, which means that $\tilde{\rho}_{1,\beta}$ and $\tilde{\rho}_{2,\beta}$ can be distinguished with success probability $\geq 2/3$. Hence, if $\tilde{\rho}_{i,\beta}$ can be constructed using the block-encoding U_i of H_i, we have a distinguisher for unitary channels associated with U_i for $i \in \{1,2\}$.

The SVDs of H_1 and H_2 are $\mathbb{I}H_1\mathbb{I}$ and $\mathbb{I}H_2\mathbb{I}$, since both are diagonal and positive semidefinite. Using Eq. (8.16) we can construct the following two block-

encodings of H_1 and H_2:

$$U_1 = \begin{bmatrix} \frac{1}{2}+\frac{1}{\beta} & 0 & \sqrt{1-\left(\frac{1}{2}+\frac{1}{\beta}\right)^2} & 0 \\ 0 & \frac{1}{2}-\frac{1}{\beta} & 0 & \sqrt{1-\left(\frac{1}{2}-\frac{1}{\beta}\right)^2} \\ \sqrt{1-\left(\frac{1}{2}+\frac{1}{\beta}\right)^2} & 0 & -\frac{1}{\beta}-\frac{1}{2} & 0 \\ 0 & \sqrt{1-\left(\frac{1}{2}-\frac{1}{\beta}\right)^2} & 0 & \frac{1}{\beta}-\frac{1}{2} \end{bmatrix},$$

$$U_2 = \begin{bmatrix} \frac{1}{2}-\frac{1}{\beta} & 0 & \sqrt{1-\left(\frac{1}{2}-\frac{1}{\beta}\right)^2} & 0 \\ 0 & \frac{1}{2}+\frac{1}{\beta} & 0 & \sqrt{1-\left(\frac{1}{2}+\frac{1}{\beta}\right)^2} \\ \sqrt{1-\left(\frac{1}{2}-\frac{1}{\beta}\right)^2} & 0 & -\frac{1}{\beta}+\frac{1}{2} & 0 \\ 0 & \sqrt{1-\left(\frac{1}{2}+\frac{1}{\beta}\right)^2} & 0 & -\frac{1}{\beta}-\frac{1}{2} \end{bmatrix}.$$

Evaluating the operator norm distance of U_1 and U_2 gives us

$$\|U_1 - U_2\| = \frac{\sqrt{3\beta^2 - \sqrt{9\beta^4 - 40\beta^2 + 16} + 4}}{\sqrt{2}\beta} \leq \frac{3}{\beta}$$

for $\beta \geq \sqrt{\frac{14}{3}}$. Moreover, for these values of β we have $\|U_1 - U_2\| \leq \frac{3}{\sqrt{14/3}} \leq 1.4$, which means that $0 \notin \mathrm{conv}(\mathrm{eig}(U_1^\dagger U^2)$ by Lemma 8.3.4. Therefore, by Lemma 8.3.3 we then have $T \geq \Omega(\beta)$. □

This matches the upper bound of some known quantum Gibbs samplers, see for example Ref. [CKG23].

No fast-forwarding for Hamiltonian simulation. In Hamiltonian simulation, one has access to some Hamiltonian H and is given a time $t \in \mathbb{R}$, with the goal of implementing the unitary $\tilde{U}$ which approximates $U = e^{-itH}$ up to diamond distance ϵ. It is well-known that it is generally not possible to do so-called Hamiltonian *fast-forwarding*, which refers to a Hamiltonian simulation algorithm which implements $\tilde{U}$ in time sub-linear in t [BACS07].

8.4.8. Proposition (No fast-forwarding for Hamiltonian simulation). *Let H with $\|H\| \leq 1$ be some Hamiltonian to which we have access through a block-encoding U_H. Suppose $t \geq 1/2\pi$. Then it takes at least*

$$T \geq \Omega(t)$$

queries to U_H to implement e^{-iHt} up to diamond distance $\leq 1/3$.

Proof:
Let $H_1 = \frac{\mathbb{I}_2}{2}$, and $H_2 = H_1 + \frac{1}{2t} |1\rangle\langle 1|$, such that $\|H_1\| \leq 1$ and $\|H_2\| \leq 1$. Define $t = 2\pi t'$ with $t' \geq 0$. Suppose we could implement $U_i = e^{-iH_i t}$ perfectly for $i \in \{1, 2\}$. We then have that

$$e^{-2\pi i H_1 t'} |+\rangle = e^{-i\pi t'} |+\rangle$$

and

$$e^{-2\pi i H_2 t'} |+\rangle = e^{-i\pi t'} |-\rangle ,$$

which are perfectly distinguishable by a measurement in the Hadamard basis. If instead we can implement some $\tilde{U}_i$ such that $\frac{1}{2} \left\| \tilde{U}_i - U_i \right\|_\diamond \leq 1/3$ for $i \in \{1, 2\}$, our Hadamard basis measurement will be able to distinguish both cases with success probability $\geq 2/3$. Hence, we have that a Hamiltonian simulation algorithm that uses U_H as a subroutine can be used to distinguish U_{H_1} from U_{H_2}. For the block-encodings U_1 and U_2 of H_1 and H_2, the matrices are given by

$$\begin{bmatrix} \frac{1}{2} & 0 & \frac{\sqrt{3}}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & 0 & -\frac{1}{2} & 0 \\ 0 & \frac{\sqrt{3}}{2} & 0 & -\frac{1}{2} \end{bmatrix} \text{ and } \begin{bmatrix} \frac{1}{2} & 0 & \frac{\sqrt{3}}{2} & 0 \\ 0 & \frac{1+1/t'}{2} & 0 & \sqrt{1 + \left(\frac{1}{2} + \frac{1}{2t'^2}\right)^2} \\ \frac{\sqrt{3}}{2} & 0 & -\frac{1}{2} & 0 \\ 0 & \sqrt{1 + \left(\frac{1}{2} + \frac{1}{2t'^2}\right)^2} & 0 & -\frac{1+1/t'}{2} \end{bmatrix},$$

respectively. Again by Lemma 8.3.3, evaluating the distance with respect to the operator norm, we find

$$\|U_1 - U_2\| = \frac{\sqrt{3 - \sqrt{-\frac{3}{t'^2} - \frac{6}{t'} + 9} - \frac{1}{t'}}}{\sqrt{2}} \leq \frac{1}{t'},$$

for which the inequality can be shown to hold after some straightforward algebraic manipulation, assuming that $t' \geq 1$. Again, for $t' \geq 1$ (and thus $t \geq 1/2\pi$) we have $\|U_1 - U_2\| \leq 1 \leq 1.4$, so $0 \notin \text{conv}(\text{eig}(U_1^\dagger U^2)$ by Lemma 8.3.4. Hence, by Theorem 8.3.7, $T = \Omega(t)$. □

See [GSLW19] for a matching upper bound in the simulation time t.

Hamiltonian learning in the Heisenberg limit. In Hamiltonian learning, the problem is to output a classical description of a Hamiltonian H' that is ϵ-close to H with respect to some distance measure d in the space of Hamiltonians, with probability $\geq$ by making queries to its time evolution operator $U(t) = e^{-itH}$,

where $t \geq 0$ is tunable. The total evolution time is then defined as the sum of all different evaluation times used in the different queries. As a distance measure, we will take the operator norm distance $\left\|\tilde{H} - H\right\|$.

8.4.9. Proposition. *Given access to an unknown Hamiltonian H with $\|H\| \leq 1$ through its time evolution operator $U(t) = e^{-itH}$, where $t \geq 0$ is a tunable parameter, outputting a matrix $\tilde{H}$ such that $\left\|\tilde{H} - H\right\| \leq \epsilon$, for some $0 < \epsilon \leq \frac{1}{2}$, with probability $\geq 2/3$ requires a total evolution time of at least*

$$\Omega(1/\epsilon).$$

Proof:
Let $H_1 = \mathbb{I}$ and $H_2 = \mathbb{I} - 2\epsilon\,|0\rangle\langle 0|$. Since $0 < \epsilon \leq \frac{1}{2}$, we have $\|H_1\| \leq 1$ and $\|H_2\| \leq 1$ holds. Learning an unknown U up to ϵ in operator distance allows one to distinguish H_1 from H_2. Suppose that we make m queries to the time evolution operator, each with some evolution time t_j with $j \in [m]$. We have that $U_1(t_j) = e^{-it_j}\mathbb{I}$ and

$$U_2(t_j) = \begin{bmatrix} e^{-it_j(1-2\epsilon)} & 0 \\ 0 & e^{-it_j} \end{bmatrix}.$$

We have that

$$U_1(t_j)^\dagger U_2(t_j) = \begin{bmatrix} e^{2it_j\epsilon} & 0 \\ 0 & 1 \end{bmatrix}.$$

which means that $D = \frac{1}{2}|1 + e^{2it_j\epsilon}|$. Now suppose that $t_j \geq \frac{\pi}{2\epsilon}$. Then this would match our desired lower bound. However, if $t_j\epsilon < \frac{\pi}{2\epsilon}$, the condition $0 \notin \mathrm{conv}(\mathrm{eig}(U_1^\dagger U_2))$ is satisfied. Therefore, by Lemma 8.3.1 we have

$$\frac{1}{2}\left\|\mathcal{U}(U_1(t_j)) - \mathcal{U}(U_2(t_j))\right\|_\diamond = \sqrt{1 - D^2} = |\sin(t_j\epsilon)| \leq t_j\epsilon.$$

The proof of Theorem 8.3.7 then directly implies that $T = \sum_{j\in[m]} t_j = \Omega(1/\epsilon)$. □

See [HTFS23] for a (local) Hamiltonian learning algorithm which achieves the Heisenberg scaling for its total evolution time.

Ground state preparation of spectrally-gapped Hamiltonians. In approximate ground state preparation problems one is given a Hamiltonian H with some unknown ground state $|\psi\rangle$, for which one has to prepare a state $|\psi\rangle$ such that $|\langle\psi_0|\psi\rangle|^2 \geq 1 - \epsilon$. It is well known that the *spectral gap* $\gamma(H)$, that is the difference between the energies of the ground state and the first excited state of H, plays an important role in how difficult this problem is [LT20b, ALVV17, DGF22].

8.4.10. Proposition (Spectrally gapped ground state preparation). *Let H with $\|H\| \leq 1$ be a Hamiltonian to which we have access through a block-encoding U_H. Suppose $\gamma(H) \leq \Delta$ for some $0 < \Delta \leq 1$. Then it takes at least*

$$T \geq \Omega(1/\Delta)$$

queries to U_H to implement a state $|\phi\rangle$ that approximates the ground state $|\psi\rangle$ up to fidelity at least $2/3$.

Proof:
For a Hilbert space $\mathcal{H} := \mathbb{C}^3$, let $H_1 = \Delta\,|1\rangle\langle 1| + |2\rangle\langle 2|$ and $H_2 = \Delta\,|0\rangle\langle 0| + |2\rangle\langle 2|$. Since $0 < \Delta \leq 1$, we have that $\|H_1\| \leq 1$ and $\|H_2\| \leq 1$ holds. The ground states of H_1 and H_2 are given by $|\psi_1\rangle = |0\rangle$ and $|\psi_2\rangle = |1\rangle$, respectively. Note that indeed $\gamma(H_1) = \gamma(H_2)) = \Delta$. Suppose that we have an algorithm could prepare the ground state of an unknown H up to fidelity $\geq 2/3$ by making queries to the block-encoding of H, then we could distinguish H_1 from H_2 by preparing the ground state and performing a measurement $\{|0\rangle\langle 0|\,, \mathbb{I} - |0\rangle\langle 0|\}$. For some real number $c = |\langle\phi|\psi\rangle|$ with $|c|^2 \geq 2/3$, and $\theta \in [0, 1]$, we can write the prepared state as $|\phi\rangle = ce^{i\theta}\,|\psi\rangle + (1-c)|\psi^{\perp}\rangle$, where $|\psi^{\perp}\rangle$ lives in the space orthogonal to $|\psi\rangle$. Clearly, if $|\psi\rangle = |0\rangle$ then we have that $|\,\langle\phi|0\rangle\,| = |c|^2 \geq 2/3$ and when $|\psi\rangle = |1\rangle$ then $|\,\langle\phi|0\rangle\,|^2 \leq 1 - |c|^2 \leq 1/3$. Therefore, we have that the probability of measuring $|0\rangle$ when $H = H_1$ is $\geq 2/3$ and when $H = H_2$ it is $\leq 1/3$, which means that we can distinguish between both cases with an overall success probability $\geq 2/3$. For the block-encodings U_1 and U_2 of H_1 and H_2, the matrices are given by

$$\begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & \Delta & 0 & 0 & \sqrt{1-\Delta^2} & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sqrt{1-\Delta^2} & 0 & 0 & -\Delta & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} \text{ and } \begin{bmatrix} 0 & 0 & 0 & \sqrt{1-\Delta^2} & 0 & 0 \\ 0 & \Delta & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \sqrt{1-\Delta^2} & 0 & 0 & -\Delta & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix},$$

respectively. A direct computation shows that

$$\|U_1 - U_2\| = \frac{1}{2}\left(\Delta + \sqrt{8 - 3\Delta^2 - 8\sqrt{1-\Delta^2}}\right) \leq 2\Delta.$$

For $\Delta \leq 0.7$ we therefore have that $\|U_1 - U_2\| \leq 1/4$ and thus, by virtue of Lemma 8.3.4, $0 \notin \mathrm{conv}(\mathrm{eig}(U_1^{\dagger}U_2))$. By Theorem 8.3.7, $T \geq \Omega(1/\Delta)$. □

For an algorithm that uses H in its block-encoding and achieves $\tilde{\mathcal{O}}(1/\Delta)$ scaling, see [LT20b].

8.5 Quantum oracle separations with SBQP

In the final section, we will use our lower bound technique to prove two quantum oracle separations. We will first argue that lower bounds on unitary property testing can generally be used to show quantum oracle separations using the following strategy:

1. Take a language L which is not in C_1 nor in C_2 (if it is already in C_1, this already implies a separation).

2. Define a unitary property $\mathcal{P} = (\mathcal{P}_{\text{yes}}, \mathcal{P}_{\text{no}})$ in such a way that for a family of quantum oracles $\{U_x\}$, parametrised by the input x, the following two conditions hold:

   - If $x \in L$, then $U_x \in \mathcal{P}_{\text{yes}}$, and if $x \notin L$, then $U_x \in \mathcal{P}_{\text{no}}$.
   - Deciding whether $U_x \in \mathcal{P}_{\text{yes}}$ or $U_x \in \mathcal{P}_{\text{no}}$ can be done by a C_1-tester, using appropriate bounds on the available resources (e.g., workspace size, proofs, advice, etc. allowed in C_1), and without exceeding the maximum allowed query complexity. However, it cannot be done by any C_2-tester: the query complexity necessarily exceeds the allowed bound for all $n \geq n_0$, for some $n_0 \in \mathbb{Z}_+$, even when unlimited other resources are permitted.

3. Use the standard technique of diagonalization [BGS75] to show a quantum oracle separation between C_1 and C_2.

However, in our case, it turns out the separation can be shown in an even simpler way. Since we want oracle separations with SBQP, we can construct an oracle that is very close in diamond distance to the identity operator. For any class that makes only a polynomial number of queries and can only distinguish between inverse polynomial output probabilities (and has error reduction), the unitary oracle will look just like the identity operator at large values of n. Hence, if a language L is not in C_2, it will also not be in C_2^U since this would imply that it would also be in $\mathsf{C}_2^{\mathbb{I}} = \mathsf{C}_2$. We will need the following lemma of one-bit quantum phase estimation to show containment in SBQP.

8.5.1. Lemma (One-bit quantum phase estimation [Kit95]). *Let $U \in \mathbb{U}(2^n)$ and let $|\psi\rangle$ be an n-qubit quantum state (which can be prepared exactly in an efficient manner) such that $U|\psi\rangle = e^{2\pi i\theta}|\psi\rangle$ for some $\theta \in [0,1)$. Then there exists a quantum algorithm that makes one controlled query to U and outputs 1 with probability $\sin^2(\pi\theta)$.*

Proof:
The one-bit quantum phase estimation circuit for a n-qubit unitary starts with

$n+1$ qubits initialised in $|0\rangle|\psi\rangle$, applies a Hadamard to the first qubit, followed by an application of U to the remaining n qubits controlled on the first qubit, and finishes by applying yet another Hadamard to first qubit followed by a measurement in the computational basis of this qubit. We have that the output state of the one-bit quantum phase estimation circuit is given by

$$\frac{1}{2}(|0\rangle + |1\rangle)|\psi\rangle + \frac{e^{2\pi i\theta}}{2}(|0\rangle - |1\rangle)|\psi\rangle,$$

which means that the probability of measuring '1' as a function of $\theta \in [0,1)$ can be expressed as

$$\Pr[\text{One-bit QPE outputs 1}] = \left|\frac{1-e^{2i\pi\theta}}{2}\right|^2 = \sin^2(\pi\theta).$$

□

We are now ready to state the proof, which uses the above argument to turn the lower bound of Proposition 8.4.1 into a quantum oracle separation.

8.5.2. Theorem. *There exists a quantum oracle $\boldsymbol{\mathcal{U}}$ such that*

$$\mathsf{QMA}(2)^{\boldsymbol{\mathcal{U}}} \not\supset \mathsf{SBQP}^{\boldsymbol{\mathcal{U}}}.$$

Proof:
Since $\mathsf{QMA}(2) \subseteq \mathsf{NEXP} \neq \mathsf{ALL}$, there must exist a unary language L for which $L \notin \mathsf{QMA}(2)$. Pick any such L, let $p(n)$ be a large polynomial, and define the quantum oracle $\boldsymbol{\mathcal{U}} = \{U_n : n \in \mathbb{Z}_+\}$ with $U_n = e^{2\pi i\theta_n}|0\rangle\langle 0| + |1\rangle\langle 1|$ for some $\theta_n \in [0,1)$ as follows:

- If $0^n \in L$, then $\theta_n = 2^{-p(n)}$.
- If $0^n \notin L$, then $\theta_n = 0$.

Clearly, we have that $L \in \mathsf{SBQP}^{\boldsymbol{\mathcal{U}}}$, since running the one-bit quantum phase estimation protocol from Lemma 8.5.1 with $U = U_n$ and $|\psi\rangle = |0\rangle$, and accepting when it outputs '1', satisfies:

- If $0^n \in L$, then $\Pr[\text{One-bit QPE accepts}] = \sin^2(\pi 2^{-p(n)}) \geq 2^{-2p(n)} \geq 2^{-p'(n)}$,
- If $0^n \notin L$, then $\Pr[\text{One-bit QPE accepts}] = 0 \leq 2^{-p'(n)-1}$,

for some polynomial $p'(n)$.

We will now show that $L \notin \mathsf{QMA}(2)^{\boldsymbol{\mathcal{U}}}$. Let $q(n)$ be a polynomial which bounds the runtime of the $\mathsf{QMA}(2)^{\boldsymbol{\mathcal{U}}}$ verifier, and let $|\psi_1\rangle$,$|\psi_2\rangle$ be the provided quantum

proofs. Observe that the identity operator has a diamond distance $\leq \pi 2^{-p'(n)}$ from the oracle $U_n \in \boldsymbol{\mathcal{U}}$. Hence, if we replace the oracle U_n with the identity operator $\mathbb{I}$ in for all $n \geq 1$, we have that by Theorem 8.3.7:

$$\left|\Pr\left[V^{U_n} \text{ accepts } (x, |\psi_1\rangle\,|\psi_2\rangle)\right] - \Pr\left[V^{\mathbb{I}} \text{ accepts } (x, |\psi_1\rangle\,|\psi_2\rangle)\right]\right| \leq \frac{\pi q(n)}{2^{p'(n)}} \leq 0.01,$$

for all $n \geq n_0$, where n_0 is some constant depending on the polynomials $p(n)$ and $q(n)$. Hence, if $L \in \mathsf{QMA}(2)^U$, then $L \in \mathsf{QMA}(2)$ as well. This follows because for all $n < n_0$, we make at most $2^{p'(n_0)}$ queries (which is constant for a fixed n_0) and replace the oracle with the identity for $n \geq n_0$, while maintaining bounded error. We can apply error reduction as in [HM13] to boost the success probability back to $\geq 2/3$. This contradicts the fact that $L \notin \mathsf{QMA}(2)$. □

8.5.3. Theorem. *There exists a quantum oracle* $\boldsymbol{\mathcal{U}}$ *such that*

$$\mathsf{QMA/qpoly}^{\boldsymbol{\mathcal{U}}} \not\supset \mathsf{SBQP}^{\boldsymbol{\mathcal{U}}}.$$

Proof:
This follows from a similar proof as Theorem 8.5.2, but now we choose a binary language L (to be specified later). Let the quantum oracle $\boldsymbol{\mathcal{U}} = \{U_x : x \in \{0,1\}^n, n \in \mathbb{Z}_+\}$ with $U_x = e^{2\pi i\theta_x}\,|0^n\rangle\langle 0^n| + (\mathbb{I} - |0^n\rangle\langle 0^n|)$ be a family of n-qubit unitaries parametrized by some $\theta_x \in [0,1)$, which is given as follows:

$$\theta_x = \left(1 + (-1)^{L(x)}\right) 2^{-p(|x|)-1},$$

where $L(x) = 1$ if $x \in L$ and $L(x) = 0$ otherwise. Hence, we have that

- If $x \in L$, then $\theta_x = 0$;
- If $x \notin L$, then $\theta_x = 2^{-p(|x|)}$.

Clearly, $L \in \mathsf{SBQP}^{\boldsymbol{\mathcal{U}}}$ for any choice of language L by the same argument as in Theorem 8.5.2. Observe that the lower bound of Proposition 8.4.1 also hold for these specific instances of U_x where the number of qubits acts on varies, as the diamond distance only changes with different n because the eigenphases change with different values of n (if $\theta = \epsilon$ for some fixed $\epsilon > 0$, then the diamond distance would be fixed for all n). Since $\mathsf{QMA/qpoly} \subseteq \mathsf{PSPACE/poly} \neq \mathsf{ALL}$ [Aar06], we can use the same argument as in Theorem 8.5.2 to show that there must exist a $L \notin \mathsf{QMA/qpoly}^{\boldsymbol{\mathcal{U}}}$. □

It is not clear if we can combine both unentangled quantum proofs as advice and arrive at a similar oracle separation, as we do not know whether $\mathsf{QMA}(2)/\mathsf{qpoly} = \mathsf{ALL}$ or not [Aar06].

8.6 Concurrent work and open problems

After the work in this chapter appeared as a pre-print on the arXiv, Chen, Wang, and Zhang [CWZ24] improved the lower bound for the entanglement entropy problem by also incorporating the dimension parameter, meaning that the best lower bound now depends on both the dimension and the precision. However, they obtained the same scaling in the precision parameter Δ, so in terms of this parameter the best-known upper and lower bounds still do not match. Since three different techniques ([CWZ24], [WZ23], and ours) all result in the same lower bound in terms of Δ, we believe that the upper bound might not be tight; which would be the more interesting case, as this would suggest that the standard technique, as explained in Section 8.4, is not optimal.

Finally, we wonder whether there are more examples of unitary property testing problems, besides the quantum search problem in [AK07], where one can show that quantum proofs provably reduce the quantum query complexity? What about quantum advice?

Chapter 9

Complement sampling

9.1 Introduction

In 2019, Google claimed to have demonstrated quantum advantage for the first time: their Sycamore processor performed a computational task in a matter of minutes that was estimated to take the world's most powerful classical computer approximately 10,000 years to complete [AAB+19]. The experiment was based on *random circuit sampling*, where, given a classical description of a quantum circuit, the goal is to sample from the probability distribution defined by the output state in the computational basis. This claim sparked a wave of research on both sides of the spectrum: new advantage claims were made based on other sampling experiments [ZWD+20, MLA+22, ZCC+22], and new classical simulation techniques [HZN+20, BCG21, PCZ22, OJF23, GKC+24] were designed to attack the experiments—challenging whether quantum advantage had truly been achieved.

For a convincing demonstration of quantum advantage, the task at hand should ideally satisfy the following three criteria:

1. It can be solved efficiently using a near-term quantum experiment.
2. It is *provably* hard to solve classically, i.e., it requires superpolynomial time as the system size scales.
3. The solution can be efficiently verified by a classical computer with minimal trust in the quantum device.

In the case of random circuit sampling, the main issue is that verification seems unavoidably tied to the underlying classical hardness; as a result, it seems to inherently require exponential time to classically verify. A recent proposal to sidestep this problem is to consider *random peaked circuits* [AZ24], though these are currently not sufficiently understood to assess their potential as a quantum advantage experiment. Many other proposals, such as Shor's algorithm [Sho94], Yamakawa–Zhandry search [YZ24], variational quantum eigensolvers [PMS+14],

and cryptographic tests of quantumness [BCM+21], tend to satisfy at most two of the three criteria.

However, quantum resources offer more than just potential *computational* speedups, where ultimately *time* is the resource one wants to minimise. For example, it is known (or in some cases strongly conjectured) that superpolynomial advantages exist in communication complexity [BCWdW01, GM23], sample complexity [BJ95, GKZ19], and space complexity [GKK+07, KPV24]. All these advantages stem from the fact that quantum information processing not only provides potential benefits in computation, but also in the way data can be stored, communicated, and accessed in a quantum setting. Therefore, a natural question is whether all three of the above criteria can be more easily achieved in a setting beyond the "standard" computational setting.

9.1.1 Results in this chapter

In this chapter, we introduce and study a new sample-to-sample problem that we call *complement sampling*. In this problem, one is given (quantum) sampling access to elements drawn uniformly from a subset S of size K from a universe of N elements, and the task is to output any element from the complementary subset $\bar{S}$, which contains the remaining $N - K$ elements. For the uniform distribution over any set S, a quantum sample $|S\rangle$ is defined as the uniform superposition over the elements in S (so each amplitude is the square root of the corresponding probability).

We prove the following results:

- For $K = N/2$, complement sampling can be solved with a single quantum sample, achieving success probability 1 by *swapping* $|S\rangle$ to $|\bar{S}\rangle$. Classically, the sample complexity is $\Theta(N)$, even when allowing any constant error probability bounded away from $1/2$. Moreover, we show that this classical sample lower bound also holds in an average-case setting.

