Efficiently Learning, Testing, and Simulating Quantum Many-Body Systems

by

Mehdi Soleimanifar

B.Sc., Sharif University of Technology (2016)

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Physics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2022

Author .....
Department of Physics
May 13, 2022

Certified by .....
Aram W. Harrow
Associate Professor of Physics
Thesis Supervisor

Accepted by .....
Deepto Chakrabarty
Associate Department Head of Physics

Efficiently Learning, Testing, and Simulating Quantum Many-Body Systems

by

Mehdi Soleimanifar

Submitted to the Department of Physics
on May 13, 2022, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Physics

Abstract

This thesis focuses on quantum information and quantum computing, and their applications in studying quantum many-body systems. A remarkable interplay between computer science and quantum physics in the past few decades has revealed that a precise control and manipulation of interacting quantum systems enables us to process information and perform computations that go beyond the reach of conventional digital computers. This novel form of information processing has also resulted in a conceptually new toolkit for tackling fundamental questions about the physics of quantum many-body systems. This thesis studies new features of interacting quantum systems through the lens of computational complexity and information theory. We will see how using these new features in turn allows us to develop efficient classical and quantum algorithms for learning, testing, and simulating quantum many-body systems. Below are the main results of this thesis:

1. We develop an algorithm for reliably testing the amount of entanglement in a pure many-body quantum state. This algorithm tests whether a quantum state is a matrix product state of certain bond dimension in the property testing model. We provide both upper and lower bounds on the number of identical copies of the quantum state required by this algorithm.

2. We prove that a quantum information quantity, known as the entanglement spread, satisfies an area law in the ground state of any gapped local Hamiltonian with an arbitrary geometry. This new feature of ground-state entanglement is obtained using a connection to the seemingly different problem of finding the communication complexity of testing bipartite states.

3. We devise an algorithm for learning the local Hamiltonian that governs the interactions in a quantum many-body system. This algorithm uses the results of local measurements on the thermal state of the system, and provably only requires a number of samples that scales polynomially with the number of particles.

4. A quasi-polynomial time algorithm is developed that estimates the quantum partition function at temperatures above the phase transition point. We also study different characterizations of the thermal phase transition by connecting the exponential decay of correlations to the analyticity of the free energy in the high-temperature phase.

5. We rigorously bound the improvement that low-depth quantum circuits can provide over methods based on product states in estimating the ground-state energy of local Hamiltonians.

Thesis Supervisor: Aram W. Harrow
Title: Associate Professor of Physics

Acknowledgments

I am grateful to my advisor, Aram Harrow, whose keen intuition about complex problems and vast technical wisdom have guided me in my research, and whose friendliness and modesty have been a source of inspiration in my day-to-day life. I also thank Aram for encouraging me to work on topics of my own interest and giving me independence to form fruitful collaborations.

I am indebted to John Wright for years of close friendship and his continuous mentorship in my graduate studies. John's clarity of thought, hands-on guidance, innovative perspective, and his lucid writing and presentation styles have taught me many attributes of a great researcher. Outside research, spending time with John has been always refreshing and fun, especially during the gloomy days of the pandemic.

I have also been fortunate to get to know Anurag Anshu during my PhD. Anurag is a brilliant researcher, an incredible collaborator, and a great friend. He has been available whenever I needed guidance and his kindness and generosity have greatly inspired me.

This thesis is the result of many wonderful collaborations. I am thankful to my coauthors Anurag Anshu, Srinivasan Arunachalam, David Gosset, Aram Harrow, Tomotaka Kuwahara, Saeed Mehraban, Karen Morenz Korol, and John Wright for the great deal that I learned from them. I especially thank David Gosset who has not only been an amazing source of insight in our joint project, but also generously helped me during my postdoctoral applications. I also thank the members of my thesis committee, Soonwon Choi, Anand Natarajan, and Senthil Todadri, for their time and helpful suggestions.

I thank Thomas Vidick for arranging my visit at the Simons Institute at Berkeley where I had many insightful discussions in particular with Umesh Vazirani, Zeph Landau, and Elizabeth Crosson. I am also grateful to Fernando Brandão for hosting me during a visit at Caltech where I had wonderful interactions with the IQIM group members.

With their unyielding efforts, the administrative staff in the physics department and the CTP have made my experience at MIT smooth and enjoyable. For this I especially thank Catherine Modica, Scott Morley, Charles Suggs, and Syndey Miller.

Going further back, I have been lucky to have great mentors in my undergraduate studies in physics and electrical engineering at Sharif University. In physics, I learned most of what I know about quantum physics and quantum information from Vahid Karimipour with whom I also had my first research project in quantum computing. Vahid's lectures gave me an intuitive grasp of physics and his exceptional presentation skills, in writing and talks, taught me how to communicate and share my scientific findings. In electrical engineering, I worked with Masoumeh Nasiri-Kenari and Amin Gohari on molecular communications and information theory. From them I learned how to effectively collaborate within a research group and saw first hand the traits of successful scientific leadership.