- For general values of K, we present a zero-error (Las Vegas) quantum algorithm that is provably optimal for swapping $|S\rangle$ to $|\bar{S}\rangle$ when restricted to using only one ancilla qubit, which serves to flag whether the swap succeeded.

- Assuming the existence of one-way functions, the required number of classical samples remains superpolynomial in $\log N$, which makes the problem classically intractable when N is exponentially large.[1]

[1]Thus, our result implies that classical cryptography can only exist if quantum sample advantage is possible, since the existence of one-way functions is a minimal assumption for classical cryptography [IL89].

Our approach seems to make optimal use of two key principles: (i) quantum computers are inherently samplers, and (ii) they can exploit quantum resources in ways that have no classical analogue.

Finally, as the quantum algorithm for complement sampling also has very low circuit complexity, we conclude by arguing that complement sampling is a promising candidate for demonstrating a form of *quantum sample advantage*, as it appears to meet all three criteria discussed above when the subsets are generated via a sufficiently random, yet easy-to-implement, permutation.

9.2 Complement sampling

We begin by introducing some notation specific to this chapter. We then formally introduce the complement sampling problem and briefly discuss its relation to existing work in the literature.

For some $n \in \mathbb{Z}_+$, let $\mathcal{X} := \{0,1\}^n$ be the set of all bit strings of length n. We write $\mathcal{S}_K$ for the family of all subsets $S \subset \mathcal{X}$ with cardinality K. Given an $S \in \mathcal{S}_K$, let $\bar{S} = \mathcal{X} \setminus S$ be the complement set with cardinality $N - K$. The task of complement sampling is then defined as follows:

Complement Sampling: Given (quantum) sampling access to the uniform distribution over a subset $S \in \mathcal{S}_K$, output an element $y \in \bar{S}$ where $\bar{S}$ is the complement of S, i.e., $\mathcal{X} \setminus S$.

We will be interested in the power of quantum samples and quantum computation, as compared to classical samples, in solving the complement sampling problem. Given a probability distribution $D : S \subseteq \mathcal{X} \to [0,1]$, a quantum sample from D is defined as having access to a single copy of the state [BJ95, AdW17]

$$|S_D\rangle = \sum_{x \in S} \sqrt{D(x)}\, |x\rangle\,. \tag{9.1}$$

When D is the uniform distribution, we simply write $|S\rangle$. Since repeatedly measuring copies of $|S_D\rangle$ in the computational basis is identical to classical sampling from the distribution $D(x)$, there would be no advantage to a classical algorithm in having access to the quantum samples if all the classical algorithm can do is computational basis measurements.

Complement sampling can be viewed as a "sample-to-sample" problem where, given samples from some set A according to a distribution μ_A, one has to output something from a related set B, according to a distribution μ_B. In this chapter, however, we relax the condition of having a desired output distribution, and any element in the support on the distribution μ_B will do (though we will see that our

quantum algorithm is in fact able to provide a sample according to the uniform distribution over $\bar{S}$).

Our setting is different from, for example, most problems in classical [Hau92, Han16] and quantum [Aar07, AdW17, CB24] machine learning that use the number of samples as a complexity measure. For a concrete example, [ABC$^+$20] studies the sample complexity of the quantum coupon collector, where the task is to fully identify a subset A given quantum sampling access to its elements. In contrast, complement sampling does not require learning anything about the subset A. Moreover, [ABC$^+$20] also considers a stronger input model which essentially gives access to the state preparation circuit for the quantum sample. In complement sampling we do not consider this more powerful access model.

9.3 Quantum complement sampling

We will discuss solutions to the complement sampling task where a quantum computer swaps the input state as

$$|S\rangle = \frac{1}{\sqrt{K}} \sum_{x \in S} |x\rangle \qquad \longrightarrow \qquad \left|\bar{S}\right\rangle = \frac{1}{\sqrt{N-K}} \sum_{x \notin S} |x\rangle \,.$$

We will call such a transformation *complement swapping*. Hence, if $\left|\bar{S}\right\rangle$ was successfully prepared, a measurement in the computational basis yields a uniformly random sample from $\bar{S}$.

9.3.1 Aaronson, Atia and Susskind's construction

We will start by reviewing Aaronson, Atia and Susskind's connection between swapping and distinguishing complexities of quantum states [AAS20], which formed our basis to arrive at our solution for the quantum sample variant to complement sampling as explained in Section 9.3. Though we found alternative proofs and easier constructions for most results in this chapter—except for those in Section 9.5—which do not rely on this framework, we feel that it still provides the best intuition of why complement sampling can be perfectly solved on a quantum computer when $K = N/2$.

We start by recalling their four notions of complexities on quantum states: relative complexity, circuit complexity, swap complexity and distinguishability complexity.

9.3.1. Definition (Relative complexity). The relative complexity $\mathcal{C}_\epsilon(|a\rangle, |b\rangle)$ of two n-qubit pure quantum states $|a\rangle$,$|b\rangle$ is defined as the minimal number of gates in a circuit C such that

$$|\langle b| \langle 0 \dots 0| C |a\rangle |0 \dots 0\rangle| \geq 1 - \epsilon.$$

9.3.2. Definition (Circuit complexity). The circuit complexity $\mathcal{C}_\epsilon(|a\rangle)$ of an n-qubit pure quantum state $|a\rangle$ is defined as the relative complexity with the n-qubit all zero state $|0\dots0\rangle$, i.e.,

$$\mathcal{C}_\epsilon(|a\rangle) = \mathcal{C}_\epsilon(|0\dots0\rangle, |a\rangle).$$

9.3.3. Definition (Swap complexity). The swap complexity $\mathcal{S}_\epsilon(|a\rangle, |b\rangle)$ of two n-qubit pure quantum states $|a\rangle$,$|b\rangle$ is defined as the minimal number of gates in a circuit C such that

$$\frac{1}{2}\big|\langle a|\langle 0\dots0|C|b\rangle|0\dots0\rangle + \langle b|\langle 0\dots0|C|a\rangle|0\dots0\rangle\big| \geq 1-\epsilon.$$

It can be verified that $\mathcal{S}_\epsilon \geq \mathcal{C}_\epsilon$ holds for any $0 \leq \epsilon \leq 1$ [AAS20].

9.3.4. Definition (Distinguishability complexity). The distinguishability complexity $\mathcal{D}_\epsilon(|a\rangle, |b\rangle)$ of two n-qubit pure quantum states $|a\rangle$,$|b\rangle$ is defined as the minimal number of gates in a circuit C such that

$$\big|\||\Pi_1 C|a\rangle|0\dots0\rangle\|^2 - \|\Pi_1 C|b\rangle|0\dots0\rangle\|^2\big| \geq 1-\epsilon,$$

where $\Pi_1 = |1\rangle\langle 1| \otimes \mathbb{I}$.

The key result from [AAS20] is an equivalence (in terms of the order of circuit complexity) between a perfect swapper of two orthogonal states, and a perfect distinguisher for the corresponding conjugate states.

9.3.5. Lemma (Adapted from [AAS20], Theorem 2). *Let $0 \leq \epsilon < 1$, and let $n \in \mathbb{Z}_+$. Let $|a\rangle$, $|b\rangle$ be orthogonal n-qubit states and let $|\phi^+\rangle = \frac{1}{\sqrt{2}}(|a\rangle + |b\rangle)$ and $|\phi^-\rangle = \frac{1}{\sqrt{2}}(|a\rangle - |b\rangle)$. The following holds:*

If $\mathcal{S}_\epsilon(|a\rangle, |b\rangle) = T_1$, then we have that $\mathcal{D}_\epsilon(|\phi^+\rangle, |\phi^-\rangle) = \mathcal{O}(T_1)$.

If $\mathcal{D}_\epsilon(|\phi^+\rangle, |\phi^-\rangle) = T_2$, then we have that $\mathcal{S}_\epsilon(|a\rangle, |b\rangle) = \mathcal{O}(T_2)$.

The following claim is also given in [AAS20, Corollary 1], but we include it to specify the values of ϵ, as this will be needed in Section 9.5.

9.3.6. Proposition. *Let $\mathcal{G}$ be a finite, self-inverse gate set. Let $|a\rangle, |b\rangle$ be two orthogonal n-qubit quantum states. If $\mathcal{D}_{4\epsilon}(|a\rangle, |b\rangle) = T$, then $\mathcal{C}_\epsilon(|a\rangle) \geq T$ and $\mathcal{C}_\epsilon(|b\rangle) \geq T$.*

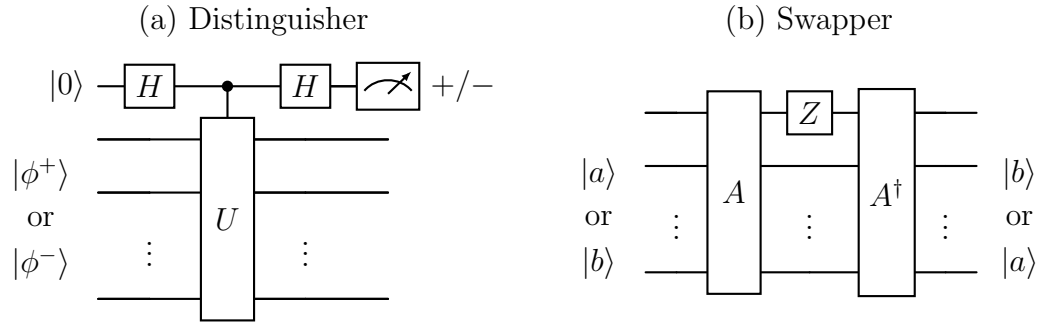


Figure 9.1: Quantum circuits for Aaronson, Atia and Susskind's construction [AAS20]. (a) A circuit distinguishing $|\phi^+\rangle$ from $|\phi^-\rangle$ using a unitary U that swaps $|a\rangle$ with $|b\rangle$. (b) A circuit swapping $|a\rangle$ and $|b\rangle$ using a circuit A that distinguishes $|\phi^+\rangle$ and $|\phi^-\rangle$.

Proof:
We prove this by contradiction. Suppose, without loss of generality, that $\mathcal{C}_\epsilon(|a\rangle) < T$. Then, there exists a quantum circuit C with gate complexity strictly less than T such that

$$|\langle a| C |0 \dots 0\rangle| \geq 1 - \epsilon.$$

Defining $|o\rangle = C|0 \dots 0\rangle$, we can write $|o\rangle$ as

$$|o\rangle = \sqrt{1-\alpha} e^{i\theta_1} |a\rangle + \sqrt{\alpha} e^{i\theta_2} |a^\perp\rangle$$

for some $\theta_1, \theta_2 \in [0, 2\pi]$ and $\alpha \in [0, 1]$. The assumption $|\langle a|o\rangle| \geq 1 - \epsilon$ implies that $\sqrt{1-\alpha} \geq 1 - \epsilon$. Now, consider the circuit that receives an unknown state $|z\rangle$, where $z \in \{a, b\}$, applies $C^\dagger$, and measures in the computational basis. The probability of obtaining the all-zero outcome is then given by $|\langle z|o\rangle|^2$. For $|z\rangle = |a\rangle$, we have

$$|\langle a|o\rangle|^2 \geq (1-\epsilon)^2.$$

For $|z\rangle = |b\rangle$, orthogonality implies that

$$|\langle b|o\rangle|^2 = 1 - |\langle a|o\rangle|^2 \leq 1 - (1-\epsilon)^2.$$

Thus, the bias of this measurement-based distinguisher is

$$\left| |\langle a|o\rangle|^2 - |\langle b|o\rangle|^2 \right| \geq (1-\epsilon)^2 - (1 - (1-\epsilon)^2) = 1 - 4\epsilon + 2\epsilon^2 \geq 1 - 4\epsilon.$$

Since $C^\dagger$ has the same gate complexity as C, this gives a distinguisher with gate complexity $< T$ and bias at least $1 - 4\epsilon$, contradicting the assumption that $\mathcal{D}_{4\epsilon}(|a\rangle, |b\rangle) = T$. The same argument applies if $\mathcal{C}_\epsilon(|b\rangle) < T$ or both $\mathcal{C}_\epsilon(|a\rangle)$ and $\mathcal{C}_\epsilon(|b\rangle)$ are smaller than T. Thus, we conclude that $\mathcal{C}_\epsilon(|a\rangle) \geq T$ and $\mathcal{C}_\epsilon(|b\rangle) \geq T$ must hold. □

9.3.2 A perfect swapper for subsets of cardinality $K = N/2$

Let us now show that the task of complement sampling, as stated in the beginning of Section 9.3, can be perfectly solved using only a single quantum sample in the case where $K = N/2$.

9.3.7. THEOREM (Quantum complement swapper). *Consider a subset $S \in \mathcal{S}_K$ with $K = N/2$. Then there exists a polynomial-time quantum algorithm which prepares the state $|\bar{S}\rangle$ from $|S\rangle$ and vice versa.*

Proof:
Any function $f : \mathcal{X} \to \{0,1\}$ induces a quantum phase state of the form

$$|y_f\rangle = \frac{1}{\sqrt{N}} \sum_{x \in \mathcal{X}} (-1)^{f(x)} |x\rangle . \tag{9.2}$$

In particular, the constant function $f_{\text{con}}(x) = 0$ yields the state

$$|y_{f_{\text{con}}}\rangle = \frac{1}{\sqrt{N}} \sum_{x \in \{0,1\}^n} |x\rangle . \tag{9.3}$$

For every S and $\bar{S}$ with $K = N/2$ there exists a balanced function f_{bal} such that $f_{\text{bal}}(x) = 0$ if $x \in S$ and $f_{\text{bal}}(x) = 1$ otherwise. Then we can write

$$|S\rangle = \frac{1}{\sqrt{2}} \left(|y_{f_{\text{con}}}\rangle + |y_{f_{\text{bal}}}\rangle \right) ,$$
$$|\bar{S}\rangle = \frac{1}{\sqrt{2}} \left(|y_{f_{\text{con}}}\rangle - |y_{f_{\text{bal}}}\rangle \right) .$$

It is well known that the Deutsch–Jozsa algorithm [DJ92] can perfectly distinguish the two phase states induced by constant and balanced functions. The final measurement in the Deutsch–Jozsa algorithm (which is performed on the phase state resulting from the query) consists of a short quantum circuit V that applies $H^{\otimes n}$ followed by the two-outcome measurement operator $\Lambda = \{|0 \ldots 0\rangle\langle 0 \ldots 0| , \mathbb{I} - |0 \ldots 0\rangle\langle 0 \ldots 0|\}$, of which the outcome is copied to an ancilla qubit by a multi-controlled NOT gate (where the controls are on being in $|0\rangle$). This circuit is illustrated in Fig. 9.2 (a). Then, by Lemma 9.3.5, there exists another quantum circuit V' which only uses two applications of V to swap $|S\rangle$ and $|\bar{S}\rangle$. This circuit is illustrated in Fig. 9.2 (b) where we ignore the ancilla qubit since it starts from and returns to the $|0\rangle$ state. Since V' consists of two applications of V, the multi-controlled NOT gate and two Z gates, it can be executed in time polynomial in n. □

Let us take a closer look at what the circuit used in the proof of Theorem 9.3.7 actually does, fixing the first qubit (so in the register which contains the qubit to

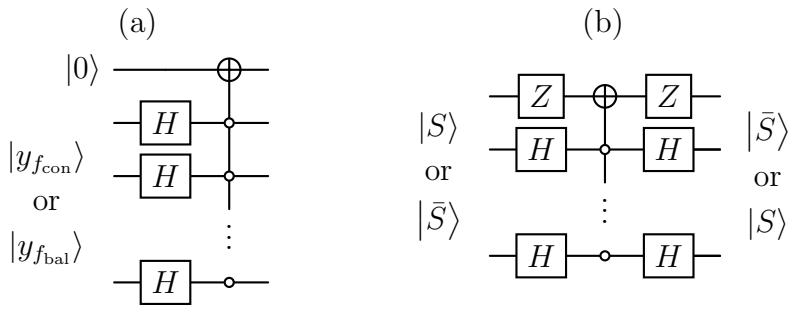


Figure 9.2: Distinguisher and swapper for complement sampling. (a) Our distinguisher circuit can be thought of as the Deutsch–Jozsa algorithm followed by a copy of the measurement outcome into an ancilla qubit. (b) The swapper circuit is obtained by plugging our distinguisher circuit into Fig. 9.1 (b) and rearranging gates. We do not show the ancilla qubit as it starts from and returns to $|0\rangle$. Our swapper circuit corresponds to the Grover diffusion operator up to a global phase of -1.

which the outcome of Λ is "copied") to be in $|0\rangle$ initially. Let the second n-qubit register be in an arbitrary state $|\psi\rangle$. A quick derivation shows that when the first qubit is fixed to be in $|0\rangle$ the circuit implements the mapping

$$|0\rangle\,|\psi\rangle \mapsto -\,|0\rangle\,U\,|\psi\rangle\,,$$

where $U = 2\,|+^n\rangle\langle+^n| - \mathbb{I}$ is the Grover diffusion operator [Gro96]. Noting that $|+^n\rangle = \frac{1}{\sqrt{N}}\left(\sqrt{N-K}\,|\bar{S}\rangle + \sqrt{K}\,|S\rangle\right)$, it can be verified that

$$U\,|S\rangle = 2\sqrt{\frac{K}{N}\left(1-\frac{K}{N}\right)}\,|\bar{S}\rangle + \left(2\frac{K}{N}-1\right)|S\rangle\,.$$

Thus, the Grover diffusion operator is an imperfect swapper with error $\epsilon = \left(2\frac{K}{N}-1\right)^2$. We see that zero error is achieved when S contains exactly half of the elements, $K = N/2$.

9.3.3 Impossibility of a perfect swapper for $K \neq N/2$

We now prove that there exists no perfect swapper for any other case than $K = N/2$, given the condition that all auxiliary qubits must return to their initial state. In Section 9.3.5 we will give a simpler proof which yields the same result (but it is used for different purposes), but we will still include the proof in this section as it uses the construction used to obtain the result for $K = N/2$, providing more intuition into the problem at hand.

9.3.8. Proposition. *Let $S \in \mathcal{S}_K$. Then for any $K \neq N/2$, there does not exist a quantum circuit which perfectly maps $|S\rangle$ to $|\bar{S}\rangle$ under the condition that all auxiliary qubits return to the $|0\rangle$ state.*

Proof:
We will prove our claim using a reductio ad absurdum: suppose that there exists a circuit C which only depends on the cardinality of S (and of course N), and can swap any $|S\rangle$ to $|\bar{S}\rangle$ and vice versa (having $\epsilon = 0$ as per Definition 9.3.3). Again, define the states $|S\rangle$ and $|\bar{S}\rangle$ as

$$|S\rangle = \frac{1}{\sqrt{K}} \sum_{x \in S} |x\rangle\,, \qquad |\bar{S}\rangle = \frac{1}{\sqrt{N-K}} \sum_{x \in \bar{S}} |x\rangle\,,$$

where $\bar{S} = \mathcal{X} \setminus S$. As before, define $|\phi^+\rangle := \frac{1}{\sqrt{2}}\left(|S\rangle + |\bar{S}\rangle\right)$ and $|\phi^-\rangle := \frac{1}{\sqrt{2}}\left(|S\rangle - |\bar{S}\rangle\right)$. Then by Lemma 9.3.5, there exists another circuit C', such that for all sets $S \in \mathcal{S}_K$:

$$\left| \left\| \Pi_1 C' |\phi^+\rangle |0\ldots0\rangle \right\|^2 - \left\| \Pi_1 C' |\phi^-\rangle |0\ldots0\rangle \right\|^2 \right| = 1,$$

which implies that either of the following holds:

$$\left\| \Pi_1 C' |\phi^+\rangle |0\ldots0\rangle \right\|^2 = 1 \text{ (resp. } 0) \qquad \text{and} \qquad \left\| \Pi_1 C' |\phi^-\rangle |0\ldots0\rangle \right\|^2 = 0 \text{ (resp. } 1). \tag{9.4}$$

In the following, we assume that the first case holds (the argument for the second case in parentheses is identical). Now consider two subsets $S_1, S_2 \in \mathcal{S}_K$ and let again $|\phi_i^+\rangle := \frac{1}{\sqrt{2}}\left(|S_i\rangle + |\bar{S}_i\rangle\right)$ and $|\phi^-\rangle := \frac{1}{\sqrt{2}}\left(|S_i\rangle - |\bar{S}_i\rangle\right)$ for $i \in \{1,2\}$. Eq. (9.4) tells us that

$$\left\| \Pi_1 C' |\phi_1^+\rangle |0\ldots0\rangle \right\|^2 = 1 \qquad \text{and} \qquad \left\| \Pi_1 C' |\phi_2^-\rangle |0\ldots0\rangle \right\|^2 = 0, \tag{9.5}$$

which implies that $|\phi_1^+\rangle$ and $|\phi_2^-\rangle$ can be perfectly discriminated. Writing out the states in full, we have

$$|\phi_1^+\rangle = \frac{1}{\sqrt{2}} \left(\frac{1}{\sqrt{K}} \sum_{x \in S_1} |x\rangle + \frac{1}{\sqrt{N-K}} \sum_{x \in \bar{S}_1} |x\rangle \right),$$

and

$$|\phi_2^-\rangle = \frac{1}{\sqrt{2}} \left(\frac{1}{\sqrt{K}} \sum_{x \in S_2} |x\rangle - \frac{1}{\sqrt{N-K}} \sum_{x \in \bar{S}_2} |x\rangle \right).$$

We have that the fidelity between $|\phi_1^+\rangle$ and $|\phi_2^-\rangle$ can be computed as

$$\begin{aligned} \left|\langle \phi_1^+ | \phi_2^- \rangle\right|^2 &= \frac{1}{4} \left| \frac{|S_1 \cap S_2|}{K} - \frac{|S_1 \cap \bar{S}_2|}{\sqrt{K(N-K)}} + \frac{|\bar{S}_1 \cap S_2|}{\sqrt{K(N-K)}} - \frac{|\bar{S}_1 \cap \bar{S}_2|}{N-K} \right|^2 \\ &= \frac{1}{4} \left| \frac{|A_1|}{K} - \frac{|A_2|}{\sqrt{K(N-K)}} + \frac{|A_3|}{\sqrt{K(N-K)}} - \frac{|A_4|}{(N-K)} \right|^2, \end{aligned}$$

where $A_1 := S_1 \cap S_2, A_2 := S_1 \cap \bar{S}_2, A_3 := \bar{S}_1 \cap S_2$ and $A_4 := \bar{S}_1 \cap \bar{S}_2$. A simple Venn diagram shows that $\bigcup_j A_j = \mathcal{X}$ and $\bigcap_j A_j = \emptyset$. Now set $|A_1| = x$, where x should satisfy $\max\{2K - N, 0\} \leq x \leq K$. We can then express the sizes of the other sets A_2, A_3 and A_4 as functions of N, k and x. We have $|A_2| = |S_1 \cap \bar{S}_2| = |S_1 \setminus (S_2 \cap S_1)| = K - x$. A similar argument gives $|A_3| = K - x$. Using the principle of inclusion and exclusion, we have $|S_1 \cup S_2| = |S_1| + |S_2| - A_1 = 2K - x$. Hence, $A_4 = N - |S_1 \cup S_2| = N - 2K + x$. Our expression for the overlap thus becomes

$$\left|\left\langle \phi_1^+ \middle| \phi_2^- \right\rangle\right|^2 = \frac{1}{4}\left|\frac{x}{K} - \frac{N - 2K + x}{(N - K)}\right|^2 = \frac{1}{4}\left(\frac{(2K - N)(K - x)}{K(N - K)}\right)^2. \tag{9.6}$$

Note that Eq. (9.6) becomes 0 when $K = N/2$ (independent of x), as expected. For a fixed $1 \leq K \leq N - 1$, we now want to maximise the overlap as a function of x, i.e. solve

$$F_{\max} = \max_{S_1, S_2} \left|\left\langle \phi_1^+ \middle| \phi_2^- \right\rangle\right|^2 = \max_{\max\{0, 2K-N\} \leq x \leq K} \frac{1}{4}\left(\frac{(2K - N)(K - x)}{K(N - K)}\right)^2. \tag{9.7}$$

The double derivative test gives

$$\frac{\partial^2}{\partial x^2}\frac{1}{4}\left(\frac{(2K - N)(K - x)}{K(N - K)}\right)^2 = \frac{(2K - N)^2}{2(K(N - K))^2} \geq 0,$$

which implies that the function is convex in x. Therefore, the maximum value is obtained at the extremal points of the interval. For simplicity, assume for now that $K \leq N/2$ so that $\max\{0, 2K - N\} = 0$. For $x = K$, the overlap becomes 0 (this is because $S_1 = S_2$ and thus $\left|\phi_1^+\right\rangle$ is orthogonal to $\left|\phi_2^-\right\rangle$), but for $x = 0$ we find that the overlap becomes $\frac{1}{4}\left(\frac{N-2K}{K-N}\right)^2$. Note that this expression is > 0 whenever $K < N/2$ and it is valid (≤ 1) for $K \leq 3N/4$, which is true by assumption. For $K > N/2$ the two extremal points are $x = K$, which again gives overlap 0, and $x = 2K - N$, which gives overlap $\frac{1}{4}\left(2 - \frac{N}{K}\right)^2$. The latter is valid ($\leq 1$) for $K \geq N/4$, which is again true by assumption. Combining both, we find that the maximum overlap as a function of N and K can be

$$\begin{cases} \frac{1}{4}\left(\frac{N-2K}{K-N}\right)^2 & \text{for } 1 \leq K \leq N/2, \\ \frac{1}{4}\left(2 - \frac{N}{K}\right)^2 & \text{for } N/2 < K \leq N - 1. \end{cases}$$

This is symmetric around $K = N/2$, as substituting $K = N - K'$ with $1 \leq K' \leq N-1$ in $\frac{1}{4}\left(2 - \frac{N}{K}\right)^2$ gives back $\frac{1}{4}\left(\frac{N-2K'}{K'-N}\right)^2$, which means that our assumption that $K \leq N/2$ can be made without loss of generality. Therefore, for any $K \neq N/2$ we have that $\left|\phi_1^+\right\rangle$ and $\left|\phi_2^-\right\rangle$ will not be perfectly orthogonal. However, Eq. (9.5) tells us that we can discriminate these two non-orthogonal states perfectly, which is impossible by the Helstrom bound [Hel69]. Hence, there cannot be a perfect

swapper for any other case than $K = N/2$. □

Hence, the intuition of why a perfect swapper is only possible when $K = N/2$ follows from the fact that it can be used to construct a distinguisher for the conjugate states, which are generally non-orthogonal unless $K = N/2$ (where it holds for any choice of $S \in \mathcal{S}_{N/2}$).

9.3.4 Probabilistic zero-error algorithm for any subset

Now that we know a perfect swapper is not possible for $K \neq N/2$, we want to create the "next best thing": a swapper that sometimes fails, but only knowingly so. In the following, we show that one can trade success probability to achieve this zero-error requirement (an algorithm satisfying this property is known as a *Las Vegas algorithm*). The idea is to rebalance the Grover diffusion operator $U = 2\left|+^n\right\rangle\!\left\langle+^n\right| - \mathbb{I}$ by adding a term proportional to $\mathbb{I}$. To this end, we use one auxiliary qubit and the parameterised gate

$$W(q) = e^{i \arccos\left(\sqrt{q}\right) Y} = \begin{pmatrix} \sqrt{q} & -\sqrt{1-q} \\ \sqrt{1-q} & \sqrt{q} \end{pmatrix}, \tag{9.8}$$

to construct the following circuit:

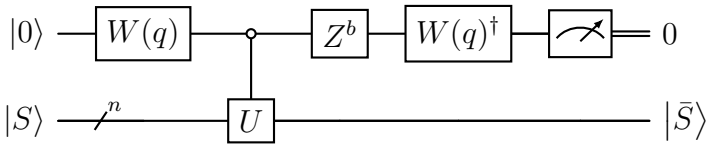


Figure 9.3: Probabilistic zero-error algorithm to swap $|S\rangle$ of cardinality K to $\left|\bar{S}\right\rangle$ of cardinality $N - K$.

Here we have that U is controlled on the zero state of the auxiliary qubit, denoted by C_U, and Z is the Z-gate with some power $b \in \{0, 1\}$. The role of b is to allow for both positive and negative deviations from the optimal case of $K = N/2$.

9.3.9. Theorem (Zero-error swapper). *Let $S \in \mathcal{S}_K$. Then for any $1 \leq K \leq N - 1$, there exists a zero-error (Las Vegas) quantum algorithm that maps $|S\rangle$ to $\left|\bar{S}\right\rangle$, with probability*

$$\frac{\min\{K, N-K\}}{N - \min\{K, N-K\}}.$$

Proof:
We will show that for suitable choices of q and b, the quantum circuit in Fig. 9.3

achieves the desired transformation and corresponding success probability. The state just before the measurement of the auxiliary qubit can be written as[2]

$$W(q)^\dagger Z^b C_U W(q) \left|0\right\rangle \left|S\right\rangle = \left|0\right\rangle \left(qU + (-1)^b(1-q)\mathbb{I}\right)\left|S\right\rangle + \left|1\right\rangle \left|g\right\rangle$$

with $\left|g\right\rangle$ some state we do not care about. Using that $U = 2\left|+^n\right\rangle\!\left\langle+^n\right| - \mathbb{I}$ gives

$$qU + (-1)^b(1-q)\mathbb{I} = 2q\left|+^n\right\rangle\!\left\langle+^n\right| + \left((-1)^b(1-q) - q\right)\mathbb{I}.$$

We will use

$$\left|+^n\right\rangle\!\left\langle+^n\right|\left|S\right\rangle = \frac{K}{N}\left|S\right\rangle + \sqrt{\frac{K(N-K)}{N^2}}\left|\bar{S}\right\rangle,$$

to obtain

$$\left(qU + (-1)^b(1-q)\mathbb{I}\right)\left|S\right\rangle = \left(2q\frac{K}{N} + (-1)^b(1-q) - q\right)\left|S\right\rangle + 2q\sqrt{\frac{K(N-K)}{N^2}}\left|\bar{S}\right\rangle. \tag{9.9}$$

To ensure that $\left(qU + (-1)^b(1-q)\right)\left|S\right\rangle$ is proportional to $\left|\bar{S}\right\rangle$, we set the coefficient of $\left|S\right\rangle$ to zero:

$$2q\frac{K}{N} + (-1)^b(1-q) - q = 0.$$

This gives two valid solutions:

- If $b = 0$, the equation becomes $2q\frac{K}{N} + 1 - 2q = 0$, which gives

$$q = \frac{1}{2(1-K/N)}.$$

- If $b = 1$, the equation becomes $2q\frac{K}{N} - 1 = 0$, which gives

$$q = \frac{N}{2K}.$$

It remains to check when these choices yield a valid success probability[3], i.e., when $\| \left(qU + (-1)^b(1-q)\right)\left|S\right\rangle \|^2 \leq 1$. Using that in Eq. (9.9) we made the amplitude on $\left|S\right\rangle$ zero and the $\left|\bar{S}\right\rangle$-part only depends implicitly on b via q, we obtain

$$\| \left(qU + (-1)^b(1-q)\right)\left|S\right\rangle \|^2 = 4q^2\frac{K(N-K)}{N^2}.$$

[2] A word of caution to the reader working through the details: do not forget that cU is controlled on $\left|0\right\rangle$.

[3] Or equivalently, for what values of q the operator $W(q)$ is indeed unitary.

We find that for the value of q corresponding to $b = 0$, the norm squared is $\frac{K}{N-K}$, which is at most 1 when $K \leq N/2$. When $b = 1$, we obtain $\frac{N-K}{K}$, which is at most 1 when $K \geq N/2$. To summarise, we set

$$\begin{cases} b = 0,\ q = \frac{1}{2(1-K/N)} & \text{if } K < \frac{N}{2}, \\ b = 1,\ q = \frac{N}{2K} & \text{if } K \geq \frac{N}{2}, \end{cases}$$

in Fig. 9.3. The overall success probability is given by the postselection success probabilities in both cases (the norms squared), which can be captured in a single expression as

$$\frac{\min\{K, N-K\}}{N - \min\{K, N-K\}}.$$

□

Let us briefly compare the probabilistic (zero-error) algorithm to directly applying the Grover diffusion operator to $|S\rangle$ (which we will refer to as "with error"). When $K = N/2$, the algorithms are identical, except for the use of an auxiliary qubit in the probabilistic case. For $K \neq N/2$, let us define β as the deviation from the ideal ratio $K/N = 1/2 + \beta$. Then, applying the Grover diffusion operator outputs complementary samples at a rate of $1 - 4\beta^2$ (Fig. 9.4, solid red line), but does not provide information on the correctness of the sample. Instead, the probabilistic algorithm outputs complementary samples at a reduced rate of $\frac{1-2|\beta|}{1+2|\beta|}$ (Fig. 9.4, dashed blue line), with the guarantee that the output is correct. A disadvantage of the probabilistic algorithm is that the swap operation is not symmetric: swapping $|S\rangle$ to $\left|\bar{S}\right\rangle$ and vice versa requires different circuits, making it incompatible with Aaronson, Atia and Susskind's construction from Section 9.3.1.

9.3.5 Zero error: optimality, auxiliary qubits and multiple samples

In this section we will show that the algorithm from Theorem 9.3.9 is in fact optimal when considering its restricted setting: we assume that there is some designated flag qubit, which when measured to be in the $|0\rangle$ state indicates that the state has been successfully created to ensure zero error (putting the flag being in $|0\rangle$ is without loss of generality). Except for this flag qubit, we require that the circuit uses no other extra auxiliary qubits. We prove the following proposition.

9.3.10. Proposition. *For all $1 \leq K \leq N-1$, under the zero-error condition and the use of only one extra ancilla (used as a flag to indicate that the operation was successful), Theorem 9.3.9 is optimal in terms of achievable success probability for swapping $|S\rangle$ to $\left|\bar{S}\right\rangle$.*

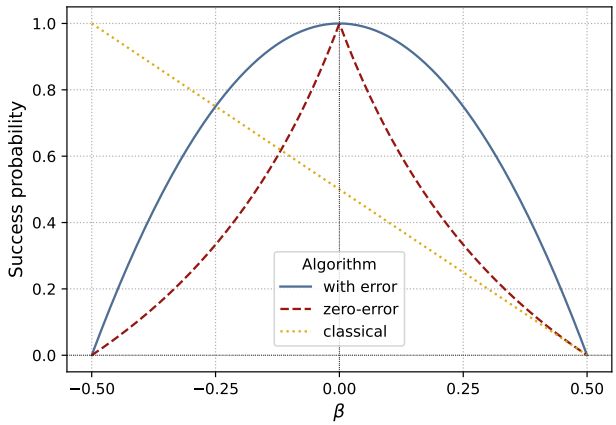


Figure 9.4: Rate at which the proposed algorithms output samples from the complementary set when given only a single sample. The horizontal axis is the deviation β from the ideal ratio, i.e. $K/N = 1/2 + \beta$. When $\beta = 0$, the quantum algorithms (red and blue lines) are exact. When $\beta \neq 0$, the quantum algorithm with error (solid red line) outputs samples at a higher rate, but does not flag the correctness of the sample. The zero-error quantum algorithm (dashed blue line) guarantees samples from the complement while paying a price in terms of success probability. The classical algorithm (dotted green line), achieves a high success rate for large negative deviations (Section 9.4).

Proof:
Splitting the success probability of Theorem 9.3.9 back to two cases, we have the following expression:

$$\begin{cases} \frac{N-K}{K} & \text{if } K \geq \frac{N}{2}, \\ \frac{K}{N-K} & \text{if } K < \frac{N}{2}. \end{cases}$$

For cardinality K, let C_K be an arbitrary circuit which probabilistically swaps $|S\rangle$'s from subsets $S \in \mathcal{S}_K$ to $\left|\bar{S}\right\rangle$. For full generality, we allow the preparation of $\left|\bar{S}\right\rangle$ up to an arbitrary global phase factor $e^{i\theta}$, $\theta \in [0, 2\pi]$. Let $0 \leq \epsilon \leq 1$. To achieve success probability $1 - \epsilon$, C_K should be able to perform a mapping of the form

$$C_K |0\rangle |S_j\rangle = e^{i\theta_j} \sqrt{1-\epsilon} |0\rangle \left|\bar{S}_j\right\rangle + \sqrt{\epsilon} |1\rangle |g_j\rangle$$

for some arbitrary subset $S_j \in \mathcal{S}_K$, phase $\theta_j \in [0, 2\pi]$, and some garbage state $|g_j\rangle$ (which can absorb its own phase factor). Now take two subsets S_1 and S_2 defined as $S_1 = [K]$ and $S_2 = \mathcal{X} \setminus [N-K]$ where, with a slight abuse of notation, we write $[K]$ to indicate the K first strings in $\mathcal{X}$ under lexicographical ordering. We have that the overlap between both input states is given by $|\langle S_1|S_2\rangle|$, and that of both output states after applying C_K is given by

$$\left|(1-\epsilon)e^{i(\theta_2-\theta_1)} \left\langle \bar{S}_1 \middle| \bar{S}_2 \right\rangle + \epsilon \langle g_1|g_2\rangle\right|.$$

We consider the two cases separately:

- $K \geq N/2$. In this case, we have that the inner product between $|S_1\rangle$ and $|S_2\rangle$ is given by (recall that each $|S_j\rangle$ only has positive amplitudes on its computational basis states)

$$\langle S_1|S_2\rangle = \frac{2K-N}{K},$$

but $\langle \bar{S}_1 | \bar{S}_2 \rangle = 0$. Since the inner product must be conserved under unitary transformations, we have

$$\frac{2K-N}{K} = \epsilon |\langle g_1|g_2\rangle|. \tag{9.10}$$

To find an upper bound on the success probability $1-\epsilon$, we can solve the following optimisation problem:

$$\begin{aligned} \max_{\epsilon, |\langle g_1|g_2\rangle|} \quad & 1-\epsilon \\ \text{s.t.} \quad & \frac{2K-N}{K} = \epsilon |\langle g_1|g_2\rangle|, \\ & \epsilon \in [0,1], \\ & |\langle g_1|g_2\rangle| \in [0,1]. \end{aligned}$$

This gives us

$$1-\epsilon \leq 1 - \frac{2K-N}{K} = \frac{N-K}{K}.$$

- $K \leq N/2$. In this case, we make the observation that $C_K^\dagger$ gives us an approximate swapper to go from $|\bar{S}\rangle$ to $|S\rangle$, where the cardinality of $\bar{S}$ is $K' = N-K$ with $K' \geq N/2$. We have that

$$\langle \bar{S}_1 | \bar{S}_2 \rangle = \frac{N-2K}{N-K}$$

Let $|\psi_j\rangle = e^{i\theta_j}\sqrt{1-\epsilon}\,|0\rangle\,|\bar{S}_j\rangle + \sqrt{\epsilon}\,|1\rangle\,|g_j\rangle$. Note that $C_K^\dagger\,|\psi_j\rangle = |0\rangle\,|S_j\rangle$ and $\left|\left(\langle 0|\,\langle \bar{S}_j|\right)|\psi_j\rangle\right| = \sqrt{1-\epsilon}$. Again by conservation of (the absolute value of) the inner product under unitary transformations and the fact that $|S_j\rangle$ and $|\bar{S}_j\rangle$ only contain real amplitudes, we must have that

$$\begin{aligned} \left|\left(\langle 0|\,\langle \bar{S}_j|\right) C_K C_K^\dagger\,|\psi_j\rangle\right| &= \left|\left(\langle 0|\,\langle \bar{S}_j|\right)|0\rangle\, C_K\,|S_j\rangle\right| \\ &= \left|\langle 0|\,\langle S_j|\, C_K^\dagger\left(|0\rangle\,|\bar{S}_j\rangle\right)\right| \\ &= \sqrt{1-\epsilon} \end{aligned}$$

which means we can write the action of $C_K^\dagger$ on $|0\rangle \left|\bar{S}_j\right\rangle$ to be of the form

$$C_K^\dagger |0\rangle \left|\bar{S}_j\right\rangle = e^{i\bar{\theta}_j}\sqrt{1-\epsilon}\,|0\rangle\,|S_j\rangle + \sqrt{\epsilon}\,|1\rangle\,|\bar{g}_j\rangle\,,$$

again for some garbage state $|\bar{g}_j\rangle$ and some phase $\bar{\theta}_j \in [0, 2\pi]$. By the same argument as the case for $K \geq N/2$ using K' instead of K, we find

$$1 - \epsilon \leq 1 - \frac{N-2K}{N-K} = \frac{K}{N-K}.$$

Hence, this shows that under the conditions of the statement Theorem 9.3.9 is optimal. □

The above proof is an arguably easier proof of the result in Section 9.3.3, as it only relies on the conservation of the inner product under unitary transformations. There are several more observations we can make in relation to the proof of Proposition 9.3.10.

Adding extra auxiliary qubits. The proof of Proposition 9.3.10 shows that, under the condition of having only one additional ancilla to work as flag qubit, the algorithm of Section 9.3.4 achieves the optimal success probability. What happens when extra auxiliary qubits are allowed (initialised in $|0\ldots0\rangle$), but we still require a single flag qubit to make the algorithm zero-error (yet still probabilistic)?

In the $K \geq N/2$ case, it is easy to show that the above proof still holds. For simplicity, we remove the relaxation that the state can be prepared up to an arbitrary global phase, as it does not change the argument. We can write the action of the swapping circuit in this case as

$$C_K\,|0\rangle\,|S_j\rangle\,|0\ldots0\rangle = \sqrt{1-\epsilon}\,|0\rangle\left|\bar{S}_j\right\rangle|h_j\rangle + \sqrt{\epsilon}\,|1\rangle\,|G_j\rangle\,,$$

for some garbage states $|h_j\rangle$ and $|G_j\rangle$, so we still have that the inner product of the output state after applying C_K to two input subset states $|S_1\rangle$ and $|S_2\rangle$ is given by

$$(1-\epsilon)\left\langle\bar{S}_1\middle|\bar{S}_2\right\rangle\langle h_1|h_2\rangle + \epsilon\,\langle G_1|G_2\rangle = \epsilon\,\langle G_1|G_2\rangle\,,$$

again using that $\left\langle\bar{S}_1\middle|\bar{S}_2\right\rangle = 0$ when $K \geq N/2$.

However, the argument for $K \leq N/2$ no longer works since $|h_1\rangle$ and $|h_2\rangle$ might be two completely different states. This means that to swap backwards using $C_K^\dagger$ one might have to use a different initial state for each different input $\bar{S}$.

An easy example that contradicts the bound in the case when extra auxiliary qubits can be used is the case of $K = 1$ (where the bound gives a maximum success probability very close to 0). An algorithm that only depends on K and perfectly makes $\left|\bar{S}\right\rangle$ is easily given: (i) read out the only element $x \in S$, (ii)

create the uniform superposition, (iii) use an additional flag qubit to mark the state $|\bar{S}\rangle$, which can be done by applying a marking operation that flips the phase of all basis states except $|x\rangle$, and (iv) use exact Grover's algorithm to create $|\bar{S}\rangle$ [Zal99]. This algorithm clearly needs additional auxiliary qubits, as it cannot both have a uniform superposition on n qubits and a flag controlled on $|x\rangle$ (which also needs to be stored in a n-qubit register). Note that in this case, the problem is also trivial classically, because after observing a single sample one has perfectly learned $\bar{S} = \mathcal{X} \setminus \{x\}$. This is shown in Fig. 9.4, where the classical algorithm (dotted green line) achieves success probability 1.

Multiple copies and no additional auxiliary qubits. The above argument also shows that a joint measurement on multiple copies does not help if one does not have additional auxiliary qubits. That is, suppose one again requires a probabilistic zero-error algorithm that is able to perform the mapping

$$C_K |0\rangle |S_j\rangle^{\otimes k} = \sqrt{1-\epsilon}\,|0\rangle |\bar{S}_j\rangle |h_j\rangle + \sqrt{\epsilon}\,|1\rangle |G_j\rangle ,$$

using k copies of the input state. Considering $K \geq N/2$, the same argument as before gives us a maximum success probability of

$$1 - \epsilon \leq 1 - \left(\frac{2K-N}{K}\right)^k ,$$

which is the same success probability one would achieve by repeating our algorithm k times, resetting the flag qubit to $|0\rangle$ after each failed attempt (i.e., when $|1\rangle$ is measured). We currently do not know whether a joint measurement on multiple copies and the use of extra auxiliary qubits would yield an algorithm with a higher success probability.

9.4 Classical samples: lower and upper bounds

9.4.1 Index query model

We will start by adopting a slightly more powerful classical setting, as it allows for easier proofs and it will be useful later when we construct S from pseudorandom permutations, which naturally assumes a query complexity setting. Instead of assuming that the algorithm is given samples from S at random, we assume that it has access to an oracle for which it can query elements of S (assuming lexicographical ordering on the set S). That is, it has access to an oracle O_{index} such that $O_{\text{index}}(i) = y$ with y the ith element from S. Clearly, this is a stronger access model, as the player can generate the random samples by querying O_{index} for randomly generated indices.

Intuitively, one would immediately expect that complement sampling with classical samples is intractable: when there is no structure to the set S (and

therefore no structure to $\bar{S}$), the best strategy seems to be to just gather (or in this case, query) as many distinct samples from S and just output anything which is not in the set of collected samples. We will now formalise this intuition and show that it is indeed the correct way to look at the problem. We will use Yao's minimax principle [Yao77] to prove our lower bound. For a fixed input size N and cardinality K, let $F(x, y)$ be a function describing the relation for a subset S from a subset family $\mathcal{S}_K$ and candidate element y such that

$$F(S, y) = \begin{cases} 1 \text{ if } y \in \bar{S} \\ 0 \text{ otherwise.} \end{cases}$$

Let A be the set of all possible deterministic algorithms allowed to make T queries to the index oracle for S. Write $P(a, S) = F(S, a(S))$, which means that $P = 1$ if and only if algorithm $a \in A$ on input S outputs any $y \in \bar{S}$. One can view $P(a, S)$ as a $|A| \times |\mathcal{S}_K|$- matrix where the rows label the different choices of deterministic query algorithms, the columns label the different instances of S, and the corresponding entry is 1 only if the algorithm provides an output satisfying the relation. By the minimax theorem, we have

$$\min_{\mu} \max_{a} e_a^\intercal P \mu = \max_{\rho} \min_{S} \rho^\intercal P e_S,$$

where e_a (resp. e_S) is a $|A|$-dimensional (resp. $|\mathcal{S}_K|$-dimensional) unit vector, and ρ (resp. μ) is a $|A|$-dimensional (resp. $|\mathcal{S}_K|$-dimensional) vector of non-negative reals that sum to one. Since μ is a probability distribution, we have that the left-hand side describes the probability that the best deterministic T-query algorithm is correct on the hardest distribution over inputs μ. On the right-hand side, $\rho^\intercal P e_S$ is the success probability on input S achieved by the randomised algorithm given by probability distribution ρ over deterministic algorithms. Hence, the right-hand side gives the highest worst-case success probability achievable by randomised T-query algorithms. Since this holds for all T, we have that

$$R_\epsilon(F) = \max_{\mu} D_\epsilon^\mu(F),$$

where R_ϵ is the worst-case randomized query complexity, achieving success probability $1-\epsilon$ on all inputs, and D_ϵ^μ the deterministic query complexity on succeeding on a $(1-\epsilon)$-fraction over all inputs weighted by μ. Hence, for any candidate distribution μ, we always have

$$R_\epsilon(F) \geq D_\epsilon^\mu(F).$$

We first prove the following lemma, showing that when starting with the uniform distribution over all families of subsets $\mathcal{S}_K$ (and thus also a uniform distribution over complementary subsets $\bar{S}$), learning entries of S will leave the resulting conditional distribution over the family of complementary subsets that do not include these elements still uniform.

9.4.1. LEMMA. *Let $\mu_{\mathcal{S}_K}$ be the uniform distribution over all $S \in \mathcal{S}_K$. Let $X \in \mathcal{S}_K$ be sampled according to $\mu_{\mathcal{S}_K}$, and define $\bar{X} = \mathcal{X} \setminus X$. Fix a set $Q = \{x_1, \ldots, x_q\}$ containing $q \leq K$ distinct elements from $\mathcal{X}$. Then, conditioned on $Q \subseteq X$, the distribution of $\bar{X}$ is uniform over $\mathcal{S}^Q_{N-K}$, the family of all subsets of size $N-K$ that do not contain any element from Q.*

Proof:
By Bayes' rule,

$$\Pr\big[\bar{X} = \bar{S} \mid x_1, \ldots, x_q \in X\big] = \frac{\Pr\big[\bar{X} = \bar{S} \wedge x_1, \ldots, x_q \in X\big]}{\Pr[x_1, \ldots, x_q \in X]}.$$

Using that

$$\Pr[x_1, \ldots, x_q \in X] = \frac{\binom{N-q}{K-q}}{\binom{N}{K}},$$

and the fact that

$$\Pr\big[\bar{X} = \bar{S} \wedge x_1, \ldots, x_q \in X\big] = \begin{cases} \binom{N}{N-K}^{-1} & \text{if } x_1, \ldots, x_q \notin \bar{S}, \\ 0 & \text{otherwise,} \end{cases}$$

we obtain

$$\Pr\big[\bar{X} = \bar{S} \mid x_1, \ldots, x_q \in X\big] = \begin{cases} \binom{N-q}{N-K}^{-1} & \text{if } x_1, \ldots, x_q \notin \bar{S}, \\ 0 & \text{otherwise,} \end{cases}$$

which is uniform over all $\bar{S} \in \mathcal{S}^Q_{N-K}$. □

We can now use Yao's principle to derive our lower bound.

9.4.2. THEOREM. *Let $\delta \in [0, \frac{1}{2}]$ and $1 \leq K \leq N-1$. Then we have that any randomised algorithm that returns an element from $\bar{S}$ with probability $\geq 1/2 + \delta$ on all inputs $S \in \mathcal{S}_K$ needs to make at least*

$$N - \frac{2(N-K)}{2\delta + 1}$$

queries to the index oracle for S.

Proof:
Consider $\mu_{\mathcal{S}_K}$ to be the uniform distribution over all possible inputs S coming from the family $\mathcal{S}_K$. Let $X \in \mathcal{S}_K$ be sampled according to $\mu_{\mathcal{S}_K}$. Let A be a deterministic algorithm with access to the index oracle O_{index}. We can assume that no element of X is queried twice by A, as there would be no benefit in

doing so. Then, after q queries, the algorithm has seen q elements of X. Let $Q = \{x_1, \ldots, x_q\}$ be the set of all strings $x_j \in \{0,1\}^n$ that were observed, and write $\mathcal{S}^Q_{N-K}$ for the family of all sets $\bar{S}$ that do not contain the strings $x_j \in Q$. By Lemma 9.4.1, we have that the conditional distribution $\mu_{\mathcal{S}^Q_{N-K}}$ over $\mathcal{S}^Q_{N-K}$ is still uniform over these elements. The cardinality of the set $\mathcal{S}^Q_{N-K}$ is then given by

$$\left|\mathcal{S}^Q_{N-K}\right| = \binom{N-q}{N-K}.$$

So for any output y, the probability that it is part of $\bar{S}'$ for some $\bar{S}' \in \mathcal{S}^Q_{N-K}$ is then given by

$$\frac{\binom{N-q-1}{N-K-1}}{\binom{N-q}{N-K}} = \frac{N-K}{N-q}, \tag{9.11}$$

which holds independently of the strings that are in Q (except for how many of them are there). Hence, A can only be correct on any such fraction of the inputs. By Yao's principle, this is then also the best success probability a q-query randomised algorithm achieves on the worst-case input. To be correct on at least $\frac{1}{2} + \delta$ of the inputs, we require $(N-K)/(N-q) \geq \frac{1}{2} + \delta$, which is satisfied when

$$q \geq N - \frac{2(N-K)}{2\delta + 1}.$$

□

We proceed by giving a matching upper bound, showing that Theorem 9.4.2 is tight.

9.4.3. Proposition. *For any subset $S \in \mathcal{S}_K$ and any $\delta \in [0, \frac{1}{2}]$, there exists a randomized algorithm which makes $N - \frac{2(N-K)}{2\delta+1}$ queries to O_{index} and outputs an element $y \in \bar{S}$ with probability $\frac{1}{2} + \delta$.*

Proof:
Again, let $Q = \{x_1, \ldots, x_q\}$ be the set of strings $x_j \in S$ that are observed after q queries. The algorithm will now simply sample a uniformly random y from the set $\mathcal{X} \setminus Q$. Given any S, the probability that this y is from $\bar{S}$ is then given by

$$\Pr\left[y \in \bar{S}\right] = \frac{N-K}{N-q} = \frac{N-K}{N - (N - \frac{2(N-K)}{2\delta+1})} = \frac{1}{2} + \delta.$$

□

Note that if $K = \Omega(N)$ and $\delta = \Omega(1)$, then we have that $q = \Omega(N)$.