My best achievement in graduate school has been meeting the love of my life, Carina Belvin. I am grateful to her for all the adventures and lovely memories we have had together in the past couple of years, and look forward to all the exciting things that we are going to do in the future. I am also thankful to Carina's family for their unwavering support and making me feel welcome in my years away from home.

The pursuit of my scientific interests would not have been possible without early encouragement from my parents and their continuous support throughout my life. They have sacrificed in their life more than I can ever repay so that my sister and I can follow our dreams. I dedicate this thesis to my family.

Contents

List of Figures

Chapter 1

Introduction

1.1 Experiment versus simulation and what goes wrong

Much of our modern understanding of the physics of microscopic phenomena is gained by studying quantum many-body systems: a collection of interacting quantum particles which often exhibit collective phenomena not present when only considering individual subsystems. In a conventional experiment that explores the physics of these systems, a sample in the form of a crystal, thin film, or polymer is grown. These compounds are then probed with various techniques such as spectroscopy and microscopy techniques, each specialized in revealing certain aspects of the material such as its magnetic, electric, and optical properties. Although these have resulted in vast achievements, such experiments are also hampered by significant limitations: It is quite challenging to synthesize quantum materials that host a particular form of interactions or exhibit a specific phase of matter, and at the same time, be able to prob them with available experimental techniques which demand specific traits in their input samples to accurately operate. Even when successfully implemented, such experimental setups are only capable of revealing limited features of the systems under study, leaving many of their aspects unexplored.

Some of these issues can be bypassed if instead of undertaking such experiments, we study quantum many-body systems by simulating them on available digital computers. Such simulations offer a versatile tool: Once an algorithm for simulating a target system is successfully implemented, it is often not hard to adapt it for a variety of other systems with similar characteristics. It is also possible to read off various features of the simulated system at once, obtaining a comprehensive picture of the underlying physics. This approach, however, faces a substantial barrier of its own, namely, the steep cost of simulating quantum systems on conventional computers. Rooted deeply in the postulates of quantum mechanics, as we will see soon, such simulations of quantum many-body systems are believed to generally require computational resources that scale exponentially with the number of particles, a cost that quickly becomes prohibitive in practice. A host of heuristic and approximate numerical methods have been devised to get around this restrictive complexity, including density functional theory, Monte-Carlo sampling, dynamical mean-field theory, and tensor networks. Nevertheless, accurately simulating many quantum systems of interest such as those exhibiting the infamous sign-problem or involving highly correlated segments, as well as dynamical aspects of such systems, still remains out of reach.

1.2 A new fix: quantum computation

Some of the shortcomings of the current experimental and computational methods in studying interacting quantum systems could potentially be overcome via a distinctive paradigm known as quantum computing. This approach relies on the premise that it is possible to engineer scalable multi-partite quantum systems (often an assembly of two-level systems known as quantum bits or qubits) that admit a high level of control, remain coherent for long times, and can be readily initialized, evolved according to a desired set of interactions, and measured. Once practically realized, such a highly adaptive quantum device can be programmed to perform a new model of computation and information processing implied by the laws of quantum mechanics. Achieving this holds great promises: We would have access to sophisticated measurement probes that allow us to gain detailed information about a diverse set of properties of interacting quantum systems at the resolution of individual constituent particles. The study of collective dynamics and non-equilibrium phenomena, thermalization, and static properties at low and finite temperatures of a wider range of quantum many-body systems than currently possible could become feasible.

Beyond providing an adaptable tool for simulating quantum physics, the quest for building quantum computers has been part of a broader investigation of the connection between physics, in particular quantum mechanics, and the nature of information and computation. The resulting theoretical framework provides a fresh perspective on many old problems and has led to the formulation of new questions that seem difficult to come by with the traditional paradigms.

1.3 Challenges to address

To unlock the potential of quantum information science in studying quantum many-body physics, certain challenges are yet to be addressed. These challenges, as we will see, cut across different disciplines and to be successfully met, require adapting tools from computer science, information theory, and mathematical physics. This thesis focuses on the theoretical questions that arise at this interface, including:

1. **Power of classical and quantum computers.** What are the ultimate limits of classical and quantum computers in simulating quantum many-body systems? Under what conditions can quantum systems be simulated with modest computational resources and when is their simulation computationally intractable? Can quantum computers achieve a provable computational advantage in simulating physically-relevant interacting quantum systems compared to classical computers?

2. **Statistical analysis of quantum data.** In a future application of quantum computers, or in a more limited form specialized quantum simulators, the state of some quantum many-body system which may be in or out of equilibrium at low or finite temperatures could be prepared. Given such many-body states, how do we actually probe a specific property of interest or test the validity of a particular theoretical model? What is the best measurement to perform? What is the optimal algorithm for analyzing the measurement data obtained in such experiments?

3. **Many-body physics through the lens of information processing.** What new constraints can be imposed on the behavior of quantum many-body systems using a computational

and information-theoretic perspective? How can the well-developed theories of randomness and correlations in these fields be lifted to study the correlations that can arise from quantum interactions and the inherent randomness in quantum mechanics? Can these tools lead to new characterizations of different phases of matter and the transition between these phases?