9.4.2 Exact sample complexity bounds

In the above, we used an index query model because it simplifies the analysis while providing lower bounds for the sample complexity setting (as it is a stronger access model). However, it is possible to translate these bounds to the sample complexity setting by computing the probability of observing q unique samples after drawing d samples. Since the results of Section 9.4.1 hold in terms of *unique* observations of elements from S, we can exactly compute matching upper and lower bounds on the sample complexity.

Sampling from a subset of size K, the probability of getting $q \leq d$ unique samples from d draws (with replacement, each with equal probability) is given by (see, for example, [MZHO16] for a derivation)

$$\Pr[X = q] = \frac{K \left\{ {d \atop q} \right\}}{(K-q)! K^d},$$

where $\left\{ {d \atop q} \right\}$ is the Stirling number of the second kind, representing the number of ways to partition d objects into q non-empty subsets. Combining this with Eq. (9.11), which gives the probability of outputting a successful sample conditioned on observing q unique elements, gives a lower bound on the overall success probability of

$$\sum_{q=0}^{q=d} \frac{N \left\{ {d \atop q} \right\}}{(K-q)! K^d} \frac{N-K}{N-q}.$$

This in principle gives the exact expression and has a matching upper bound by the same argument as the previous subsection. However, since the expression is cumbersome to work with and we are interested in proving asymptotic lower bounds, we will continue with the query model in the following sections to consider more practically relevant notions of hardness of the problem.

9.4.3 Average-case lower bounds

A consequence of Theorem 9.4.2 is that it also shows that the problem is hard on average (with respect to the uniform distribution), as a worst-to-average case reduction with respect to the uniform distribution can easily be constructed. We will use that applying a random permutation on N elements as an operation to any fixed subset of K out of the N elements will result in a uniformly random subset of K elements.

9.4.4. Corollary. *Let $S \in \mathcal{S}_K$, $K = N/2$, be sampled according to $\mu_{\mathcal{S}_K}$. Then any classical randomized algorithm A which makes q queries to O_{index} and satisfies*

$$\Pr\left[A \text{ outputs } y \in \bar{S}\right] \geq \frac{1}{2} + \delta \tag{9.12}$$

has

$$q \geq N\left(1 - \frac{1}{2\delta + 1}\right),$$

where the probability in Eq. (9.12) *is taken over* S *and the randomness of* A*.*

Proof:
Let σ be a permutation on $\mathcal{X}$ chosen uniformly at random, and write σ^{-1} for its inverse. Starting from any $S = \{x\} \in \mathcal{S}_{N/2}$, $\bar{S} = \{y\} = \mathcal{X} \setminus S$, we have that the subset S' and its complement $\bar{S}' = \mathcal{X} \setminus S'$ given by

$$S' = \{x' \mid x' = \sigma(x), x \in S\}, \qquad \bar{S}' = \{y' \mid y' = \sigma(y), y \in \bar{S}\},$$

are uniformly at random from all possible $S \in \mathcal{S}_{N/2}$. Since we are only interested in query (or sample) complexity, we do not care that the time complexity of implementing this permutation σ, nor its inverse σ^{-1}, generally scales exponentially in n (recall $n = \log N$). Now suppose there exists a q-query classical algorithm A that succeeds with expected probability at least $\frac{1}{2} + \delta$ over this distribution of inputs. We can then use A to create a q-query algorithm B that works with a certain success probability on all inputs in the following way:

1. Pick a permutation σ uniformly at random and let σ^{-1} be its inverse.

2. Let $O'_{\text{index}}(i) = \sigma(O_{\text{index}}(i))$. Run algorithm A with $O'_{\text{index}}(i)$, resulting in some y'.

3. Return $y = \sigma^{-1}(y')$.

We then have that B succeeds on any input S with probability at least $\frac{1}{2} + \delta$ as well, since

$$\Pr\left[B \text{ outputs } y \in \bar{S}\right] = \mathbb{E}_{S'}\left[\Pr\left[A \text{ outputs } y' \in \bar{S}'\right]\right] \geq \frac{1}{2} + \delta,$$

using that the expectation of a probability is again a probability. By Theorem 9.4.2, we then need at least

$$q \geq N\left(1 - \frac{1}{2\delta + 1}\right)$$

queries to O_{index}, completing the proof. □

9.4.4 Hardness for strong pseudorandom subsets

We will now prove our strongest hardness result: we will show that under the existence of one-way functions, there should be families of subsets that are easy to generate and verify, but still hard to perform complement sampling on classically. To achieve this, we will use strong pseudorandom permutations. Following the convention in cryptography, we say that a function f is negligible ($f(n) = \text{negl}(n)$), if for every constant c, we have $f(n) = o\left(\frac{1}{n^c}\right)$.

9.4.5. Definition (Strong pseudorandom permutations). For some $n \in \mathbb{Z}_+$, let $P : \{0,1\}^n \times \{0,1\}^\lambda \to \{0,1\}^n$ be a family of efficiently computable permutations indexed by a key $k \in \{0,1\}^\lambda$. We say that P is a *strong pseudorandom permutation* if for all probabilistic polynomial-time distinguishers D, there exists a negligible function $\text{negl}(\cdot)$ such that

$$\left|\Pr\left[D^{P_k(\cdot),P_k^{-1}(\cdot)}(1^n)\right] - \Pr\left[D^{\sigma(\cdot),\sigma^{-1}(\cdot)}(1^n)\right]\right| \leq \text{negl}(\lambda), \tag{9.13}$$

where $k \leftarrow \{0,1\}^\lambda$ is chosen uniformly at random, and σ is chosen uniformly at random from the set of permutations on $\{0,1\}^n$.

It is known that strong pseudorandom permutations can be constructed from pseudorandom permutations [LR88], which exist if and only if one-way functions exist [KL20].

We will show the hardness result for cardinality $K = N/2$, as it forms the basis of our proposed advantage experiment. However, the argument can be easily extended to any K superpolynomial in N.

9.4.6. Theorem. *For some $n \in \mathbb{Z}_+$, let $P : \{0,1\}^n \times \{0,1\}^\lambda \to \{0,1\}^n$ be a strong pseudorandom permutation with security parameter $\lambda = n$, and P^{-1} its inverse. Given a key $k \in \{0,1\}^\lambda$, let S be the image of $\{0ik : i \in \{0,1\}^{n-1}\}$ under P, and $\bar{S}$ be the image of $\{1ik : i \in \{0,1\}^{n-1}\}$ under P. Let $O_{\text{index}} : [2^{n-1}] \to \{0,1\}^n$ be the oracle that on input i returns the i-th element in S. Picking a key uniformly at random, for all polynomial-time algorithms A that make a polynomial number of queries to O_{index} it must hold that*

$$\Pr\left[A^{O_{\text{index}}} \text{ outputs a } y \in \bar{S}\right] \leq \frac{1}{2} + \text{negl}(n).$$

Proof:
Suppose that there exists a polynomial-time algorithm A with query access to some O_{index}, which, when picking a key uniformly at random to construct S, outputs a $y \in \bar{S}$ with probability $\geq \frac{1}{2} + 1/\text{poly}(n)$. We show that if such aa A

existed, it would imply a distinguisher between a strong pseudorandom permutation and a truly random one, violating the assumption that S was generated by a strong pseudorandom permutation. First, note that defining S to be the set of all outputs of a uniformly random permutation for which the preimage has "0" as the first bit ensures that S is chosen uniformly at random from $\mathcal{S}_{N/2}$. Second, note that any given P already acts as O_{index} when applied to strings from $\{0ik : i \in \{0,1\}^{n-1}\}$. Let F be a samplable family of either (i) permutations or (ii) strong pseudorandom permutations. Then, we can design a distinguisher using $A^{O_{\text{index}}}$ as a subroutine in the following way:

1. We sample a permutation $f \in F$, which is either strong pseudorandom or truly random, and construct O_{index} from f.

2. We run $A^{O_{\text{index}}}$ to obtain a candidate string $\hat{y}$, supposedly from the complement $\bar{S}$.

3. We check whether $f^{-1}(\hat{y})$ has the first bit equal to "1". If this is the case, we return ✓. If not, we return ✗.

Let $\delta = 1/\sqrt{N}$, $N = 2^n$, such that $\delta = \text{negl}(n)$. Then, Corollary 9.4.4 readily implies that whenever $q = \text{poly}(n)$ and S is uniformly random (i.e. f comes from a family of truly random permutation), then any q-query classical algorithm (even without the requirement that it runs in polynomial time) succeeds with probability smaller than $\frac{1}{2} + \delta = \frac{1}{2} + \text{negl}(n)$. However, if f comes from a family of strong pseudorandom permutations, then by assumption it must hold that we observe a ✓with probability $\frac{1}{2} + 1/\text{poly}(n)$. Hence, under this assumption, we can distinguish strong pseudorandom permutations from truly random ones by repeating the above distinguisher a polynomial number of times to detect the small polynomial bias towards returning ✓in the pseudorandom case. Since this is not possible by the definition of strong pseudorandomness, such an A cannot exist. Thus, the statement must be true. □

The above argument does not extend to pseudorandom permutations that are not strong, as we need to use P^{-1} to check whether the generated string is indeed from $\bar{S}$.

9.4.5 Circuit lower bounds for uniform samplers

Finally, we conclude our classical hardness results by showing that we can even prove an exponential *circuit lower bound* on any sampler (quantum or classical) that uses only a polynomial number of classical samples and produces a uniformly random sample from $\bar{S}$ with probability 1. Unlike the previous lower bounds, which hold even when the uniformly random criterion is removed, this bound relies crucially on the requirement that every element from $\bar{S}$ has a non-zero

probability to be produced by the sampler. The key idea is to use the Kolmogorov complexity of binary strings to arrive at a circuit lower bound.

9.4.7. Definition. The Kolmogorov complexity $\mathcal{K}(x)$ of any binary string $x \in \{0,1\}^*$ is the length of the shortest computer program x^* that can produce this string on the Universal Turing Machine and then halts.

It is easy to show that there must exist incompressible strings (so with $\mathcal{K}(x) \geq |x|$) for every input length n: there are 2^n binary strings of length n but only $2^n - 1$ binary strings of length strictly less than n to describe the program.

We will formulate our bound in terms of the circuit complexity of a Boolean circuit. Since a circuit using a universal gate set of a constant number of logic gates having size s can be described by at most $\mathcal{O}(s \log s)$ bits, any circuit of size s can be converted into a program x^* for the Universal Turing Machine of length at most $|x^*| = \mathcal{O}(s \log s)$. Hence, if a string has Kolmogorov complexity $\mathcal{K}(x)$ as per Definition 9.4.7, then the size of the smallest possible circuit has $s = \Omega\left(\mathcal{K}(x)/\log(\mathcal{K}(x))\right)$.

We will use this to show a circuit lower bound on any classical sampler that uses only a polynomial number of samples and can sample uniformly at random from the complementary subset $\bar{S}$. We stress that this is a stronger setting than we have considered so far classically, but we argue that it is still valid, since the quantum algorithm can indeed provide a single classical sample uniformly at random.

9.4.8. Theorem (Circuit complexity lower bound with classical samples). *There exists an $S \in \mathcal{S}_K$, with $K = N/2$, such that any sampler, which given as an input $l = \mathrm{poly}(n)$ samples from S produces samples y from $\bar{S}$ uniformly at random, has circuit complexity $\tilde{\Omega}(N)$.*

Proof:
We will simply assume that all l samples are distinct, since this not being true only increases the lower bound that follows from our argument. Take a subset $\bar{S}$ with Kolmogorov complexity $\mathcal{K}(\bar{S}) = N/2$. Such a set can be constructed by taking a Kolmogorov random string $z \in \{0,1\}^{N/2}$ of length $N/2$, with each element z_i indexed by a string $i \in \{0,1\}^{n-1}$. We define $y_i = (z_i, i)$ as the concatenation of bit z_i and the string i. For each $i \in \{0,1\}^{n-1}$, we add the string y_i to the set $\bar{S}$. This way, one can fully reconstruct z if one knows all the strings in $\bar{S}$. Then, any sampler M that produces samples from $\bar{S}$ uniformly at random, given the description of any l elements from S, is a description of the set $\bar{S}$ (and thus of the string z), since one can run the sampler until all $N/2$ distinct labels from $\bar{S}$ are observed. Since l elements from S can be described with $l \log_2 N$ bits, the description length of the program for M given l such elements has to be at least $N/2 - l \log_2 N - \mathcal{O}(1)$. Since any circuit description (of the elementary gates) of

size s can represent a program for M in $\mathcal{O}(s \log s)$ bits, we must have that the circuit complexity of any such circuit which implements M is lower bounded by

$$\Omega\left(\frac{N/2 - l\log_2 N - \mathcal{O}(1)}{\log\left(N/2 - l\log_2 N - \mathcal{O}(1)\right)}\right) = \tilde{\Omega}\left(N\right),$$

when $l = \text{poly}(n)$. □

We believe the Kolmogorov complexity argument of Theorem 9.4.8 is particularly insightful of *why* complement sampling works in a quantum setting, as it highlights the key difference between quantum and classical samples. In the process of providing a sample from the complement, our quantum algorithm destroys the quantum sample, meaning that it needs a new input sample for every generated output sample. On the other hand, classical samples can be reused indefinitely as they can be cloned.

9.5 Circuit complexity and distinguishability of quantum samples

In the final section we will show that even though we have proven in the above sections that a quantum algorithm can easily perfectly swap between any $|S\rangle$ and $\left|\bar{S}\right\rangle$, there exist pairs of $|S\rangle$ and $\left|\bar{S}\right\rangle$ that cannot be distinguished by any efficient quantum circuit. As an immediate corollary, this result implies that generally states of the form of $|S\rangle$ must have exponential circuit complexity. This means that in order to turn complement sampling into a suitable quantum advantage experiment, the hardness result for strong PRPs is indeed necessary. We will first prove the easier case of the exact setting and then move on to the general case. Again, we fix the cardinality to be $K = N/2$ where $N = 2^n$ and n is the number of qubits.

9.5.1 Warm-up: exact case

As a warm-up, we first consider the simple example where we care about being able to distinguish perfectly. Consider again the states

$$|y_{f_{\text{con}}}\rangle = \frac{1}{\sqrt{N}}\sum_{x\in\mathcal{X}}|x\rangle, \qquad |y_{f_{\text{bal}}}\rangle = \frac{1}{\sqrt{N}}\sum_{x\in\mathcal{X}}(-1)^{f_{\text{bal}}(x)}|x\rangle, \tag{9.14}$$

for some Boolean function $f_{\text{bal}} : \mathcal{X} \to \{0,1\}$ that is balanced. Note that for all such f_{bal}, we have that

$$\langle y_{f_{\text{con}}}|y_{f_{\text{bal}}}\rangle = \frac{1}{N}\sum_{x\in\mathcal{X}}(-1)^{f_{\text{bal}}(x)} = 0.$$

The total number of balanced functions is lower bounded by

$$\binom{N}{N/2} \geq 2^{\Omega(N)}.$$

Now, let $\mathcal{G}$ be any gate set with $g = n^{\mathcal{O}(1)}$ possible choices of gates and qubit indices on which a single (multi-qubit) gate acts. Starting from $|y_{f_{\text{con}}}\rangle$, the total number of states that can be constructed by circuits using M gates from $\mathcal{G}$ is at most g^M. Thus, for some balanced function, the corresponding circuit must have $M = \tilde{\Omega}(N)$ gates, which is the relative complexity of $|y_{f_{\text{con}}}\rangle$ and $|y_{f_{\text{bal}}}\rangle$.

For every choice of f_{bal}, a circuit can only *exactly* swap $|y_{f_{\text{con}}}\rangle$ to a single $|y_{f_{\text{bal}}}\rangle$. Since the swap complexity is lower bounded by the relative complexity [AAS20], there must exist a $|y_{f_{\text{bal}}}\rangle$ such that

$$\mathcal{S}(|y_{f_{\text{con}}}\rangle, |y_{f_{\text{bal}}}\rangle) \geq \tilde{\Omega}(N). \tag{9.15}$$

By Lemma 9.3.5, we have that if the states $|\phi^+\rangle$ and $|\phi^-\rangle$, given by

$$\left|\phi^+\right\rangle = \frac{1}{\sqrt{2}}(|y_{f_{\text{con}}}\rangle + |y_{f_{\text{bal}}}\rangle) = \frac{1}{\sqrt{N/2}} \sum_{x\in S} |x\rangle =: |S\rangle,$$
$$\left|\phi^-\right\rangle = \frac{1}{\sqrt{2}}(|y_{f_{\text{con}}}\rangle - |y_{f_{\text{bal}}}\rangle) = \frac{1}{\sqrt{N/2}} \sum_{x\in \bar{S}} |x\rangle =: \left|\bar{S}\right\rangle,$$

where

$$S = \{x : f_{\text{bal}}(x) = 0\}, \qquad \bar{S} = \{x : f_{\text{bal}}(x) = 1\},$$

have exact distinguishing complexity $\mathcal{D}(|\phi_+\rangle, |\phi_-\rangle) = T$, then $\mathcal{S}(|y_{f_{\text{con}}}\rangle, |y_{f_{\text{bal}}}\rangle) = \mathcal{O}(T)$. Hence, from Eq. (9.15) it follows that

$$\mathcal{D}(|\phi_+\rangle, |\phi_-\rangle) \geq \tilde{\Omega}(N).$$

9.5.2 Approximate case

We consider the same $|y_{f_{\text{con}}}\rangle$ and $|y_{f_{\text{bal}}}\rangle$ as in Eq. (9.14), but will now investigate the relative circuit complexity when an error ϵ is allowed. Let us consider two balanced functions $f_{\text{bal}}, g_{\text{bal}} : \mathcal{X} \to \{0,1\}$. From now on, we omit the "bal"-subscript when possible, and write $f = f_{\text{bal}}$ and $g = g_{\text{bal}}$ to ease the presentation. We define $I_{f,g} = |\{x : f(x) = g(x)\}|$ and $N_{f,g} = |\{x : f(x) \neq g(x)\}|$. The overlap between $|y_f\rangle$ and $|y_g\rangle$ is given by

$$|\langle y_f|y_g\rangle| = \frac{1}{N}|I_{f,g} - N_{f,g}| = \frac{1}{N}|2I_{f,g} - N|,$$

using the relation $I_{f,g} + N_{f,g} = N$. If we want $|\langle y_f|y_g\rangle| \leq \epsilon$, then we must have that

$$\frac{1}{N}|2I_{f,g} - N| \leq \epsilon.$$

We consider two cases:

(i) $I_{f,g} \geq \frac{N}{2}$. We find

$$\frac{1}{N}(2I_{f,g} - N) \leq \epsilon \quad \text{if and only if} \quad I_{f,g} \leq \frac{(1+\epsilon)N}{2}.$$

(ii) $I_{f,g} \leq \frac{N}{2}$. Likewise, we have

$$\frac{1}{N}(N - 2I_{f,g}) \leq \epsilon \quad \text{if and only if} \quad I_{f,g} \geq \frac{(1-\epsilon)N}{2}.$$

Combining cases (i) and (ii), we obtain a necessary and sufficient condition for $|\langle y_f|y_g\rangle| \leq \epsilon$ is that

$$\frac{(1-\epsilon)N}{2} \leq I_{f,g} \leq \frac{(1+\epsilon)N}{2}. \tag{9.16}$$

The question now is how large the set of functions $\{f\}$ is such that for any pair $f \neq g$ we have that Eq. (9.16) holds. We can formulate this as a problem in extremal set theory in the following way. Let $\mathcal{F} = \{F\}$ be a family containing $N/2$-subsets from $\mathcal{X}$. We say that if $x \in F$, then a corresponding function f satisfies $f(x) = 1$. For example, if for $n = 2$ we have $f(00) = 1$, $f(01) = 0$, $f(10) = 0$ and $f(11) = 1$, then $F = \{00, 11\}$. Hence, if $|F \cap G| = l$ for some $F, G \in \mathcal{F}$ with associated functions f, g, respectively, we must have that by definition

$$|\{x : f(x) = g(x) = 1\}| = l.$$

However, we also know that for both F and G the remaining $N/2 - l$ elements in each set must all be distinct: for those we have either $f(x) = 1$ and $g(x) = 0$ or $f(x) = 0$ and $g(x) = 1$. For all remaining x which have $f(x) = g(x) = 0$, i.e., $x \notin F$ and $x \notin G$, we know that there are a total of

$$|\{x : f(x) = g(x) = 0\}| = N - l - 2(N/2 - l) = l.$$

Combining both, we have

$$I_{f,g} = |\{x : f(x) = g(x) = 0\}| + |\{x : f(x) = g(x) = 1\}| = 2l.$$

Therefore, the intersection between two sets F and G completely determines the possible overlaps between the corresponding quantum states $|y_f\rangle$ and $|y_g\rangle$: we have by condition Eq. (9.16) that if for all $F, G \in \mathcal{F}$

$$\frac{(1-\epsilon)N}{4} \leq |F \cap G| \leq \frac{(1+\epsilon)N}{4}, \tag{9.17}$$

then for all corresponding f, g we have that $|\langle y_f|y_g\rangle| \leq \epsilon$.

We will now use results from extremal set theory to prove our circuit lower bound. Let $\mathcal{H}$ be a family of subsets of N elements, and L be a set of non-negative integers. We say that $\mathcal{H}$ is k-uniform if $|H| = k$ for each $H \in \mathcal{H}$. We have that $\mathcal{H}$ is L-intersecting if $|H \cap G| \in L$ for each pair of distinct $G, H \in \mathcal{H}$. The following lemma was proven by Ray-Chaudhuri and Wilson and provides an upper bound on the size of $\mathcal{H}$ knowing the cardinality of L.

9.5.1. Lemma (Ray-Chaudhuri–Wilson [RCW75]). *Let $\mathcal{H}$ be a L-intersecting k-uniform family of subsets of a set of N elements, where $s = |L| \leq k$. Then*

$$|\mathcal{H}| \leq \binom{N}{s}.$$

In terms of N and s, the lemma is easily shown to be tight by considering the family of all s-subsets of any set of size N.

Using Lemma 9.5.1, we can lower bound the size of our family $\mathcal{F}$. At first glance, the bound of Lemma 9.5.1 provides the wrong inequality: we are interested in a *lower* bound on the size of $\mathcal{F}$, whilst Lemma 9.5.1 provides an *upper* bound. The key idea is that we can consider a family $\mathcal{H}$ for all subsets that do *not* meet the condition of Eq. (9.17), and upper bound the size of that family to lower bound the size of the remaining family of subsets. After all, for a fixed k, the union of L_1-intersecting and L_2-intersecting k-uniform families of subsets of N elements results in the family of all k-uniform subsets if the union of L_1 and L_2 forms the set of all possible intersections.

9.5.2. Lemma. *For $\epsilon = \omega(1/N) > 0$, let $\mathcal{F}$ be an L-intersecting set of $N/2$-subsets of a set of N elements, where*

$$L = \{\lceil (1-\epsilon)N/4 \rceil, \lceil (1-\epsilon)N/4 \rceil + 1, \ldots, \lfloor (1+\epsilon)N/4 \rfloor\}.$$

Then we have that

$$|\mathcal{F}| \geq \Omega(2^{N/2}).$$

Proof:
Let $\mathcal{H}$ be an L'-intersecting $N/2$-uniform family of subsets of a set of N elements with $L' = \{0, 1, \ldots, N/2\} \setminus L$. We have that the cardinality of L', i.e., $s = |L'|$, can be upper bounded as

$$\begin{aligned} s &\leq 1 + N/2 - ((1+\epsilon)N/4 - (1-\epsilon)N/4 - 2) \\ &= N(1-\epsilon)/2 + 2 \\ &= 3 + \frac{N}{2}(1-\epsilon). \end{aligned}$$

For $\epsilon \geq \frac{6}{N}$ we have $s \leq N/2$ and, using Lemma 9.5.1, we get

$$|\mathcal{H}| \leq \binom{N}{s} \leq \binom{N}{3 + \frac{N}{2}(1-\epsilon)}.$$

This means that

$$\begin{aligned} |\mathcal{F}| &= \binom{N}{N/2} - |\mathcal{H}| \\ &\geq \binom{N}{N/2} - \binom{N}{3 + \frac{N}{2}(1-\epsilon)} \\ &= \Omega(2^{N/2}), \end{aligned}$$

if $N(1-\epsilon)/2 + 3 < N/2$, which holds asymptotically if $\epsilon = \omega\left(1/N\right)$. □

This implies that for $\epsilon = \omega\left(1/N\right) > 0$ there are a total of $\Omega(2^{N/2})$ functions $f = f_{\text{bal}}$ such that the corresponding balanced phase states from the set $\Psi^{\epsilon}_{\text{bal}} = \{|y_{f_{\text{bal}}}\rangle\}$ all have at most ϵ pairwise overlap.

9.5.3. Theorem. *For any $\epsilon = \omega\left(1/N\right)$, there exists a subset $S \in \mathcal{S}_K$, $K = N/2$, such that*

$$\mathcal{D}_{\epsilon}(|S\rangle, |\bar{S}\rangle) = \tilde{\Omega}(N).$$

Proof:
By Lemma 9.5.2, for any $\epsilon = \omega\left(1/N\right)$, there are $\Omega(2^{N/2})$ possible balanced functions f_{bal} such that all states $|y_{f_{\text{bal}}}\rangle$ have pairwise overlap $\leq \epsilon$. Since any circuit can only ϵ-approximately swap $|y_{f_{\text{con}}}\rangle$ to a single state $|y_{f_{\text{bal}}}\rangle \in \Psi^{\epsilon}_{\text{bal}}$, we need to consider $\Omega(2^{N/2})$ different possible circuits. Again, let $\mathcal{G}$ be any gate set with $g = n^{\mathcal{O}(1)}$ possible choices of gates and qubit indices on which a single (multi-qubit) gate acts. Starting from $|y_{f_{\text{con}}}\rangle$, the total number of states that can be constructed by circuits using M gates from $\mathcal{G}$ is at most g^M. Thus, for some balanced function, the swapper circuit must have $M = \tilde{\Omega}(N)$ gates, which implies $\mathcal{S}_{\epsilon} \geq \Omega(N)$. By Lemma 9.3.5, the result immediately follows. □

From Proposition 9.3.6, the following corollary immediately follows.

9.5.4. Corollary. *For any $\epsilon = \omega\left(1/N\right)$, there exists a subset $S \in \mathcal{S}_K$, $K = N/2$, such that*

$$\mathcal{C}_{\epsilon}(|S\rangle) = \tilde{\Omega}(N).$$

As a final remark, Corollary 9.5.4 can also qualitatively be obtained as a corollary of Theorem 9.4.8, as any provided samples can be hardcoded into a circuit.

9.5.3 Provable, verifiable and NISQable advantage in sample complexity?

Equipped with the above results, we propose an experiment to demonstrate the advantage of quantum over classical resources in a *provable*, *verifiable* and *NISQable* manner.

Quantum sample advantage in Complement Sampling

For a total of r rounds, a player and referee play the following interactive game:

1. The referee picks a pseudorandom permutation P with inverse P^{-1}. We say that $S = \{y : y = P(x) \text{ and the first bit of } x \in \mathcal{X} \text{ is } 0\}$.
2. The (quantum) player asks for k (quantum) samples corresponding to the uniform distribution over elements in S. A quantum player can perform coherent operations on the quantum samples. The player sends back a candidate element $\hat{y}$ supposedly from $\bar{S}$.
3. The referee verifies that $\hat{y} \in \bar{S}$ by checking if $\hat{x} = P^{-1}(\hat{y})$ has the first bit equal to 1.

As we know from Section 9.4, the best strategy a classical player has is to output any sample different from the observed samples. For $k = 1$, this random guessing strategy has success probability of $\frac{2^{n-1}}{2^n-1}$, quickly approaching $1/2$ for large n. By repeating the experiment a total of r times, each time with a different pseudorandom permutation, the probability of success is approximately $1/2^r$. On the other hand, the quantum player can use our algorithm to obtain $\left|\bar{S}\right\rangle$ from $|S\rangle$. This strategy succeeds with probability 1 at each of the r rounds. In reality, the quantum player only needs to succeed with probability $\geq \frac{1}{2} + 1/\text{poly}(n)$ at each round in order to beat the classical player over $r = \text{poly}(n)$ rounds. Thus, the player can tolerate a small rate of noise in the quantum computation. For example, in the simplistic noise model where each gate fails with probability λ, the overall circuit fidelity is of the order $(1-\lambda)^m$ where m is the number of gates. The quantum circuit of Fig. 9.2 only needs $\mathcal{O}(n)$ gates and $\mathcal{O}(\log n)$ circuit depth; the bottleneck being the n-qubit Toffoli gate, which can be implemented in $\mathcal{O}(\log n)$ circuit depth with 1 additional ancilla using the construction from [NZS24].

The preparation of the state $|S\rangle$ by the referee is computationally more expensive due to the pseudorandom permutation. As a near-term toy example, one could use the S-AES protocol, the simpler version of the Advanced Encryption Standard (AES) family. S-AES uses a block size of 16 bits (as opposed to 128

bits for AES) and a key size of 16 bits as well (128, 192 or 256 for AES). Though in this regime, complement sampling is still classically tractable—as it only needs $2^{15} = 32768$ distinct samples to solve the problem exactly—it could still display features of quantum advantage. A proof of concept based on the construction in [WWL22] requires only 48 qubits, 168 Toffoli gates, 364 CNOT gates and 75 NOT gates. When noise affects the computation, the referee is allowed to use quantum error detection codes [Kni04] before sending the state $|S\rangle$ to the player. By discarding circuit runs where an error is detected, the referee is expected to increase the fidelity of the state preparation. This comes at the cost of introducing more qubits, gates, and runs, but near-term protocols have shown promising results with low overheads [SBA24].

We expect that in the first complement sampling experiments all computations from both referee and player will be performed on the same device. Smart choices of families of subsets, and the use of circuit optimisations and error detection, may lead to a NISQable advantage experiment.

Appendices

Appendix A

This dissertation's Complexity Zoo

A.1 Deterministic classes

A.1.1. Definition (P). A promise problem $A = (A_{\text{YES}}, A_{\text{NO}})$ is in P if and only if there exists a deterministic polynomial-time Turing machine M, which takes as input a string $x \in \{0,1\}^*$, such that:

- if $x \in A_{\text{YES}}$ then M accepts x.
- if $x \in A_{\text{NO}}$ then M rejects x.

A.1.2. Definition (NP). A promise problem $A = (A_{\text{YES}}, A_{\text{NO}})$ is in NP if and only if there exists a deterministic polynomial-time Turing machine M and a polynomial p, where M takes as input a string $x \in \{0,1\}^*$ and a $p(|x|)$-bit witness y, such that:

- if $x \in A_{\text{YES}}$ then there exists a $y \in \{0,1\}^{p(|x|)}$ such that M accepts (x, y).
- if $x \in A_{\text{NO}}$ then for every $y \in \{0,1\}^{p(|x|)}$ we have that M rejects (x, y).

A.1.3. Definition (NqP). The class NqP is defined analogously to NP, but where the Turing machine is allowed to run in quasi-polynomial time (i.e., $2^{\text{polylog}(n)}$).

A.1.4. Definition (PSPACE). A promise problem $A = (A_{\text{YES}}, A_{\text{NO}})$ is in PSPACE if and only if there exists a deterministic polynomial-space Turing machine M, which takes as input a string $x \in \{0,1\}^*$, such that:

- if $x \in A_{\text{YES}}$, then M accepts x.
- if $x \in A_{\text{NO}}$, then M rejects x.

A.2 Randomized classes

A.2.1. Definition (BPP). A promise problem $A = (A_{\text{YES}}, A_{\text{NO}})$ is in BPP if and only if there exists a probabilistic polynomial-time Turing machine M, which takes as input a string $x \in \{0,1\}^*$, such that:

- If $x \in A_{\text{YES}}$, then M accepts x with probability at least $2/3$.
- If $x \in A_{\text{NO}}$, then M accepts x with probability at most $1/3$.

A.2.2. Definition (PCP$[r,q]$). A promise problem $A = (A_{\text{YES}}, A_{\text{NO}})$ has a (r,q)-PCP verifier if there exists a polynomial-time probabilistic Turing machine M, which takes an input $x \in \{0,1\}^*$, has random access to a string $\pi \in \{0,1\}^*$ of length at most $q(|x|)2^{r(|x|)}$, uses at most $r(|x|)$ random coins and makes at most $q(|x|)$ non-adaptive queries to locations of π, such that:

- If $x \in A_{\text{YES}}$, then there is a proof π such that $M^\pi(x)$ accepts with certainty.
- If $x \in A_{\text{NO}}$, then for all proofs π we have that $M^\pi(x)$ accepts with probability at most $1/2$.

A promise problem $A = (A_{\text{YES}}, A_{\text{NO}})$ belongs to PCP$[r,q]$ if and only if it has a (r,q)-PCP verifier.

A.2.3. Definition (MA). A promise problem $A = (A_{\text{YES}}, A_{\text{NO}})$ is in MA if and only if there exists a probabilistic polynomial-time Turing machine M and a polynomial p, such that M takes as input a string $x \in \{0,1\}^*$ and a $p(|x|)$-bit witness y and satisfies:

- if $x \in A_{\text{YES}}$, there exists a $y \in \{0,1\}^{p(|x|)}$ such that M accepts (x,y) with probability at least $2/3$.
- if $x \in A_{\text{NO}}$, then for every $y \in \{0,1\}^{p(|x|)}$, M accepts (x,y) with probability at most $1/3$.

A.2.4. Definition (IP$[k]$ and IP). A promise problem $A = (A_{\text{YES}}, A_{\text{NO}})$ is in IP$[k]$ if and only if there exists a polynomial p and interactive proof system, involving a probabilistic polynomial-time Turing machine M that is allowed to exchange k messages of length at most $p(|x|)$ with a computationally unbounded prover $P : \{0,1\}^* \to \{0,1\}^*$, such that on input $x \in \{0,1\}^*$ the following conditions hold:

- If $x \in A_{\text{YES}}$, there exists a prover P that causes M to accept with probability at least $2/3$.
- If $x \in A_{\text{NO}}$, then for every prover P, M accepts with probability at most $1/3$.

We define $\mathsf{IP} = \bigcup_{\alpha \geq 1} \mathsf{IP}[n^\alpha]$.

A.2.5. Definition ($\mathsf{AM}[k]$ and AM). For every k, the complexity class $\mathsf{AM}[k]$ is defined as the subset of $\mathsf{IP}[k]$ (Definition A.2.4), where the verifier's message to the prover consists only of its random coin tosses, and it does not use any further randomness during the protocol.

A.2.6. Definition (PP). A promise problem $A = (A_{\text{YES}}, A_{\text{NO}})$ is in PP if and only if there exists a probabilistic polynomial-time Turing machine M, which takes as input a string $x \in \{0,1\}^*$, such that:

- If $x \in A_{\text{YES}}$, then M accepts x with probability greater than $1/2$.
- If $x \in A_{\text{NO}}$, then M accepts x with probability at most $1/2$.

A.2.7. Remark. PP can also be defined as a non-deterministic complexity class where if $x \in A_{\text{YES}}$, then at least half of the witnesses must be accepted and if $x \in A_{\text{NO}}$, then strictly fewer than half of the witnesses are accepted.

A.3 Quantum classes

We adopt the convention that, given a circuit U_n and an input $x \in \{0,1\}^n$, the statement "U_n accepts x" means that the basis state $|x\rangle$ forms part of a larger initial state to which U_n is applied, and that a single designated output qubit is measured in the computational basis. The circuit U_n is said to accept x if and only if the measurement outcome is $|1\rangle$. The same convention applies when other strings y are part of the input to U_n. All circuits have access to ancilla registers initialised to $|0^m\rangle$ for some $m = \text{poly}(n)$, and the output qubit may depend arbitrarily on both the input and ancillas.

A.3.1. Definition (BQP). A promise problem $A = (A_{\text{YES}}, A_{\text{NO}})$ is in BQP if and only if there exists a polynomial q and P-uniform family of quantum circuits $\mathcal{U} = \{U_n : n \in \mathbb{N}\}$, where each circuit U_n takes as input a string $x \in \{0,1\}^n$ and uses $q(n)$ ancilla qubits initialised in $|0^{q(n)}\rangle$, such that:

- If $x \in A_{\text{YES}}$, then U_n accepts x with probability at least $2/3$.
- If $x \in A_{\text{NO}}$, then U_n accepts x with probability at most $1/3$.

A.3.2. Definition (QMA). A promise problem $A = (A_{\text{YES}}, A_{\text{NO}})$ is in QMA if and only if there exist polynomials p, q and P-uniform family of quantum circuits $\mathcal{U} = \{U_n : n \in \mathbb{N}\}$, where each circuit U_n takes as input a string $x \in \{0,1\}^n$, a $p(n)$-qubit quantum witness state $|\psi\rangle$, and $q(n)$ ancilla qubits initialised in $|0^{q(n)}\rangle$, such that:

- If $x \in A_{\text{YES}}$, then there exists a quantum witness $|\psi\rangle$ such that U_n accepts $(x, |\psi\rangle)$ with probability at least 2/3.
- If $x \in A_{\text{NO}}$, then for all quantum witnesses $|\psi\rangle$, U_n accepts $(x, |\psi\rangle)$ with probability at most 1/3.

A.3.3. Definition ($\mathsf{QMA}(k)$). A promise problem $A = (A_{\text{YES}}, A_{\text{NO}})$ is in $\mathsf{QMA}(k)$ if and only if there exist polynomials p, q, a P-uniform family of quantum circuits $\mathcal{U} = \{U_n : n \in \mathbb{N}\}$ such that each circuit U_n takes as input a string $x \in \{0,1\}^n$, k quantum witnesses $\{|\psi_i\rangle : i \in [k]\}$, each of $p(n)$ qubits, and $q(n)$ ancilla qubits initialised in $|0^{q(n)}\rangle$, such that:

- If $x \in A_{\text{YES}}$, then there exist quantum witnesses $|\psi_1\rangle, \ldots, |\psi_k\rangle$ such that U_n accepts $(x, |\psi_1\rangle, \ldots, |\psi_k\rangle)$ with probability at least 2/3.
- If $x \in A_{\text{NO}}$, then for all quantum witnesses $|\psi_1\rangle, \ldots, |\psi_k\rangle$, U_n accepts $(x, |\psi_1\rangle, \ldots, |\psi_k\rangle)$ with probability at most 1/3.

A.3.4. Definition (QCMA). A promise problem $A = (A_{\text{YES}}, A_{\text{NO}})$ is in QCMA if and only if there exist polynomials p, q and a P-uniform family of quantum circuits $\mathcal{U} = \{U_n : n \in \mathbb{N}\}$, where each circuit U_n takes as input a string $x \in \{0,1\}^n$, a classical witness string y of length $p(n)$, and $q(n)$ ancilla qubits initialised in $|0^{q(n)}\rangle$, such that:

- if $x \in A_{\text{YES}}$, then there exists a classical witness $y \in \{0,1\}^{p(n)}$ such that U_n accepts (x, y) with probability at least 2/3.
- if $x \in A_{\text{NO}}$, then for all classical witnesses $y \in \{0,1\}^{p(n)}$, U_n accepts (x, y) with probability at most 1/3.

A.3.5. Definition ($\mathsf{PostBQP}$). Let $\mathcal{G}$ be the universal gate set consisting of the generators of the Clifford group and the T-gate. A promise problem $A = (A_{\text{YES}}, A_{\text{NO}})$ is in $\mathsf{PostBQP}$ if and only if there exists a polynomial q, and a P-uniform family of polynomial-size quantum circuits $\mathcal{U} = \{U_n : n \in \mathbb{N}\}$, where each U_n uses only gates from $\mathcal{G}$ and takes as input a string $x \in \{0,1\}^n$, such that the following conditions hold:

- After U_n is applied to the initial state $|0^{q(n)}\rangle \otimes |x\rangle$, the first qubit has a nonzero probability of being measured as $|1\rangle$.
- If $x \in A_{\text{YES}}$, then conditioned on the first qubit being measured as $|1\rangle$, the second qubit is $|1\rangle$ with probability at least 2/3.
- If $x \in A_{\text{NO}}$, then conditioned on the first qubit being measured as $|1\rangle$, the second qubit is $|1\rangle$ with probability at most 1/3.

A.3.6. Definition (SBQP). A promise problem $A = (A_{\text{YES}}, A_{\text{NO}})$ is in SBQP if and only if there exist polynomials p, q and a P-uniform family of quantum circuits $\mathcal{U} = \{U_n : n \in \mathbb{N}\}$, where each circuit U_n takes as input a string $x \in \{0,1\}^n$, $q(n)$ ancilla qubits initialised in $|0^{q(n)}\rangle$, such that:

- If $x \in A_{\text{YES}}$, then U_n accepts x with probability at least $2^{-p(n)}$.
- If $x \in A_{\text{NO}}$, then U_n accepts x with probability at most $2^{-p(n)-1}$.

Appendix B

Omitted classification proofs

As all the examples of families of states in Section 3.4 have a classical description in standard English text from which it can be verified that they indeed belong to the correct class, we can simply take the function rep$(\cdot)$ to be any standard text-to-binary converter, e.g., ASCII.

B.1 Clustered product states

We only consider the case where $k = \mathcal{O}(\log n)$. Being classically evaluatable directly follows, as commuting expectation values of $\mathcal{O}(\log n)$-local observables only involves matrix-vector products of dimensions poly(n).

We will reduce the question of all the remaining properties for k-clustered product states to the MPS setting. For $k = \max_{i\in[m]} |V_i| = \mathcal{O}(\log n)$, let $\mathcal{H}_A$ be the Hilbert space such that $|u\rangle \in \mathcal{H}_A$. We define an extended k-clustered product state $|u_{\text{ext}}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ as

$$|u_{\text{ext}}\rangle = \bigotimes_{i\in[m]} \left(|u_i\rangle_{V_i} \otimes |0\rangle^{\otimes q_i} \right),$$

where $q_i = 2^{k-|V_i|}$, such that $\text{tr}_B[|u_{\text{ext}}\rangle\langle u_{\text{ext}}|] = |u\rangle\langle u|$. Since $k = \mathcal{O}(\log n)$, we have that $|u_{\text{ext}}\rangle$ can be viewed as an MPS with local dimension $d = 2^k = \text{poly}(n)$. Being perfectly samplable and uniformly quantumly preparable then follows directly from the MPS case.

B.2 Bounded MPS

Perfectly samplable. Let u be an $N = 2^n$-dimensional vector described by an MPS of n particles, with bounded bond dimension D and local particle dimension d. We will see that matrix product states are even samplable for any $1 \leq d \leq$

poly(n), which is crucial to make our reduction from clustered product states in Appendix B.1.

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

(ii) Expectation values of tensor products of 1-local observables can be computed efficiently for an MPS in time $\mathcal{O}(nd^2D^3)$ [VMC08]. Assuming m is already known (see item (iii)) and that the MPS is normalized, the algorithm Q_u^{samp} works as follows: for each $j \in \{0, 1, \dots, d-1\}$, compute the probability p_j that the first qudit is in state j by evaluating the expectation value of the 1-local projector $\Pi_j = |j\rangle\langle j|_1$. The algorithm then samples a number $b_1 \in \{0, 1, \dots, d-1\}$ according to the distribution $\{p_j\}$, and computes the expectation values of the 2-local projectors $\Pi_{b_1 j} = |b_1\rangle\langle b_1|_1 \otimes |j\rangle\langle j|_2$ to obtain the distribution $\{p_{b_1 j}\}$, from which the next bit is sampled. This process is repeated for all $n-2$ remaining sites, yielding a sample b with probability $|u_b|^2$. The total time complexity is $\mathcal{O}(n^2d^3D^3) = \mathcal{O}(\text{poly}(\log N))$, when $d = \mathcal{O}(\text{poly}(n))$ and $D = \mathcal{O}(\text{poly}(n))$, as desired.

(iii) The value of m can be computed by evaluating the overlap of the MPS with itself, which is equivalent to computing the expectation value of a 0-local observable. This can be done in time $\mathcal{O}(ndD^3)$.

Classically evaluatable. One can compute the inner product $\langle u|O|u\rangle$ in time at most $n(2D^3\chi d + D^2\chi^2 d^2)$ for any (even n-local) operator O that admits a matrix product operator (MPO) decomposition with bond dimension χ [Sch11, Orú14]. Since O is k-local, it can be represented by an MPO with bond dimension at most d^k, so $\langle u|O|u\rangle$ can be computed in time at most $n(2D^3d^{k+1} + D^2d^{2k+2}) = \text{poly}(n, D, 2^k)$ when d is constant.

Uniformly quantumly preparable. An MPS on n qubits with bond dimension D can be prepared efficiently when $d = \mathcal{O}(1)$ and $D = \text{poly}(n)$ using the techniques from [SSV+05].

B.3 Constant depth quantum circuits

Approximately samplable. 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 $\mathsf{BQP} \subseteq \mathsf{AM}$. This means that, under the same condition, constant-depth quantum circuits are not ξ-samplable for any $\xi < 1/3$.

Classically evaluatable. We have that $\langle u| O |u\rangle = \langle 0| U^\dagger OU |0\rangle$, where $U^\dagger OU$ is a $k2^t$-local observable (via a light-cone argument). Hence, we can compute $\langle 0| U^\dagger OU |0\rangle$ in time $\mathcal{O}\left(2^{\mathcal{O}(k2^t)} \cdot \text{poly}(n)\right)$, which is $\text{poly}(2^k)$ if $t = \mathcal{O}(1)$.

Uniformly quantumly preparable. This holds by definition.

B.4 2D isoTNS

Samplable. This follows from the next argument for approximately classically evaluatable states and Proposition 3.5.10.

Approximately classically evaluatable. It is shown in [MT25] that computing expectation values of a 2D isoTNS up to constant additive error is BQP-hard, so this would not be possible under the assumption that $\mathsf{BPP} \neq \mathsf{BQP}$.

Uniformly quantumly preparable. This is shown in [SC21] and [MT25].

B.5 PEPS

It is known that computing the expectation values of local observables given the description of a PEPS is $\#\mathsf{P}$-hard, and that preparing PEPS as a quantum state is $\mathsf{PostBQP}$-hard [SWVC07], which means that under the assumption $\#\mathsf{P} \neq \mathsf{FBPP}$, PEPS are not samplable or approximately classically evaluatable. Additionally, under the assumption that $\mathsf{PP} \notin \mathsf{BQP}$ (noting that $\mathsf{PP} = \mathsf{PostBQP}$ [Aar05]), PEPS are not uniformly quantumly preparable.

Appendix C

Classical energy estimation with classically evaluatable states

In this appendix, we present an alternative algorithm to the one proposed in [GL22] for $\mathrm{GLH}(k, a, b, \zeta)$ and $\mathrm{GaLH}(k, a, b, \zeta)$, in the setting where the guiding state is classically evaluatable rather than samplable. Besides the assumed access model, our algorithm also differs from the one proposed in [GL22] in two additional ways:

- We focus on local Hamiltonians rather than arbitrary sparse complex matrices. This simplifies the algorithm, as functions applied to such Hamiltonians can be interpreted directly in terms of their *spectrum*, rather than their *singular values*.

- We further simplify the algorithm by tailoring it to ground state *decision* problems, as opposed to *estimation* problems.

We introduce and analyse the complexity of the *spectral amplification* algorithm in Appendix C.1, and discuss how, together with the results in Chapter 4, it has certain implications for the quantum PCP conjecture (see Chapter 6).

C.1 Low-energy projectors via spectral amplification

Let $H = \sum_{i\in[m]} H_i$ be a Hamiltonian on n qubits, where each H_i is an at most k-local Hermitian operator, and assume $\|H\| \leq 1$. Since H is Hermitian, it admits a spectral decomposition:

$$H = \sum_i \lambda_i \left|\psi_i\right\rangle\!\left\langle\psi_i\right|,$$

where $\lambda_i \in [-1, 1]$ are the eigenvalues, and $|\psi_i\rangle$ the corresponding eigenvectors. Let $P \in \mathbb{R}[x]$ be a polynomial of degree d, written as

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

We define the *polynomial spectral amplification* of H with respect to P as

$$\begin{aligned} P(H) &= a_0 \mathbb{I} + a_1 H + \cdots + a_d H^d \\ &= \sum_i P(\lambda_i) \, |\psi_i\rangle\langle\psi_i| \, , \end{aligned}$$

where we used that $H^k = \sum_i \lambda_i^k \, |\psi_i\rangle\langle\psi_i|$ for all $k \geq 0$. For any $\alpha \in [-1, 1]$, we define the *low-energy projector* of H as

$$\Pi_\alpha = \sum_{\{i : \lambda_i \leq \alpha\}} |\psi_i\rangle\langle\psi_i| \, . \tag{C.1}$$

Note that for any $\alpha \geq \lambda_0$, we have $\Pi_{\text{gs}} \Pi_\alpha = \Pi_\alpha \Pi_{\text{gs}} = \Pi_{\text{gs}}$, where Π_{gs} denotes the projection onto the ground space of H. The operator Π_α can in principle be used to solve CGaLH(k, a, b, ζ) by computing $\| \Pi_\alpha \, |u\rangle \|^2$ for $\alpha = a$, given a classically evaluatable state $|u\rangle$:

- In the YES-case, there exists a classically evaluatable state $|u\rangle$ such that $\| \Pi_a \, |u\rangle \|^2 \geq \| \Pi_{\text{gs}} \, |u\rangle \|^2 \geq \zeta$.
- In the NO-case, by the promise, all eigenvalues of H are greater than a, hence $\Pi_a = 0$, and so $\| \Pi_a \, |u\rangle \|^2 = 0$ for any state $|u\rangle$.

Thus, the two cases are separated by a gap of at least ζ.

However, there are two complications: (1) an efficient classical description of Π_a is unlikely to exist, and (2) even if it did, Π_a is generally not k-local, meaning the quantity $\| \Pi_a \, |u\rangle \|^2$ may not be efficiently computable.

The idea is now to approximate the low-energy projector Π_α by a polynomial in H. Observe that Π_α can be written exactly as

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

where $\operatorname{sgn}(x)$ is the sign function, defined for $x \in \mathbb{R}$ as

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

Thus, Π_α acts as a step function that projects onto the eigenspaces with eigenvalues at most α. The sign function is discontinuous, so it cannot be represented exactly by a polynomial. However, we can approximate it pointwise using techniques from [HC17], where polynomial approximations to $\operatorname{sgn}(x)$ on a bounded domain are constructed with a small error outside a small transition region, as indicated by the following lemma:

C.1.1. Lemma ([HC17]). *For all $\delta' > 0$ and $\epsilon' \in (0, 1/2)$, 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 $x \in [-2, 2] \setminus (-\delta', \delta')$, $|P(x) - \mathrm{sgn}(x)| \leq \epsilon'$.*

It is known that Lemma C.1.1 is optimal in its parameters δ' and ϵ' for the required conditions [HC17].

Since Lemma C.1.1 holds over the entire interval $[-2, 2]$, choosing any $\alpha \in [-1, 1]$ and scaling the $\mathrm{sgn}(x)$ function by a factor of $1/2$ ensures that the approximation error is at most $\epsilon/2$. Define $q_\alpha(x) : \mathbb{R} \to [0, 1]$ as $q_\alpha(x) = \frac{1}{2}(1 - \mathrm{sgn}(x - \alpha))$, and let $Q_\alpha \in \mathbb{R}[x]$ be a degree-d polynomial approximating it. Note that Q_α can be written as a shifted version of the polynomial P from Lemma C.1.1, i.e., $Q_\alpha(x) = \frac{1}{2}(1 - P(x - \alpha))$. We define $\tilde{\Pi}_\alpha = Q_\alpha(H)$ as the polynomial approximation to the low-energy projector Π_α. Although $\tilde{\Pi}_\alpha$ is Hermitian (since it is a polynomial in the Hermitian operator H), it is not a true projector; generally, we would have $\tilde{\Pi}_\alpha^2 \neq \tilde{\Pi}_\alpha$. Replacing Π_α by $\tilde{\Pi}_\alpha$ in the expression $\|\Pi_\alpha |u\rangle\|^2$, we obtain

$$\left\| \tilde{\Pi}_\alpha |u\rangle \right\|^2 = \sqrt{\langle u| \tilde{\Pi}_\alpha^\dagger \tilde{\Pi}_\alpha |u\rangle} = \langle u| \tilde{\Pi}_\alpha^2 |u\rangle = \langle u| \left(Q_\alpha(H)\right)^2 |u\rangle .$$

This implies that we must compute expectation values of H up to degree $2d$. The next lemma will provide an upper bound on the number of such evaluations needed to compute this expression.

C.1.2. Lemma. *Let $u \in \mathbb{C}^{2^n}$, and let $H = \sum_{i \in [m]} H_i$ be an n-qubit Hamiltonian, where each H_i is k-local and $m \geq 2$. Let $P(x)$ be a polynomial of degree d. Then, the expectation value $\langle u| P(H) |u\rangle$ can be written as a sum of $\mathcal{O}(m^d)$ terms $\langle u| O_j |u\rangle$, where each observable O_j is at most kd-local.*

Proof:
We begin by expanding the polynomial:

$$\langle u| P(H) |u\rangle = \sum_{l=0}^{d} a_l \langle u| H^l |u\rangle .$$

Each power H^l is given by

$$H^l = \left(\sum_{i \in [m]} H_i \right)^l = \sum_{(i_1, \ldots, i_l) \in [m]^l} H_{i_1} \cdots H_{i_l},$$

which expands into at most m^l terms. Each such term is a product of l k-local operators, and so acts non-trivially on at most kl qubits. To ensure Hermiticity of the expectation values, observe that for every non-Hermitian monomial $Q = H_{i_1} \cdots H_{i_l}$, its Hermitian conjugate $Q^\dagger = H_{i_l} \cdots H_{i_1}$ also appears in the expansion. Hence, we can group each such pair and consider the Hermitian observable:

$$\hat{Q} = \frac{1}{2}(Q + Q^\dagger), \tag{C.2}$$

which remains kl-local and has operator norm at most 1. The terms where all H_{i_j} commute trivially are already Hermitian. Therefore, each $\langle u| H^l |u\rangle$ can be written as a weighted sum of at most m^l expectation values of Hermitian kl-local observables O_j, where we multiply the computed expectation value by a factor of 2 when it is of the form of Eq. (C.2). The total number of such expectation values across all powers l is bounded by

$$\sum_{l=1}^{d} m^l = \frac{m(m^d - 1)}{m - 1} = \frac{m^d - 1}{1 - 1/m} = \mathcal{O}\left(m^d\right),$$

assuming $m \geq 2$. □

All that remains to show is that, for a constant promise gap δ, a sufficiently good approximation $\tilde{\Pi}_\alpha$ with an appropriate choice of α ensures that we can still distinguish between the two cases in the $\mathsf{CGaLH}(H, a, \zeta)$ problem in polynomial (resp. quasi-polynomial) time in m, when $\zeta = \Omega(1)$ (resp. $\zeta = 1/\mathrm{poly}(n)$).

C.1.3. Theorem. *Let $H = \sum_{i\in[m]} H_i$ be a Hamiltonian on n qubits, $m \geq 2$, and let* $\mathrm{desc}(u)$ *be a description of a classically evaluatable state $|u\rangle \in \mathbb{C}^{2^n}$. Let $a, b \in [-1, 1]$ such that $b - a \geq \delta$ for some $\delta > 0$, and let $\zeta \in (0, 1]$. Consider the following two cases, with the promise that one holds:*

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

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

Then there exists a classical algorithm that distinguishes between cases (i) *and* (ii) *using*

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

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

Proof:
Set $\alpha := \frac{a+b}{2}$, $\delta' := \delta/2$, $\epsilon' := \zeta/10$, and $d := \mathcal{O}(\log(1/\epsilon')/\delta')$. Let $\tilde{\Pi}_\alpha := Q_\alpha(H)$, where Q_α is the polynomial of degree d approximating the low-energy projector $\Pi_\alpha = \frac{1}{2}(1 - \mathrm{sgn}(H - \alpha\mathbb{I}))$, where $\mathrm{sgn}(\cdot)$ is approximated as per Lemma C.1.1. We want to evaluate $\left\|\tilde{\Pi}_\alpha |u\rangle\right\|^2$, which requires computing expectation values of $(Q_\alpha(H))^2$, which is a degree-$2d$ polynomial in H. We proceed as follows:

1. Compute $\|Q_\alpha(H)\,|u\rangle\|^2$.
2. If $\|Q_\alpha(H)\,|u\rangle\|^2 \geq \frac{9}{10}\zeta$, output (i); otherwise, output (ii).

By Lemma C.1.2, this requires at most $\mathcal{O}\left(m^{c(\log(1/\zeta)/\delta)}\right)$ computations of expectation values of local observables, for some universal constant $c > 0$. Since $|u\rangle$ is classically evaluatable and the observables are at most kd-local Hermitian operators with efficiently computable matrix elements, each expectation value $\langle u|\,O_j\,|u\rangle$ can be computed efficiently.

We now prove correctness. Since $Q_\alpha(H)$ is a real polynomial in the Hermitian operator H, it is Hermitian and has the same eigenvectors as H, with eigenvalues given by $Q_\alpha(\lambda_i)$ for each eigenvalue λ_i of H. Hence, we can write

$$Q_\alpha(H) = \sum_i Q_\alpha(\lambda_i)\,|\psi_i\rangle\langle\psi_i|\,,$$

which means that

$$\|Q_\alpha(H)\,|u\rangle\|^2 = \langle u|\,(Q_\alpha(H))^2\,|u\rangle = \sum_i Q_\alpha(\lambda_i)^2|\,\langle\psi_i|u\rangle\,|^2.$$

Case (i): Since $\|\Pi_{\text{gs}}\,|u\rangle\|^2 \geq \zeta$, and for each i such that $\lambda_i \leq a$, we have that Lemma C.1.1 guarantees $Q_\alpha(\lambda_i)^2 \geq (1-\zeta/20)^2$. Hence,

$$\begin{aligned}\|Q_\alpha(H)\,|u\rangle\|^2 &= \sum_i Q_\alpha(\lambda_i)^2|\,\langle\psi_i|u\rangle\,|^2 \\ &\geq \left(1-\frac{\zeta}{20}\right)^2 \sum_{i:\lambda_i\leq a} |\,\langle\psi_i|u\rangle\,|^2 \\ &\geq \left(1-\frac{\zeta}{20}\right)^2 \zeta,\end{aligned}$$

Since $\zeta \in (0,1]$, this can lower bounded further as

$$\left(1-\frac{\zeta}{20}\right)^2\zeta = \left(1-\frac{\zeta}{10}+\frac{\zeta^2}{400}\right)\zeta \geq \left(1-\frac{\zeta}{10}\right)\zeta \geq \frac{9}{10}\zeta.$$

Case (ii): All eigenvalues satisfy $\lambda_i \geq b$, so $Q_\alpha(\lambda_i)^2 \leq (\zeta/20)^2$. Hence,

$$\|Q_\alpha(H)\,|u\rangle\|^2 \leq \frac{\zeta^2}{400}\sum_i |\,\langle\psi_i|u\rangle\,|^2 = \frac{\zeta^2}{400}.$$

Therefore, the gap between the two cases is at least

$$\frac{9}{10}\zeta - \frac{\zeta^2}{400} \geq \frac{9}{10}\zeta - \frac{1}{400}\zeta \geq \frac{4}{5}\zeta,$$

which is inverse polynomial in n whenever $\zeta \geq 1/\mathrm{poly}(n)$, concluding the proof. □

The following corollary follows straightforwardly from Theorem C.1.3:

C.1.4. Corollary. *For any $k = \mathcal{O}(\log n)$, any constant $\delta \in (0, 1]$ such that $b - a = \delta$, and any constant $\zeta \in (0, 1]$, we have that* $\mathrm{CGaLH}(k, a, b, \zeta)$ *is in* NP. *If instead $\zeta \geq 1/\mathrm{poly}(n)$ holds, then* $\mathrm{CGaLH}(k, a, b, \zeta)$ *is in* NqP.

C.2 Implications to the quantum PCP conjecture

For readers unfamiliar with the quantum PCP conjecture, we recommend first reviewing Section 1.2.1 or Section 6.1.

C.2.1 Gap amplification

We consider the implications of the above sections on the possibility of gap amplification for local Hamiltonian problems. The following corollary follows from Theorem C.1.3 and Theorem 4.4.4.

C.2.1. Corollary (No-go results for gap amplification). *There cannot exist:*

1. *A polynomial-time reduction which maps any instance of* $\mathrm{CGaLH}(k, a, b, \zeta)$ *with $k \geq 2$, some constant $\zeta > 0$, and $b - a = 1/\mathrm{poly}(n)$ to an instance of* $\mathsf{CGaLH}(k', a', b', \zeta')$ *with $k' \geq 2$, some constant $\zeta' > 0$, and $b' - a' = \Omega(1)$,*

unless QCMA = NP, *and*

2. *A quasi-polynomial-time reduction which maps any instance of* $\mathsf{CGaLH}(k, a, b, \zeta)$ *with $k \geq 2$, $\zeta = 1/\mathrm{poly}(n)$, and $b - a = 1/\mathrm{poly}(n)$ to an instance of* $\mathsf{CGaLH}(k', a', b', \zeta')$ *with $k' \geq 2$, $\zeta' = 1/\mathrm{poly}(n)$, and $b' - a' = \Omega(1)$,*

unless QCMA ⊆ NqP.

One can also interpret the no-go results for gap amplification (Points 1 and 2 in the above corollary) in a more general setting: if one seeks to prove the QPCP conjecture through a gap amplification procedure a la Dinur's [Din07], the procedure must necessarily fail to preserve classically evaluatability—it cannot even maintain inverse-polynomial fidelity with such states—unless it simultaneously shows that QCMA = NP (or QCMA ⊆ NqP, which is also very unlikely). Hence, this result can be viewed as a QCMA-analogue of the result from Aharonov and Grilo [ABG19], who showed that the existence of quantum gap amplification procedures that preserve stoquasticity would imply NP = MA. We believe this

constitutes a significant obstacle to constructing a quantum gap amplification procedure, as many of the existing Hamiltonian gadget constructions do indeed preserve classical evaluatability—this was precisely what we exploited in Section 4.6.3 to extend the hardness results to a large number of different families of Hamiltonians.

C.2.2 No low-energy classically evaluatable states

We close by formulating a new conjecture, which we call the *no low-energy classically-evaluatable states (NLCES)* conjecture. This can be viewed as a strengthening of the NLTS theorem, or as an alternative to the NLSS conjecture of [GL22] in light of our results. It must hold if the quantum PCP conjecture is true and $\mathsf{QMA} \neq \mathsf{NP}$.

C.2.2. Conjecture (NLCES). *There exists a family of local Hamiltonians* $\{H_n : n \in \mathbb{Z}_+\}$*, where each* H_n *acts on* n *qubits, and a constant* $\beta > 0$*, such that for sufficiently large integer* n*, for all classically evaluatable states* $|u\rangle \in \mathbb{C}^{2^n}$ *as per Definition 3.5.5, it holds that*

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

Taking into account our results on the containment of the constant-gapped classically guidable local Hamiltonian problem in NP—in particular the insight that what really matters is the *fidelity* of a classically evaluatable state with the low-energy subspace of the Hamiltonian, rather than its energy expectation—we propose a stronger variant of the NLCES conjecture, which must also hold if the quantum PCP conjecture is true.

C.2.3. Conjecture (Strong-NLCES). *There exists a family of local Hamiltonians* $\{H_n : n \in \mathbb{Z}_+\}$*, where each* H_n *acts on* n *qubits, and a constant* $\beta > 0$*, such that for sufficiently large integer* n*, for all classically evaluatable states* $|u\rangle \in \mathbb{C}^{2^n}$ *as per Definition 3.5.5, we have*

$$\|\Pi_{\lambda_0(H_n)+\beta} |u\rangle\|^2 = o(1/\mathrm{poly}(n)),$$

where $\Pi_{\lambda_0(H_n)+\beta}$ *denotes the projector onto the subspace spanned by eigenvectors of* H_n *with energy less than* $\lambda_0(H_n) + \beta$*.*

Note that the NLCES conjecture is strictly weaker than the Strong-NLCES conjecture, and that neither necessarily implies the QPCP conjecture.

Bibliography

[AA15] Scott Aaronson and Andris Ambainis. Forrelation: A Problem that Optimally Separates Quantum from Classical Computing. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '15, page 307–316, 2015. arXiv: 1411.5729. 4, 14, 18, 167

[AA24] Anurag Anshu and Srinivasan Arunachalam. A survey on the complexity of learning quantum states. *Nature Reviews Physics*, 6(1):59–69, 2024. arXiv: 2305.20069. 16

[AAB+19] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando G.S.L. Brandao, David A. Buell, et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 2019. arXiv: 1910.11333. 229

[AALV09] Dorit Aharonov, Itai Arad, Zeph Landau, and Umesh Vazirani. The detectability lemma and quantum gap amplification. In *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*, STOC '09, page 417–426, 2009. arXiv: 0811.3412. 12, 132, 133, 158

[Aar05] Scott Aaronson. Quantum Computing, Postselection, and Probabilistic Polynomial-Time. *Proceedings: Mathematical, Physical and Engineering Sciences*, 461(2063):3473–3482, 2005. arXiv: quant-ph/0412187. 271

[Aar06] Scott Aaronson. QMA/qpoly $\subseteq$ PSPACE/poly: de-Merlinizing quantum protocols. In *21st Annual IEEE Conference on Computational Complexity (CCC'06)*, pages 13 pp.–273, July 2006. arXiv: quant-ph/0510230. 227

[Aar07] Scott Aaronson. The learnability of quantum states. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 463(2088):3089–3114, 2007. arXiv: `quant-ph/0608142`. 232

[Aar09] Scott Aaronson. Computational complexity: Why quantum chemistry is hard. *Nature Physics*, 5:707–708, October 2009. 6, 75

[AAS20] Scott Aaronson, Yosi Atia, and Leonard Susskind. On the hardness of detecting macroscopic superpositions, 2020. arXiv: `2009.07450`. 232, 233, 234, 255

[AAV13] Dorit Aharonov, Itai Arad, and Thomas Vidick. Guest column: the quantum PCP conjecture. *SIGACT News*, 44(2):47–79, June 2013. arXiv: `1309.7495`. 133, 176

[AB09] Sanjeev Arora and Boaz Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, 2009. 19

[ABC+20] Srinivasan Arunachalam, Aleksandrs Belovs, Andrew M. Childs, Robin Kothari, Ansis Rosmanis, and Ronald de Wolf. Quantum Coupon Collector. In *15th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2020)*, volume 158, pages 10:1–10:17, 2020. arXiv: `2002.07688`. 232

[ABG19] Dorit Aharonov and Alex Bredariol Grilo. Stoquastic PCP vs. Randomness. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1000–1023, 2019. arXiv: `1901.05270`. 12, 132, 278

[ABN23] Anurag Anshu, Nikolas P. Breuckmann, and Chinmay Nirkhe. NLTS Hamiltonians from Good Quantum Codes. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, STOC 2023, page 1090–1096, 2023. arXiv: `2206.13228`. 12, 63, 131

[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`. 88, 90

[AdW17] Srinivasan Arunachalam and Ronald de Wolf. Guest Column: A Survey of Quantum Learning Theory. *SIGACT News*, 48(2):41–67, June 2017. arXiv: `1701.06806`. 16, 37, 231, 232

[AF11] P.W. Atkins and R.S. Friedman. *Molecular Quantum Mechanics*. OUP Oxford, 2011. 44

[AGDLHG05] Alán Aspuru-Guzik, Anthony D. Dutoi, Peter J. Love, and Martin Head-Gordon. Simulated Quantum Computation of Molecular Energies. *Science*, 309(5741):1704–1707, 2005. arXiv: `quant-ph/0604193`. 7, 75

[AGK24] Eric R. Anschuetz, David Gamarnik, and Bobak Kiani. Combinatorial NLTS From the Overlap Gap Property. *Quantum*, 8:1527, November 2024. arXiv: `2304.00643`. 12, 131

[Aha03] Dorit Aharonov. A simple proof that Toffoli and Hadamard are quantum universal, 2003. arXiv: `quant-ph/0301040v`. 32

[AIK22] Scott Aaronson, DeVon Ingram, and William Kretschmer. The Acrobatics of BQP. In *Proceedings of the 37th Computational Complexity Conference (CCC 2022)*, volume 234 of *LIPIcs*, pages 20:1–20:17, 2022. arXiv: `2111.10409`. 191, 192, 193

[AJL06] Dorit Aharonov, Vaughan Jones, and Zeph Landau. A polynomial quantum algorithm for approximating the Jones polynomial. In *Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '06, page 427–436, 2006. arXiv: `quant-ph/0511096`. 142

[AK07] Scott Aaronson and Greg Kuperberg. Quantum versus classical proofs and advice. In *Twenty-Second Annual IEEE Conference on Computational Complexity (CCC'07)*, pages 115–128, 2007. arXiv: `quant-ph/0604056`. 13, 15, 201, 228

[AL99] Daniel S. Abrams and Seth Lloyd. Quantum Algorithm Providing Exponential Speed Increase for Finding Eigenvalues and Eigenvectors. *Physical Review Letters*, 83:5162–5165, 1999. 7, 75

[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. 11, 131

[ALVV17] Itai Arad, Zeph Landau, Umesh Vazirani, and Thomas Vidick. Rigorous RG algorithms and area laws for low energy eigenstates in 1D. *Communications in Mathematical Physics*, 356:65–105, 2017. arXiv: `1602.08828`. 223

[Amb02] Andris Ambainis. Quantum Lower Bounds by Quantum Arguments. *Journal of Computer and System Sciences*, 64(4):750–767, 2002. arXiv: `quant-ph/0002066`, Earlier version in STOC'00. 196

[Amb14] Andris Ambainis. On Physical Problems that are Slightly More Difficult than QMA. In *Proceedings of the 2014 IEEE 29th Conference on Computational Complexity*, CCC '14, page 32–43, 2014. arXiv: `1312.4758`. 109, 117

[AN02] Dorit Aharonov and Tomer Naveh. Quantum NP—a survey, October 2002. arXiv: `quant-ph/0210077`. 12, 14, 132, 136

[AS98] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: a new characterization of NP. *J. ACM*, 45(1):70–122, January 1998. 11, 131

[AZ24] Scott Aaronson and Yuxuan Zhang. On verifiable quantum advantage with peaked circuit sampling, April 2024. arXiv: `2404.14493`. 229

[BACS07] Dominic W. Berry, Graeme Ahokas, Richard Cleve, and Barry C. Sanders. Efficient quantum algorithms for simulating sparse Hamiltonians. *Communications in Mathematical Physics*, 270:359–371, 2007. arXiv: `quant-ph/0508139`. 221

[BBBV97] Charles H. Bennett, Ethan Bernstein, Gilles Brassard, and Umesh Vazirani. Strengths and Weaknesses of Quantum Computing. *SIAM Journal on Computing*, 26(5):1510–1523, 1997. arXiv: `quant-ph/9701001`. 15, 201

[BBC+95] Adriano Barenco, Charles H. Bennett, Richard Cleve, David P. DiVincenzo, Norman Margolus, Peter Shor, Tycho Sleator, John A. Smolin, and Harald Weinfurter. Elementary gates for quantum computation. *Phys. Rev. A*, 52:3457–3467, Nov 1995. arXiv: `quant-ph/9503016`. 80

[BBC+01] Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald de Wolf. Quantum lower bounds by polynomials. *J. ACM*, 48(4):778–797, July 2001. arXiv: `quant-ph/9802049`. 181, 196

[BBMC20] Bela Bauer, Sergey Bravyi, Mario Motta, and Garnet Kin-Lic Chan. Quantum Algorithms for Quantum Chemistry and Quantum Materials Science. *Chemical Reviews*, 120(22):12685–12717, 2020. arXiv: `2001.03685`. 6, 75

[BBT09] Nikhil Bansal, Sergey Bravyi, and Barbara M. Terhal. Classical approximation schemes for the ground-state energy of quantum and classical ising spin Hamiltonians on planar graphs. *Quantum Information & Computation*, 9(7):701–720, July 2009. arXiv: 0705.1115. 12

[BBW25] Marcello Benedetti, Harry Buhrman, and Jordi Weggemans. Complement Sampling: Provable, Verifiable and NISQable Quantum Advantage in Sample Complexity, 2025. arXiv: 2502.08721. vi

[BCG21] Boaz Barak, Chi-Ning Chou, and Xun Gao. Spoofing Linear Cross-Entropy Benchmarking in Shallow Quantum Circuits. In *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*, volume 185 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 30:1–30:20, 2021. arXiv: 2005.02421. 229

[BCM+21] Zvika Brakerski, Paul Christiano, Urmila Mahadev, Umesh Vazirani, and Thomas Vidick. A Cryptographic Test of Quantumness and Certifiable Randomness from a Single Quantum Device. *J. ACM*, 68(5), August 2021. arXiv: 1804.00640. 230

[BCW98] Harry Buhrman, Richard Cleve, and Avi Wigderson. Quantum vs. classical communication and computation. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, STOC '98, page 63–68, 1998. arXiv: quant-ph/9802040. 4

[BCWdW01] Harry Buhrman, Richard Cleve, John Watrous, and Ronald de Wolf. Quantum Fingerprinting. *Phys. Rev. Lett.*, 87:167902, Sep 2001. arXiv: quant-ph/0102001. 4, 215, 230

[BCY11] Fernando GSL Brandao, Matthias Christandl, and Jon Yard. Faithful squashed entanglement. *Communications in Mathematical Physics*, 306:805–830, 2011. arXiv: 1010.1750. 58

[BDL11] Sergey Bravyi, David P. DiVincenzo, and Daniel Loss. Schrieffer-Wolff transformation for quantum many-body systems. *Annals of Physics*, 326(10):2793–2826, October 2011. arXiv: 1105.0675. 83, 84

[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. 164

[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. 3

[Ben89] Charles H. Bennett. Time/Space Trade-Offs for Reversible Computation. *SIAM Journal on Computing*, 18(4):766–776, 1989. 3

[Bes05] Arvid J. Bessen. Lower bound for quantum phase estimation. *Phys. Rev. A*, 71:042313, Apr 2005. arXiv: `quant-ph/0412008`. 16, 202

[BFL90] L. Babai, L. Fortnow, and C. Lund. Nondeterministic exponential time has two-prover interactive protocols. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 16–25 vol.1, October 1990. 11

[BG94] Mihir Bellare and Shafi Goldwasser. The Complexity of Decision Versus Search. *SIAM Journal on Computing*, 23(1):97–119, 1994. 10

[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. arXiv: `1911.07782`. 10, 41, 55, 56, 108

[BGLW24] Harry Buhrman, Dmitry Grinko, Philip Verduyn Lunel, and Jordi Weggemans. Permutation tests for quantum state identity, 2024. arXiv: `2405.09626`. vi

[BGS75] Theodore Baker, John Gill, and Robert Solovay. Relativizations of the P=?NP question. *SIAM Journal on computing*, 4(4):431–442, 1975. 225

[BGT21] Adam Bouland and Tudor Giurgica-Tiron. Efficient universal quantum compilation: An inverse-free Solovay-Kitaev algorithm, December 2021. arXiv: `2112.02040`. 32

[BGW24] Harry Buhrman, François Le Gall, and Jordi Weggemans. Classical versus quantum queries in quantum PCPs with classical proofs, November 2024. arXiv: `2411.00946`. vi

[BH13a] Fernando G.S.L. Brandao and Aram W. Harrow. Quantum de Finetti Theorems under Local Measurements with Applications. In *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '13, page 861–870, 2013. arXiv: `1210.6367`. 58

[BH13b] Fernando G.S.L. Brandao and Aram W. Harrow. Product-state approximations to quantum ground states. In *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '13, page 871–880, 2013. arXiv: 1310.0017. 12, 132, 164, 165

[BH17] Sergey Bravyi and Matthew Hastings. On complexity of the quantum Ising model. *Communications in Mathematical Physics*, 349(1):1–45, 2017. arXiv: 1410.0703. 84, 98

[BHMT02] Gilles Brassard, Peter Høyer, Michele Mosca, and Alain Tapp. Quantum amplitude amplification and estimation. *Contemporary Mathematics*, 305:53–74, 2002. arXiv: quant-ph/0005055. 196, 215, 218

[BHW25] Harry Buhrman, Jonas Helsen, and Jordi Weggemans. Quantum PCPs: on Adaptivity, Multiple Provers and Reductions to Local Hamiltonians. *Quantum*, 9:1791, July 2025. arXiv: 2403.04841. v, 137, 165

[BJ95] Nader H. Bshouty and Jeffrey C. Jackson. Learning DNF over the uniform distribution using a quantum example oracle. In *Proceedings of the Eighth Annual Conference on Computational Learning Theory*, COLT '95, page 118–127, 1995. 4, 16, 37, 230, 231

[BL08] Jacob D. Biamonte and Peter J. Love. Realizable Hamiltonians for universal adiabatic quantum computers. *Physical Review A*, 78(1):012352, 2008. arXiv: 0704.1287. 102

[BLMT24] Ainesh Bakshi, Allen Liu, Ankur Moitra, and Ewin Tang. High-Temperature Gibbs States are Unentangled and Efficiently Preparable . In *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1027–1036, October 2024. arXiv: 2403.16850. 43

[BMN+21] Ryan Babbush, Jarrod R. McClean, Michael Newman, Craig Gidney, Sergio Boixo, and Hartmut Neven. Focus beyond quadratic speedups for error-corrected quantum advantage. *PRX quantum*, 2(1):010103, 2021. arXiv: 2011.04149. 3

[BRS15] Alex Bocharov, Martin Roetteler, and Krysta M. Svore. Efficient synthesis of universal repeat-until-success quantum circuits. *Physical review letters*, 114(8):080502, 2015. arXiv: 1404.5320. 114

[BV93] Ethan Bernstein and Umesh Vazirani. Quantum complexity theory. In *Proceedings of the Twenty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '93, page 11–20, 1993. 14, 195

[BW24] Martijn Brehm and Jordi Weggemans. Assessing fault-tolerant quantum advantage for k-SAT with structure, 2024. arXiv: 2412.13274. vi, 3

[CAB+21] Marco Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and Patrick J. Coles. Variational quantum algorithms. *Nature Reviews Physics*, 3:625–644, 2021. arXiv: 2012.09265. 8

[CB24] Luuk Coopmans and Marcello Benedetti. On the sample complexity of quantum Boltzmann machine learning. *Communications Physics*, 7(1), August 2024. arXiv: 2306.14969. 232

[CCD+03] Andrew M. Childs, Richard Cleve, Enrico Deotto, Edward Farhi, Sam Gutmann, and Daniel A. Spielman. Exponential algorithmic speedup by a quantum walk. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '03, page 59–68, 2003. arXiv: quant-ph/0209131. 4, 167

[CCNN23a] Nolan J. Coble, Matthew Coudron, Jon Nelson, and Seyed Sajjad Nezhadi. Hamiltonians whose low-energy states require $\Omega(n)$ T gates, 2023. arXiv: 2310.01347. 12, 131

[CCNN23b] Nolan J. Coble, Matthew Coudron, Jon Nelson, and Seyed Sajjad Nezhadi. Local Hamiltonians with No Low-Energy Stabilizer States. In *18th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2023)*, volume 266, pages 14:1–14:21, 2023. arXiv: 2302.14755. 12, 131

[CCY93] A. Chi-Chih Yao. Quantum circuit complexity. In *Proceedings of the 1993 IEEE 34th Annual Foundations of Computer Science*, SFCS '93, page 352–361, 1993. 32, 33

[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)*, volume 261, pages 32:1–32:19, 2023. arXiv: 2207.10250. v, 12

[CFNW23] Chris Cade, Marten Folkertsma, Ido Niesen, and Jordi Weggemans. Quantifying Grover speed-ups beyond asymptotic analysis. *Quantum*, 7:1133, October 2023. arXiv: 2203.04975. vi, 3

[CFNW24] Chris Cade, Marten Folkertsma, Ido Niesen, and Jordi Weggemans. Quantum algorithms for community detection and their empirical run-times. *Quantum Information & Computation*, 24(5&6):0361–0410, 2024. arXiv: 2203.06208. vi, 3

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

[CGL+22] Nai-Hui Chia, András Pal Gilyén, Tongyang Li, Han-Hsuan Lin, Ewin Tang, and Chunhao Wang. Sampling-based Sublinear Low-rank Matrix Arithmetic Framework for Dequantizing Quantum Machine Learning. *J. ACM*, 69(5), October 2022. arXiv: 190.06151. 65

[Chi14] Giulio Chiribella. A First Course in Quantum Information Theory, 2014. Tsinghua University. 26

[Chi17] Andrew M. Childs. Lecture Notes on Quantum Algorithms, 2017. University of Maryland, Fall 2017. 32

[CHM21] Jordan Cotler, Hsin-Yuan Huang, and Jarrod R. McClean. Revisiting dequantization and quantum advantage in learning tasks, 2021. arXiv: 2112.00811. 65

[Chu36a] Alonzo Church. A note on the Entscheidungsproblem. *The journal of symbolic logic*, 1(1):40–41, 1936. 2

[Chu36b] Alonzo Church. An unsolvable problem of elementary number theory. *American journal of mathematics*, 58(2):345–363, 1936. 2

[CKBG23] Chi-Fang Chen, Michael J. Kastoryano, Fernando GSL Brandão, and András Gilyén. Quantum Thermal State Preparation, March 2023. arXiv: 2303.18224. 219

[CKG23] Chi-Fang Chen, Michael J. Kastoryano, and András Gilyén. An efficient and exact noncommutative quantum Gibbs sampler, November 2023. arXiv: 2311.09207. 221

[CKM19] Earl Campbell, Ankur Khurana, and Ashley Montanaro. Applying quantum algorithms to constraint satisfaction problems. *Quantum*, 3:167, July 2019. arXiv: 1810.05582. 3

[CKMR07] Matthias Christandl, Robert König, Graeme Mitchison, and Renato Renner. One-and-a-half quantum de Finetti theorems. *Communications in Mathematical Physics*, 273(2):473–498, 2007. arXiv: `quant-ph/0602130`. 58

[CM16] Toby Cubitt and Ashley Montanaro. Complexity Classification of Local Hamiltonian Problems. *SIAM Journal on Computing*, 45(2):268–316, 2016. arXiv: `1311.3161`. 101, 102, 104

[CMP18] Toby S. Cubitt, Ashley Montanaro, and Stephen Piddock. Universal quantum Hamiltonians. *Proceedings of the National Academy of Sciences*, 115(38):9497–9502, 2018. arXiv: `1701.05182`. 6, 97, 103, 104, 105

[CNY03] Thomas Chen, Shivam Nadimpalli, and Henry Yuen. Testing and Learning Quantum Juntas Nearly Optimally. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1163–1185, 2003. arXiv: `2207.05898`. 15, 201

[CPGSV21] J. Ignacio Cirac, David Perez-Garcia, Norbert Schuch, and Frank Verstraete. Matrix product states and projected entangled pair states: Concepts, symmetries, theorems. *Reviews of Modern Physics*, 93(4):045003, 2021. arXiv: `2011.12127`. 64

[CRO+19] Yudong Cao, Jonathan Romero, Jonathan P. Olson, Matthias Degroote, Peter D. Johnson, Mária Kieferová, Ian D. Kivlichan, Tim Menke, Borja Peropadre, Nicolas PD Sawaya, et al. Quantum chemistry in the age of quantum computing. *Chemical reviews*, 119(19):10856–10915, 2019. arXiv: `1812.09976`. 6, 44

[CS12] André Chailloux and Or Sattath. The Complexity of the Separable Hamiltonian Problem. In *2012 IEEE 27th Conference on Computational Complexity*, pages 32–41, 2012. arXiv: `1111.5247`. 159, 161

[CWZ24] Kean Chen, Qisheng Wang, and Zhicheng Zhang. Local test for unitarily invariant properties of bipartite quantum states, 2024. arXiv: `2404.04599`. 228

[DCS+95] P. Dai, B.C. Chakoumakos, G.F. Sun, K.W. Wong, Y. Xin, and D.F. Lu. Synthesis and neutron powder diffraction study of the superconductor $HgBa_2Ca_2Cu_3O_{8+\delta}$ by Tl substitution. *Physica C: Superconductivity*, 243(3):201–206, 1995. 43

[Deu85] David Deutsch. Quantum theory, the Church–Turing principle and the universal quantum computer. *Proceedings of the Royal Society*

of London. A. Mathematical and Physical Sciences, 400(1818):97–117, 1985. 3, 5, 32, 33

[DF80] Persi Diaconis and David Freedman. Finite Exchangeable Sequences. *The Annals of Probability*, pages 745–764, 1980. 58

[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. 50, 51, 88, 89, 119, 223

[DGLM24] Joao F. Doriguello, George Giapitzakis, Alessandro Luongo, and Aditya Morolia. On the practicality of quantum sieving algorithms for the shortest vector problem, 2024. arXiv: 22410.13759. 3

[Din07] Irit Dinur. The PCP Theorem by Gap Amplification. *Journal of the ACM*, 54(3):12–es, June 2007. 11, 131, 278

[DJ92] David Deutsch and Richard Jozsa. Rapid solution of problems by quantum computation. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 439(1907):553–558, 1992. 235

[DMB+23] Alexander M Dalzell, Sam McArdle, Mario Berta, Przemyslaw Bienias, Chi-Fang Chen, András Gilyén, Connor T Hann, Michael J Kastoryano, Emil T Khabiboulline, Aleksander Kubica, et al. Quantum algorithms: A survey of applications and end-to-end complexities, 2023. arXiv: 2310.03011. 3

[dW19] Ronald de Wolf. Quantum computing: Lecture notes, 2019. arXiv: 1907.09415. 19

[Fey82] Richard P. Feynman. Simulating Physics with Computers. *International Journal of Theoretical Physics*, 21(6/7), 1982. 3, 75

[Fey86] Richard P. Feynman. Quantum mechanical computers. *Foundations of Physics*, 16(6):507–531, 1986. 75

[FGKM15] Simon Forest, David Gosset, Vadym Kliuchnikov, and David McKinnon. Exact synthesis of single-qubit unitaries over Clifford-cyclotomic gate sets. *Journal of Mathematical Physics*, 56(8), 2015. arXiv: 1501.04944. 114

[FKS21] Steph Foulds, Viv Kendon, and Tim Spiller. The controlled SWAP test for determining quantum entanglement. *Quantum Science and Technology*, 6(3):035002, 2021. arXiv: 2009.07613. 214

[FL16] Bill Fefferman and Cedric Lin. Quantum Merlin Arthur with exponentially small gap, 2016. arXiv: 1601.01975. 203

[Gha12] Sevag Gharibian. *Approximation, Proof Systems, and Correlations in a Quantum World*. PhD thesis, University of Waterloo, 2012. 45

[Gha24] Sevag Gharibian. Guest Column: The 7 faces of quantum NP. *SIGACT News*, 54(4):54–91, January 2024. arXiv: 2310.18010. 13, 132

[GHGM22] Sevag Gharibian, Ryu Hayakawa, François Le Gall, and Tomoyuki Morimae. Improved Hardness Results for the Guided Local Hamiltonian Problem, 2022. arXiv: 2207.10250v2. v

[GHST20] Hrant Gharibyan, Masanori Hanada, Brian Swingle, and Masaki Tezuka. Characterization of quantum chaos by two-point correlation functions. *Physical Review E*, 102(2):022213, 2020. arXiv: 1902.11086. 9

[GK24] Sevag Gharibian and Jonas Kamminga. BQP, Meet NP: Search-To-Decision Reductions and Approximate Counting. In *51st International Colloquium on Automata, Languages, and Programming (ICALP 2024)*, volume 297 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 70:1–70:19, 2024. arXiv: 2401.03943. 109

[GKC+24] Xun Gao, Marcin Kalinowski, Chi-Ning Chou, Mikhail D. Lukin, Boaz Barak, and Soonwon Choi. Limitations of Linear Cross-Entropy as a Measure for Quantum Advantage. *PRX Quantum*, 5:010334, Feb 2024. arXiv: 2112.01657. 229

[GKK+07] Dmitry Gavinsky, Julia Kempe, Iordanis Kerenidis, Ran Raz, and Ronald de Wolf. Exponential separations for one-way quantum communication complexity, with applications to cryptography. In *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing*, STOC '07, page 516–525, 2007. arXiv: quant-ph/0611209. 4, 230

[GKZ19] Alex B. Grilo, Iordanis Kerenidis, and Timo Zijlstra. Learning-with-errors problem is easy with quantum samples. *Phys. Rev. A*, 99:032314, Mar 2019. arXiv: 1702.08255. 4, 230

[GL22] Sevag Gharibian and François Le Gall. Dequantizing the Quantum singular value transformation: hardness and applications to

Quantum chemistry and the Quantum PCP conjecture. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2022, page 19–32, 2022. arXiv: 2111.09079. 7, 8, 12, 64, 65, 70, 75, 76, 77, 78, 80, 83, 88, 132, 273, 279

[GLM08] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Quantum Random Access Memory. *Phys. Rev. Lett.*, 100:160501, Apr 2008. arXiv: 0708.1879. 65

[GM23] Dar Gilboa and Jarrod R. McClean. Exponential Quantum Communication Advantage in Distributed Inference and Learning, 2023. arXiv: 2310.07136. 230

[Göd31] Kurt Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für mathematik und physik*, 38:173–198, 1931. 1, 2

[GPY20] Sevag Gharibian, Stephen Piddock, and Justin Yirka. Oracle Complexity Classes and Local Measurements on Physical Hamiltonians. In *37th International Symposium on Theoretical Aspects of Computer Science (STACS 2020)*, volume 154 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 20:1–20:37, 2020. arXiv: 1909.05981. 109

[GR02] Lov Grover and Terry Rudolph. Creating superpositions that correspond to efficiently integrable probability distributions, August 2002. arXiv: quant-ph/0208112. 80

[Gri18] Alex B. Grilo. *Quantum proofs, the local Hamiltonian problem and applications*. PhD thesis, Université Sorbonne Paris Cité, April 2018. 13, 132, 133, 137, 142, 158

[Gro96] Lov K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '96, page 212–219, 1996. arXiv: quant-ph/9605043. 15, 201, 236

[GSLW19] András Gilyén, Yuan Su, Guang Hao Low, and Nathan Wiebe. Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC '19, page 193–204, 2019. arXiv: 1806.01838. 77, 222

[GY19] Sevag Gharibian and Justin Yirka. The complexity of simulating local measurements on quantum systems. *Quantum*, 3:189, 2019. arXiv: 1606.05626. 85, 109, 117

[Han16] Steve Hanneke. The Optimal Sample Complexity of PAC Learning. *Journal of Machine Learning Research*, 17(38):1–15, 2016. arXiv: 1507.00473. 232

[Has07] M. B. Hastings. An area law for one-dimensional quantum systems. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(08):P08024, aug 2007. arXiv: 0705.2024. 7, 63

[HATH24] Yaroslav Herasymenko, Anurag Anshu, Barbara M. Terhal, and Jonas Helsen. Fermionic Hamiltonians without trivial low-energy states. *Phys. Rev. A*, 109:052431, May 2024. arXiv: 2307.13730. 12, 13, 131, 132, 133

[Hau92] David Haussler. Decision theoretic generalizations of the PAC model for neural net and other learning applications. *Information and Computation*, 100(1):78–150, 1992. 232

[HB93] Murray J. Holland and Keith Burnett. Interferometric detection of optical phase shifts at the Heisenberg limit. *Physical review letters*, 71(9):1355, 1993. 203

[HC17] Guang Hao Low and Isaac L. Chuang. Hamiltonian Simulation by Uniform Spectral Amplification, July 2017. arXiv: 1707.05391. 274, 275

[Hel69] Carl W. Helstrom. Quantum detection and estimation theory. *Journal of Statistical Physics*, 1:231–252, 1969. 238

[Hil00] David Hilbert. Mathematische Probleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1900:253–297, 1900. 1

[HJ12] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge University Press, 2012. 209

[HKOT23] Jeongwan Haah, Robin Kothari, Ryan O'Donnell, and Ewin Tang. Query-optimal estimation of unitary channels in diamond distance. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 363–390, nov 2023. arXiv: 2302.14066. 207

[HM13] Aram W. Harrow and Ashley Montanaro. Testing Product States, Quantum Merlin-Arthur Games and Tensor Optimization. *J. ACM*, 60(1), February 2013. arXiv: 1001.0017. 137, 205, 227

[HM17] Aram W Harrow and Ashley Montanaro. Quantum computational supremacy. *Nature*, 549(7671):203–209, 2017. arXiv: 1809.07442. 3

[Hø00] Peter Høyer. Arbitrary phases in quantum amplitude amplification. *Phys. Rev. A*, 62:052304, Oct 2000. arXiv: quant-ph/0006031. 215

[Hol73] Alexander Semenovich Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problemy Peredachi Informatsii*, 9(3):3–11, 1973. 17

[Hå01] Johan Håstad. Some optimal inapproximability results. *Journal of the ACM (JACM)*, 48(4):798–859, July 2001. 11, 190, 191

[HRC02] Aram W. Harrow, Benjamin Recht, and Isaac L. Chuang. Efficient discrete approximations of quantum gates. *Journal of Mathematical Physics*, 43(9):4445–4451, 2002. arXiv: quant-ph/0111031. 114

[HTFS23] Hsin-Yuan Huang, Yu Tong, Di Fang, and Yuan Su. Learning many-body Hamiltonians with Heisenberg-limited scaling. *Physical Review Letters*, 130(20):200403, 2023. 223

[HZN+20] Cupjin Huang, Fang Zhang, Michael Newman, Junjie Cai, Xun Gao, Zhengxiong Tian, Junyin Wu, Haihong Xu, Huanjun Yu, Bo Yuan, et al. Classical simulation of quantum supremacy circuits, 2020. arXiv: 2005.06787. 229

[IL89] Russell Impagliazzo and Michael Luby. One-way functions are essential for complexity based cryptography. In *30th Annual Symposium on Foundations of Computer Science*, pages 230–235. IEEE Computer Society, 1989. 230

[INN+22] Sandy Irani, Anand Natarajan, Chinmay Nirkhe, Sujit Rao, and Henry Yuen. Quantum Search-To-Decision Reductions and the State Synthesis Problem. In *37th Computational Complexity Conference (CCC 2022)*, 2022. arXiv: 2111.02999. 10, 107, 108, 109, 128

[IW18] Eric B. Isaacs and Chris Wolverton. Inverse band structure design via materials database screening: application to square planar thermoelectrics. *Chemistry of Materials*, 30(5):1540–1546, 2018. 6

[JGL10] Stephen P. Jordan, David Gosset, and Peter J. Love. Quantum-Merlin-Arthur–complete problems for stoquastic Hamiltonians and Markov matrices. *Phys. Rev. A*, 81:032331, Mar 2010. arXiv: 0905.4755. 6

[JLGS20] Dhawal Jethwani, François Le Gall, and Sanjay K. Singh. Quantum-Inspired Classical Algorithms for Singular Value Transformation. In *45th International Symposium on Mathematical Foundations of Computer Science (MFCS 2020)*, volume 170 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 53:1–53:14, 2020. arXiv: 1910.05699. 65

[JW93] Pascual Jordan and Eugene Paul Wigner. *Über das Paulische Äquivalenzverbot*. Springer, 1993. 44

[Kit95] Alexei Y. Kitaev. Quantum measurements and the Abelian Stabilizer Problem. *Electron. Colloquium Comput. Complex.*, TR96, 1995. arXiv: quant-ph/9511026. 8, 48, 75, 213, 225

[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. 6, 50, 83

[KL20] J. Katz and Y. Lindell. *Introduction to Modern Cryptography*. Chapman & Hall/CRC Cryptography and Network Security Series. CRC Press, 2020. 251

[Kle36] Stephen Cole Kleene. General recursive functions of natural numbers. *Mathematische annalen*, 112(1):727–742, 1936. 2

[KMM15] Vadym Kliuchnikov, Dmitri Maslov, and Michele Mosca. practical approximation of single-qubit unitaries by single-qubit quantum Clifford and T circuits. *IEEE Transactions on Computers*, 65(1):161–172, 2015. arXiv: 1212.6964. 114

[KMY03] Hirotada Kobayashi, Keiji Matsumoto, and Tomoyuki Yamakami. Quantum Merlin-Arthur proof systems: Are multiple Merlins more helpful to Arthur? In *Algorithms and Computation: 14th International Symposium, ISAAC 2003*, pages 189–198, 2003. arXiv: quant-ph/0306051. 215

[Kni04] Emanuel Knill. Fault-tolerant postselected quantum computation: Threshold analysis, 2004. arXiv: quant-ph/0404104. 260

[KPV24] John Kallaugher, Ojas Parekh, and Nadezhda Voronova. Exponential Quantum Space Advantage for Approximating Maximum

Directed Cut in the Streaming Model. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, STOC 2024, page 1805–1815, 2024. arXiv: 2311.14123. 4, 230

[KR03] Julia Kempe and Oded Regev. 3-local Hamitonian is QMA-complete. *Quantum Info. Comput.*, 3(3):258–264, may 2003. arXiv: quant-ph/0302079. 50, 89, 92, 118, 119

[KR24] Isaac H. Kim and Daniel Ranard. Classifying 2D topological phases: mapping ground states to string-nets, 2024. arXiv: 2405.17379. 63

[Kre88] Mark W. Krentel. The complexity of optimization problems. *Journal of Computer and System Sciences*, 36(3):490–509, 1988. 107

[KST12] Johannes Kobler, Uwe Schöning, and Jacobo Torán. *The graph isomorphism problem: its structural complexity*. Springer Science & Business Media, 2012. 170, 171

[KSV02] Alexei Y. Kitaev, Alexander Shen, and Mikhail N. Vyalyi. *Classical and quantum computation*. American Mathematical Society, 2002. 45, 46, 47, 51, 75, 108, 150, 151

[KSZ93] A. Klümper, A. Schadschneider, and J. Zittartz. Matrix Product Ground States for One-Dimensional Spin-1 Quantum Antiferromagnets. *Europhysics Letters*, 24(4):293, nov 1993. arXiv: cond-mat/9307028. 63

[LBG+21] Joonho Lee, Dominic W. Berry, Craig Gidney, William J. Huggins, Jarrod R. McClean, Nathan Wiebe, and Ryan Babbush. Even More Efficient Quantum Computations of Chemistry Through Tensor Hypercontraction. *PRX Quantum*, 2:030305, 2021. arXiv: 2011.03494. 75

[LC17] Guang Hao Low and Isaac L Chuang. Optimal Hamiltonian simulation by quantum signal processing. *Physical review letters*, 118(1):010501, 2017. 219

[LC19] Guang Hao Low and Isaac L. Chuang. Hamiltonian simulation by qubitization. *Quantum*, 3:163, 2019. 219

[Lee20] Jasper Lee. Lecture 3: Concentration Inequalities and Mean Estimation. Lecture Notes, CSCI 1951-W Sublinear Algorithms for Big Data, Fall 2020, 2020. 57

[Lin22] Lin Lin. Lecture notes on quantum algorithms for scientific computation, 2022. arXiv: 2201.0830. 77

[Liu06] Yi-Kai Liu. Consistency of local density matrices is QMA-complete. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques: 9th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, APPROX 2006 and 10th International Workshop on Randomization and Computation, RANDOM 2006*, pages 438–449, 2006. arXiv: `quant-ph/0604166`. 10, 41, 55, 108, 109

[Llo96] Seth Lloyd. Universal Quantum Simulators. *Science*, 273(5278):1073–1078, 1996. 9, 42

[LPS86] Alexander Lubotzky, Ralph Phillips, and Peter Sarnak. Hecke operators and distributing points on the sphere I. *Communications on Pure and Applied Mathematics*, 39(S1):S149–S186, 1986. 114

[LR88] Michael Luby and Charles Rackoff. How to Construct Pseudorandom Permutations from Pseudorandom Functions. *SIAM Journal on Computing*, 17(2):373–386, 1988. 251

[LT20a] Lin Lin and Yu Tong. Optimal polynomial based quantum eigenstate filtering with application to solving quantum linear systems. *Quantum*, 4:361, November 2020. arXiv: `1910.14596`. 8, 77

[LT20b] Lin Lin and Yu Tong. Near-optimal ground state preparation. *Quantum*, 4:372, December 2020. arXiv: `2002.12508`. 8, 77, 223, 224

[LVV15] Zeph Landau, Umesh Vazirani, and Thomas Vidick. A polynomial time algorithm for the ground state of one-dimensional gapped local Hamiltonians. *Nature Physics*, 11(7):566–569, 2015. arXiv: `1307.5143`. 63

[MdW13] Ashley Montanaro and Ronald de Wolf. A Survey of Quantum Property Testing. *Theory of Computing*, 2016:1–81, 10 2013. arXiv: `1310.2035`. 203

[Mic96] Albert A. Michelson. Dedication of Ryerson Physical Laboratory, 1896. Originally presented in 1894. 2

[MLA+22] Lars S. Madsen, Fabian Laudenbach, Mohsen Falamarzi Askarani, Fabien Rortais, Trevor Vincent, Jacob F.F. Bulmer, Filippo M. Miatto, Leonhard Neuhaus, Lukas G. Helt, Matthew J. Collins, et al. Quantum computational advantage with a programmable photonic processor. *Nature*, 606(7912):75–81, 2022. 229

[Mon16] Ashley Montanaro. Quantum algorithms: an overview. *npj Quantum Information*, 2(1):1–8, 2016. arXiv: 1511.04206. 3

[MT25] Daniel Malz and Rahul Trivedi. Computational Complexity of Isometric Tensor-Network States. *PRX Quantum*, 6:020310, Apr 2025. arXiv: 2402.07975. 64, 271

[MW05] Chris Marriott and John Watrous. Quantum Arthur-Merlin games. *Computational Complexity*, 14(2):122–152, 2005. arXiv: cs/0506068. 91, 165

[MZHO16] Alex F. Mendelson, Maria A. Zuluaga, Brian F. Hutton, and Sébastien Ourselin. What is the distribution of the number of unique original items in a bootstrap sample?, 2016. arXiv: 1602.05822. 249

[Nai40] M. A. Naimark. Spectral functions of a symmetric operator. *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya*, 4(3):277–318, 1940. 23

[NC10] Michael A. Nielsen and Isaac L. Chuang. *Quantum computation and quantum information*. Cambridge university press, 2010. 19, 30, 114

[NOAO+18] Tammie R. Nelson, Dianelys Ondarse-Alvarez, Nicolas Oldani, Beatriz Rodriguez-Hernandez, Laura Alfonso-Hernandez, Johan F. Galindo, Valeria D. Kleiman, Sebastian Fernandez-Alberti, Adrian E. Roitberg, and Sergei Tretiak. Coherent exciton-vibrational dynamics and energy transfer in conjugated organics. *Nature communications*, 9(1):2316, 2018. 9

[NW99] Ashwin Nayak and Felix Wu. The quantum query complexity of approximating the median and related statistics. In *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*, STOC '99, page 384–393, 1999. arXiv: quant-ph/9804066. 213

[NWB+20] Tammie R. Nelson, Alexander J. White, Josiah A. Bjorgaard, Andrew E. Sifain, Yu Zhang, Benjamin Nebgen, Sebastian Fernandez-Alberti, Dmitry Mozyrsky, Adrian E. Roitberg, and Sergei Tretiak. Non-adiabatic excited-state molecular dynamics: Theory and applications for modeling photophysics in extended molecular materials. *Chemical reviews*, 120(4):2215–2287, 2020. 9

[NZS24] Junhong Nie, Wei Zi, and Xiaoming Sun. Quantum circuit for multi-qubit Toffoli gate with optimal resource, 2024. arXiv: 2402.05053. 259

[OIWF22] Bryan O'Gorman, Sandy Irani, James Whitfield, and Bill Fefferman. Intractability of Electronic Structure in a Fixed Basis. *PRX Quantum*, 3:020322, May 2022. arXiv: 2103.08215. 6

[OJF23] Changhun Oh, Liang Jiang, and Bill Fefferman. spoofing cross-entropy measure in boson sampling. *Physical Review Letters*, 131(1):010401, 2023. arXiv: 2210.15021. 229

[OOWK+21] L. A. B. Olde Olthof, J. R. Weggemans, G. Kimbell, J. W. A. Robinson, and X. Montiel. Tunable critical field in Rashba superconductor thin films. *Phys. Rev. B*, 103:L020504, Jan 2021. arXiv: 2009.14592. vi

[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. 270

[Osb12] Tobias J. Osborne. Hamiltonian complexity. *Reports on progress in physics*, 75(2):022001, 2012. arXiv: 1106.5875. 42

[OT08] Roberto Oliveira and Barbara M. Terhal. The complexity of quantum spin systems on a two-dimensional square lattice. *Quantum Information and Computation*, 8(10):900–924, November 2008. arXiv: quant-ph/0504050. 6, 103, 104

[Ou96] ZY Ou. Complementarity and fundamental limit in precision phase measurement. *Physical review letters*, 77(12):2352, 1996. 203

[Pap03] Christos H. Papadimitriou. *Computational complexity*, page 260–265. John Wiley and Sons Ltd., 2003. 34

[Pau03] Vern Paulsen. *Completely bounded maps and operator algebras.* Cambridge University Press, 2003. 23

[PCZ22] Feng Pan, Keyang Chen, and Pan Zhang. Solving the Sampling Problem of the Sycamore Quantum Circuits. *Phys. Rev. Lett.*, 129:090502, Aug 2022. 229

[PM17] Stephen Piddock and Ashley Montanaro. The complexity of antiferromagnetic interactions and 2d lattices. *Quantum Information & Computation*, 17(7-8):636–672, 2017. arXiv: 1506.04014. 102, 103

[PMS+14] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O'Brien. A variational eigenvalue solver on a photonic quantum processor. *Nature Communications*, 5(1), July 2014. arXiv: 1304.3061. 229

[PS18] Ori Parzanchevski and Peter Sarnak. Super-golden-gates for PU(2). *Advances in Mathematics*, 327:869–901, 2018. arXiv: 1704.02106. 114

[Raz99] Ran Raz. Exponential separation of quantum and classical communication complexity. In *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*, STOC '99, page 358–367, 1999. 4

[RCW75] Dijen K. Ray-Chaudhuri and Richard M. Wilson. On t-designs. *Osaka Journal of Mathematics*, 12(3):737 – 744, 1975. 257

[Ren07] Renato Renner. Symmetry of large physical systems implies independence of subsystems. *Nature Physics*, 3(9):645–649, 2007. arXiv: quant-ph/0703069. 58

[RR95] Kenneth W. Regan and James S. Royer. On closure properties of bounded two-sided error complexity classes. *Mathematical systems theory*, 28:229–243, 1995. 174

[RS16] Neil J. Ross and Peter Selinger. Optimal ancilla-free Clifford+T approximation of z-rotations. *Quantum Information & Computation*, 16(11–12):901–953, sep 2016. arXiv: 1403.2975. 114

[RT19] Ran Raz and Avishay Tal. Oracle separation of BQP and PH. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2019, page 13–23, 2019. 191, 192

[RWS+17] Markus Reiher, Nathan Wiebe, Krysta M. Svore, Dave Wecker, and Matthias Troyer. Elucidating reaction mechanisms on quantum computers. *Proceedings of the National Academy of Sciences*, 114(29):7555–7560, 2017. arXiv: 1605.03590. 75

[SBA24] Chris N. Self, Marcello Benedetti, and David Amaro. Protecting expressive circuits with a quantum error detection code. *Nature Physics*, 20(2):219–224, January 2024. arXiv: 2211.06703. 260

[SBW+21] Yuan Su, Dominic W. Berry, Nathan Wiebe, Nicholas Rubin, and Ryan Babbush. Fault-Tolerant Quantum Simulations of Chemistry in First Quantization. *PRX Quantum*, 2:040332, Nov 2021. arXiv: 2105.12767. 75

[SC21] Lucas Slattery and Bryan K. Clark. Quantum circuits for two-dimensional isometric tensor networks, 2021. arXiv: 2108.02792. 271

[Sch89] Uwe Schöning. Probabilistic complexity classes and lowness. *Journal of Computer and System Sciences*, 39(1):84–100, 1989. 168, 170, 171

[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. arXiv: cond-mat/0409292. 270

[Sha92] Adi Shamir. IP = PSPACE. *Journal of the ACM (JACM)*, 39(4):869–877, October 1992. 11

[Sho94] Peter W. Shor. Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings of the 35th Annual Symposium on Foundations of Computer Science*, SFCS '94, page 124–134, 1994. 3, 229

[Sho96] Peter W. Shor. Fault-tolerant quantum computation. In *Proceedings of 37th Conference on Foundations of Computer Science*, pages 56–65, 1996. arXiv: quant-ph/9605011. 3, 33

[Sim97] Daniel R. Simon. On the Power of Quantum Computation. *SIAM Journal on Computing*, 26(5):1474–1483, 1997. 4, 14, 167

[SSB+20] Tomohiro Soejima, Karthik Siva, Nick Bultinck, Shubhayu Chatterjee, Frank Pollmann, and Michael P. Zaletel. Isometric tensor network representation of string-net liquids. *Physical Review B*, 101(8):085117, 2020. arXiv: 1908.07545. 64

[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. 270

[SV09] Norbert Schuch and Frank Verstraete. Computational complexity of interacting electrons and fundamental limitations of density functional theory. *Nature Physics*, 5:732–735, 2009. arXiv: 0712.0483. 6, 75

[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. arXiv: quant-ph/0611050. 64, 271

[SY23] Adrian She and Henry Yuen. Unitary Property Testing Lower Bounds by Polynomials. In *14th Innovations in Theoretical Computer Science Conference (ITCS 2023)*, volume 251, pages 96:1–96:17, 2023. arXiv: 2210.05885. 15, 201, 203, 204, 205

[Tan19] Ewin Tang. A quantum-inspired classical algorithm for recommendation systems. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2019, page 217–228, 2019. arXiv: 1807.04271. 65

[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. 71, 270

[Tro12] Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12:389–434, 2012. 153

[Tur36] Alan Mathison Turing. On computable numbers, with an application to the entscheidungsproblem. *J. of Math*, 58(345-363):5, 1936. 1, 31

[Uhl76] A. Uhlmann. The "transition probability" in the state space of a *-algebra. *Reports on Mathematical Physics*, 9(2):273–279, 1976. 25

[VC04] Frank Verstraete and J. Ignacio Cirac. Renormalization algorithms for quantum-many body systems in two and higher dimensions, 2004. arXiv: cond-mat/0407066. 63

[vEB91] Peter van Emde Boas. *Machine models and simulations*, page 1–66. MIT Press, 1991. 2

[Vid08] Guifré Vidal. Class of quantum many-body states that can be efficiently simulated. *Physical review letters*, 101(11):110501, 2008. arXiv: quant-ph/0610099. 63

[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. arXiv: 0907.2796. 270

[Wan11] Guoming Wang. Property testing of unitary operators. *Phys. Rev. A*, 84:052328, Nov 2011. arXiv: 1110.1133. 15, 201

[Wat08] John Watrous. Quantum computational complexity, 2008. arXiv: 0804.3401. 19, 34

[Wat18] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018. 19, 30

[Weg24] Jordi Weggemans. Finding quantum partial assignments by search-to-decision reductions, 2024. arXiv: 2408.03986. v, 118

[Weg25] Jordi Weggemans. Lower Bounds for Unitary Property Testing with Proofs and Advice. *Quantum*, 9:1717, April 2025. arXiv: 2401.07912. vi

[WFC24] Jordi Weggemans, Marten Folkertsma, and Chris Cade. Guidable Local Hamiltonian Problems with Implications to Heuristic Ansatz State Preparation and the Quantum PCP Conjecture. In *19th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2024)*, volume 310, pages 10:1–10:24, 2024. arXiv: 2302.11578. v, 12, 73

[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. 87, 88

[WLZ23] Braden M. Weight, Xinyang Li, and Yu Zhang. Theory and modeling of light-matter interactions in chemistry: current and future. *Phys. Chem. Chem. Phys.*, 25:31554–31577, 2023. 9

[WMN10] Tzu-Chieh Wei, Michele Mosca, and Ashwin Nayak. Interacting Boson Problems Can Be QMA Hard. *Phys. Rev. Lett.*, 104:040501, Jan 2010. arXiv: 0905.3413. 6, 75

[Wol23] Michael M. Wolf. Lecture notes: Mathematical Introduction to Quantum Information Processing, March 2023. 207

[WUR+22] Jordi Weggemans, Alexander Urech, Alexander Rausch, Robert Spreeuw, Richard Boucherie, Florian Schreck, Kareljan Schoutens, Jiří Minář, and Florian Speelman. Solving correlation clustering with QAOA and a Rydberg qudit system: a full-stack approach. *Quantum*, 6:687, April 2022. arXiv: 2106.11672. vi

[WWL22] Ze-Guo Wang, Shi-Jie Wei, and Gui-Lu Long. A quantum circuit design of AES requiring fewer quantum qubits and gate operations. *Frontiers of Physics*, 17(4):41501, 2022. arXiv: 2109.12354. 260

[WZ23] Qisheng Wang and Zhicheng Zhang. Quantum lower bounds by sample-to-query lifting, August 2023. arXiv: 2308.01794. 15, 16, 201, 214, 216, 228

[Yao77] Andrew Chi-Chin Yao. Probabilistic computations: Toward a unified measure of complexity. In *Proceedings of the 18th Annual Symposium on Foundations of Computer Science*, SFCS '77, page 222–227, 1977. 246

[YZ24] Takashi Yamakawa and Mark Zhandry. Verifiable quantum advantage without structure. *Journal of the ACM (JACM)*, 71(3), June 2024. arXiv: 2204.02063. 229

[ZA21] Leo Zhou and Dorit Aharonov. Strongly universal Hamiltonian simulators, 2021. arXiv: 2102.02991. 98, 103

[Zal99] Christof Zalka. A Grover-based quantum search of optimal order for an unknown number of marked elements, 1999. arXiv: quant-ph/9902049. 245

[ZCC+22] Qingling Zhu, Sirui Cao, Fusheng Chen, Ming-Cheng Chen, Xiawei Chen, Tung-Hsun Chung, Hui Deng, Yajie Du, Daojin Fan, Ming Gong, et al. Quantum computational advantage via 60-qubit 24-cycle random circuit sampling. *Science bulletin*, 67(3):240–245, 2022. arXiv: 2109.03494. 229

[ZLTN20] Yu Zhang, Linqiu Li, Sergei Tretiak, and Tammie Nelson. Nonadiabatic excited-state molecular dynamics for open-shell systems. *Journal of Chemical Theory and Computation*, 16(4):2053–2064, 2020. 9

[ZP20] Michael P. Zaletel and Frank Pollmann. Isometric tensor network states in two dimensions. *Physical review letters*, 124(3):037201, 2020. arXiv: 1902.05100. 64

[ZS10] Mário Ziman and Michal Sedlák. Single-shot discrimination of quantum unitary processes. *Journal of Modern Optics*, 57(3):253–259, 2010. arXiv: 1003.1488. 207

[ZWD+20] Han-Sen Zhong, Hui Wang, Yu-Hao Deng, Ming-Cheng Chen, Li-Chao Peng, Yi-Han Luo, Jian Qin, Dian Wu, Xing Ding, Yi Hu, Peng Hu, Xiao-Yan Yang, Wei-Jun Zhang, Hao Li, Yuxuan Li, Xiao Jiang, Lin Gan, Guangwen Yang, Lixing You, Zhen Wang, Li Li, Nai-Le Liu, Chao-Yang Lu, and Jian-Wei Pan. Quantum computational advantage using photons. *Science*, 370(6523):1460–1463, 2020. arXiv: 2012.01625. 229

Samenvatting

Deze dissertatie onderzoekt de wisselwerking tussen kwantum- en klassieke computationele middelen vanuit het perspectief van de computationele complexiteitstheorie. De tekst is opgebouwd uit drie delen, die elk een ander aspect van dit onderwerp behandelen.

Deel I richt zich op laag-energetische toestanden van kwantumsystemen, die een belangrijke rol spelen in kwantumveeldeeltjessystemen en kwantumchemie. We bestuderen de complexiteit van het schatten van grondtoestands- en aangeslagen energieniveaus van lokale Hamiltonianen, gegeven toegang tot een zogenaamde hulptoestand met niet-verwaarloosbare overlap met de relevante eigensubruimte. In Hoofdstuk 3 formaliseren we verschillende toegangstypen tot zulke hulptoestanden en onderzoeken we hun eigenschappen en onderlinge relaties. In Hoofdstuk 4 tonen we aan dat deze taak voor bepaalde fysisch gemotiveerde 2-lokale Hamiltonianen BQP-compleet is voor een breed scala aan parameters. We laten ook zien dat, wanneer we slechts het bestaan van de hulptoestand veronderstellen (zonder expliciete toegang ertoe), het probleem QCMA-compleet wordt. Onder gebruikelijke aannames in de computationele complexiteitstheorie laten deze resultaten zien dat er een praktisch gemotiveerde setting bestaat, in de context van het berekenen van grondtoestandsenergieniveaus, waarin kwantumcomputers aantoonbaar superpolynomiaal efficiënter zijn dan klassieke computers. In Hoofdstuk 5 onderzoeken we of beschrijvingen van grondtoestanden kunnen worden verkregen met toegang tot een QMA-orakel. In de meest algemene vorm komt dit neer op het vinden van een kwantumbewijs voor een willekeurig probleem in QMA. Hoewel bekend is dat dit niet mogelijk is ten opzichte van een orakel, tonen we aan dat het wel mogelijk is om klassieke benaderingen van alle lokale gereduceerde dichtheidsmatrices van een bijna-optimaal bewijs te berekenen met een efficiënt klassiek algoritme dat toegang heeft tot een QMA-orakel.

Deel II richt zich op kwantum-probabilistisch verifieerbare bewijssystemen (QPCP-systemen). In Hoofdstuk 6 definiëren we een algemene klasse van QPCP's waarin adaptiviteit en meerdere niet-verstrengelde bewijzen zijn toegestaan. Via

kwantumreducties tonen we aan dat deze systemen equivalent zijn aan lokale Hamiltoniaanproblemen met een constante beloftescheiding. Hieruit volgt onder meer dat adaptiviteit geen extra kracht toevoegt in dit model, en dat constante-query QPCP's voor $\mathsf{QMA}(2)$ impliceren dat $\mathsf{QMA} = \mathsf{QMA}(2)$. In Hoofdstuk 7 bestuderen we kwantum-klassieke PCP's (QCPCP's), waarin een kwantumverifieerder ofwel klassieke ofwel kwantumqueries mag uitvoeren naar een klassiek bewijs. We tonen aan dat elke constante kwantum-query QCPCP met inverse-polynoom beloftescheiding kan worden gesimuleerd door een klassieke constante-query QCPCP met constante beloftescheiding. De corresponderende klasse van problemen ligt in $\mathsf{BQ} \cdot \mathsf{NP}$, de klasse van belofteproblemen die via een kwantumreductie herleidbaar zijn tot 3-SAT. Deze resultaten geven sterke aanwijzingen dat het onwaarschijnlijk is dat alle problemen in QCMA kunnen worden opgelost met behulp van kwantum-klassieke PCP's.

Deel III onderzoekt alternatieve complexiteitsmaten, in het bijzonder unitaire querycomplexiteit en samplecomplexiteit. In Hoofdstuk 8 introduceren we een algemene techniek om ondergrenzen te bewijzen voor unitaire querycomplexiteit via reducties naar unitaire kanaaldiscriminatie. Deze techniek blijft geldig in de aanwezigheid van kwantumadvies of -bewijzen. Als direct gevolg tonen we het bestaan aan van kwantumorakels ten opzichte waarvan $\mathsf{QMA}(2) \not\supset \mathsf{SBQP}$ en $\mathsf{QMA/qpoly} \not\supset \mathsf{SBQP}$. In Hoofdstuk 9 introduceren en analyseren we een zogenaamde complement sampling-taak, waarvoor we aantonen dat een klassieke computer een exponentieel aantal samples nodig heeft, terwijl een kwantumcomputer de taak kan oplossen met slechts één kwantumsample. Dit resulteert in de maximaal mogelijke scheiding tussen klassieke en kwantum samplecomplexiteit. Daarnaast laten we zien dat, onder standaard cryptografische aannames, de taak efficiënt verifieerbaar is, klassiek moeilijk is en uitvoerbaar is op relatief eenvoudige kwantumapparatuur. Hiermee bieden we een nieuw scenario waarin quantumtechnologie taken kan uitvoeren die klassiek praktisch onuitvoerbaar zijn.

Abstract

This dissertation studies the interplay between quantum and classical computational resources through the lens of computational complexity theory. It is structured in three parts, each addressing a different aspect of this topic.

Part I considers low-energy states of quantum systems, which play an important role in quantum many-body physics and quantum chemistry. We study the complexity of estimating ground and excited state energies of local Hamiltonians, given access to a guiding state promised to have non-negligible overlap with the relevant eigenspace. In Chapter 3, we formalize access models for such guiding states and study some of their properties and relations. In Chapter 4, we show that, for certain physically motivated 2-local Hamiltonians, this task is BQP-complete for a large range of input parameter settings. If the guiding state is only promised to exist, we show that the problem becomes QCMA-complete. Under standard complexity-theoretic assumptions, these results establish a well-defined, practically motivated setting in which quantum computers are provably super-polynomially more efficient than classical ones. In Chapter 5, we ask whether approximate descriptions of ground states can be obtained given QMA oracle access. In the fully general setting, this would correspond to finding a near-optimal quantum witness for any problem in QMA. Whilst this is known not to be possible relative to an oracle, we prove that it is possible to compute classical approximations of all low-locality reduced density matrices of a near-optimal witness, using a classical polynomial-time algorithm with such oracle access.

Part II focuses on quantum probabilistically checkable proof (QPCP) systems. In Chapter 6, we define a general class of QPCPs, allowing adaptivity and multiple unentangled provers. We show, via quantum reductions, that they are equivalent to constant-gap local Hamiltonian problems. As a consequence, we prove that adaptivity does not increase verifier power in this model, and that constant-query multi-prover QPCPs for QMA(2) imply QMA = QMA(2). In Chapter 7, we study quantum–classical PCPs (QCPCPs), where a quantum verifier has either classical or quantum query access to a classical proof. We prove that any constant

quantum-query QCPCP with inverse-polynomial promise gap can be simulated by a constant-classical-query QCPCP with constant promise gap, and that the corresponding class of problems lies in $\mathsf{BQ} \cdot \mathsf{NP}$, the class of all promise problems that have a quantum reduction to 3-SAT. This gives strong evidence that the power of QCMA cannot be captured by a quantum–classical PCP.

Part III explores alternative complexity measures, namely unitary query complexity and sample complexity. In Chapter 8, we develop a general technique for proving lower bounds on unitary query complexity using reductions to unitary channel discrimination, which also apply in the presence of quantum advice or proofs. As a direct corollary of our technique, we show that there exist quantum oracles relative to which $\mathsf{QMA}(2) \not\supset \mathsf{SBQP}$ and $\mathsf{QMA/qpoly} \not\supset \mathsf{SBQP}$. In Chapter 9, we introduce and study a sample-to-sample task called complement sampling. We show that classically, this task requires an exponential number of classical samples, whilst a quantum computer can solve it using only a single quantum sample—resulting in the largest possible separation between classical and quantum sample complexity. We also argue that, under standard cryptographic assumptions, the task is efficiently verifiable, classically hard, and feasible on near-term quantum devices, providing a new path to "quantum resource advantage".

Titles in the ILLC Dissertation Series:

ILLC DS-2020-13: **Thom van Gessel**
Questions in Context

ILLC DS-2020-14: **Gianluca Grilletti**
Questions & Quantification: A study of first order inquisitive logic

ILLC DS-2020-15: **Tom Schoonen**
Tales of Similarity and Imagination. A modest epistemology of possibility

ILLC DS-2020-16: **Ilaria Canavotto**
Where Responsibility Takes You: Logics of Agency, Counterfactuals and Norms

ILLC DS-2020-17: **Francesca Zaffora Blando**
Patterns and Probabilities: A Study in Algorithmic Randomness and Computable Learning

ILLC DS-2021-01: **Yfke Dulek**
Delegated and Distributed Quantum Computation

ILLC DS-2021-02: **Elbert J. Booij**
The Things Before Us: On What it Is to Be an Object

ILLC DS-2021-03: **Seyyed Hadi Hashemi**
Modeling Users Interacting with Smart Devices

ILLC DS-2021-04: **Sophie Arnoult**
Adjunction in Hierarchical Phrase-Based Translation

ILLC DS-2021-05: **Cian Guilfoyle Chartier**
A Pragmatic Defense of Logical Pluralism

ILLC DS-2021-06: **Zoi Terzopoulou**
Collective Decisions with Incomplete Individual Opinions

ILLC DS-2021-07: **Anthia Solaki**
Logical Models for Bounded Reasoners

ILLC DS-2021-08: **Michael Sejr Schlichtkrull**
Incorporating Structure into Neural Models for Language Processing

ILLC DS-2021-09: **Taichi Uemura**
Abstract and Concrete Type Theories

ILLC DS-2021-10: **Levin Hornischer**
Dynamical Systems via Domains: Toward a Unified Foundation of Symbolic and Non-symbolic Computation

ILLC DS-2021-11: **Sirin Botan**
Strategyproof Social Choice for Restricted Domains

ILLC DS-2021-12: **Michael Cohen**
Dynamic Introspection

ILLC DS-2021-13: **Dazhu Li**
Formal Threads in the Social Fabric: Studies in the Logical Dynamics of Multi-Agent Interaction

ILLC DS-2021-14: **Álvaro Piedrafita**
On Span Programs and Quantum Algorithms

ILLC DS-2022-01: **Anna Bellomo**
Sums, Numbers and Infinity: Collections in Bolzano's Mathematics and Philosophy

ILLC DS-2022-02: **Jan Czajkowski**
Post-Quantum Security of Hash Functions

ILLC DS-2022-03: **Sonia Ramotowska**
Quantifying quantifier representations: Experimental studies, computational modeling, and individual differences

ILLC DS-2022-04: **Ruben Brokkelkamp**
How Close Does It Get?: From Near-Optimal Network Algorithms to Suboptimal Equilibrium Outcomes

ILLC DS-2022-05: **Lwenn Bussière-Carae**
No means No! Speech Acts in Conflict

ILLC DS-2022-06: **Emma Mojet**
Observing Disciplines: Data Practices In and Between Disciplines in the 19th and Early 20th Centuries

ILLC DS-2022-07: **Freek Gerrit Witteveen**
Quantum information theory and many-body physics

ILLC DS-2023-01: **Subhasree Patro**
Quantum Fine-Grained Complexity

ILLC DS-2023-02: **Arjan Cornelissen**
Quantum multivariate estimation and span program algorithms

ILLC DS-2023-03: **Robert Paßmann**
Logical Structure of Constructive Set Theories

ILLC DS-2023-04: **Samira Abnar**
Inductive Biases for Learning Natural Language

ILLC DS-2023-05: **Dean McHugh**
Causation and Modality: Models and Meanings

ILLC DS-2023-06: **Jialiang Yan**
Monotonicity in Intensional Contexts: Weakening and: Pragmatic Effects under Modals and Attitudes

ILLC DS-2023-07: **Yiyan Wang**
Collective Agency: From Philosophical and Logical Perspectives

ILLC DS-2023-08: **Lei Li**
Games, Boards and Play: A Logical Perspective

ILLC DS-2023-09: **Simon Rey**
Variations on Participatory Budgeting

ILLC DS-2023-10: **Mario Giulianelli**
Neural Models of Language Use: Studies of Language Comprehension and Production in Context

ILLC DS-2023-11: **Guillermo Menéndez Turata**
Cyclic Proof Systems for Modal Fixpoint Logics

ILLC DS-2023-12: **Ned J.H. Wontner**
Views From a Peak: Generalisations and Descriptive Set Theory

ILLC DS-2024-01: **Jan Rooduijn**
Fragments and Frame Classes: Towards a Uniform Proof Theory for Modal Fixed Point Logics

ILLC DS-2024-02: **Bas Cornelissen**
Measuring musics: Notes on modes, motifs, and melodies

ILLC DS-2024-03: **Nicola De Cao**
Entity Centric Neural Models for Natural Language Processing

ILLC DS-2024-04: **Ece Takmaz**
Visual and Linguistic Processes in Deep Neural Networks: A Cognitive Perspective

ILLC DS-2024-05: **Fatemeh Seifan**
Coalgebraic fixpoint logic Expressivity and completeness result

ILLC DS-2024-06: **Jana Sotáková**
Isogenies and Cryptography

ILLC DS-2024-07: **Marco Degano**
Indefinites and their values

ILLC DS-2024-08: **Philip Verduyn Lunel**
Quantum Position Verification: Loss-tolerant Protocols and Fundamental Limits

ILLC DS-2024-09: **Rene Allerstorfer**
Position-based Quantum Cryptography: From Theory towards Practice

ILLC DS-2024-10: **Willem Feijen**
Fast, Right, or Best? Algorithms for Practical Optimization Problems

ILLC DS-2024-11: **Daira Pinto Prieto**
Combining Uncertain Evidence: Logic and Complexity

ILLC DS-2024-12: **Yanlin Chen**
On Quantum Algorithms and Limitations for Convex Optimization and Lattice Problems

ILLC DS-2024-13: **Jaap Jumelet**
Finding Structure in Language Models

ILLC DS-2025-01: **Julian Chingoma**
On Proportionality in Complex Domains

ILLC DS-2025-02: **Dmitry Grinko**
Mixed Schur-Weyl duality in quantum information

ILLC DS-2025-03: **Rochelle Choenni**
Multilinguality and Multiculturalism: Towards more Effective and Inclusive Neural Language Models

ILLC DS-2025-04: **Aleksi Anttila**
Not Nothing: Nonemptiness in Team Semantics

ILLC DS-2025-05: **Niels M. P. Neumann**
Adaptive Quantum Computers: decoding and state preparation

ILLC DS-2025-06: **Alina Leidinger**
Towards Language Models that benefit us all: Studies on stereotypes, robustness, and values

ILLC DS-2025-07: **Zhi Zhang**
Advancing Vision and Language Models through Commonsense Knowledge, Efficient Adaptation and Transparency