1.4 Efficient versus intractable

Understanding the power of classical and quantum computers, dealing with the analysis of quantum data, and other challenges mentioned in the previous sections are questions which are computational in nature, and as such, we will apply theoretical frameworks developed in computer science to study them.

To build an abstract model of the computational resources needed to solve quantum many-body problems independent of the details of the platforms used, we consider the *asymptotic performance* of the underlying algorithm. That is, instead of focusing on a particular quantum system with a fixed set of interacting particles, we often study a general family of quantum many-body systems and track the worst-case performance of the algorithm as the number of particles in such families grow. There are different ways to characterize this performance, each being relevant in a particular setup. Here, we will be mostly interested in the time, sample, and communication complexities as overviewed in what follows.

Time complexity. The *running time* or *time complexity* is measured as the number of elementary operations executed by the algorithm. In quantum computing, these operations take the form of a series of local unitary gates applied on the input qubits as part of a *quantum circuit*. Algorithms with running times which scale as a polynomial function of the number of particles are called polynomial-time, or more colloquially, *efficient algorithms*. Problems that require running times that grow at least exponentially with the number of particles are considered *computationally intractable*. Besides the total number of local unitary gates in a quantum circuit—also known as the size of a quantum circuit—we are often also interested in the minimum depth needed to implement an algorithm on a quantum circuit. The *circuit depth* is defined as the maximum number of gates on any path that connects an input qubit to an output qubit, and corresponds to the parallel running time of an algorithm. As we will see, a limited depth is used to model the architecture of some of the quantum computers that may be available in the near future.

Sample complexity. When dealing with statistical problems involving quantum states, we assume multiple identical copies of the state are available and that a joint measurement can be performed on these copies. The number of such copies is called the *sample* (or *copy*) *complexity* which when scales polynomially with the number of particles, is another measure of efficiency that we consider. The measurements performed on these copies may simply be a series of single-qubit measurements possibly with a classical post-processing step, or more complicated multi-partite measurements which are implemented on a quantum computer. In either case, the time complexity of a statistical problem is considered as the running time of the algorithm involved.

Communication complexity. Finally, a measurement on a many-body quantum state may be performed when the quantum state is distributed among various parties, each having access to a subset of particles. Therefore, to collaboratively implement the measurement, the parties need to

engage in a communication protocol by exchanging qubits. The total number of exchanged messages is another proxy for the efficiency known as the *communication complexity*.

1.5 Pros and cons of complexity-theoretic modeling

Many computational problems that arise in the study of quantum many-body physics do not obviously admit a tractable algorithmic solution. Aiming for efficient algorithms therefore provides a guiding principle to uncover new structures present in these systems that allow us to avoid the apparent computational intractability. Such features turn out to be deeply connected to problems in physics and result in non-trivial insights. Examples of this are

1. The connection between the computational complexity of estimating quantum partition functions and thermal phase transitions examined in Chapter 5,

2. The relation between the sample complexity of learning quantum interactions and the properties of the free energy described in Chapter 4,

3. The implications of the communication complexity of testing quantum states on the structure of low-temperature many-body entanglement presented in Chapter 3.

Another advantage of evaluating the performance of algorithms in terms of time, sample, or communication complexity is that it makes the problem amenable to a rigorous mathematical analysis. For many problems, such a rigorous analysis remains an effective tool we have at our disposal since there are still no functional large-scale quantum computers that allow us to practically test ideas and benchmark some of the proposed algorithms. Even when practically feasible, having provable error bounds on the estimated quantities or limits on the resources that may be needed remains crucial.

There are also limitations to this complexity-theoretic approach. For one, the practical instances of problems that we want to efficiently solve might not be the worst-case instances. This means that in practice, the worst-complexity of a particular algorithm or a general intractability result about a problem are overly pessimistic and ignore the helpful features that are generic, but not universal. Some of the algorithms that we encounter in the next chapters, although efficient, are not optimized to be practical, and the abstractions used in their formal analysis also neglect some of the constraints that appear in different architectures for quantum computing.

1.6 What makes quantum mechanics hard?

Among the computational problems that are discussed in this thesis, some have close classical analogs which are well-studied topics in computer science and statistical physics. However, there are also purely quantum aspects to these questions, which render a direct extension of many of the known classical techniques to the quantum setting impossible or very challenging. In the following, we briefly review the challenges encountered in the quantum mechanical problems studied in the later chapters.

Entanglement and an exponentially-large Hilbert space. The state of a quantum system is described as a vector in a complex Hilbert space. The state space of multiple quantum systems is

the *tensor product* of such Hilbert spaces. This tensor product structure brings about entanglement, a rather counterintuitive feature of quantum mechanics. A bipartite quantum state is said to be entangled if it cannot be simply factored into two states, each corresponding to one of the constituent systems. Entanglement plays a multifaceted role in quantum many-body physics, ranging from providing a crucial resource for achieving a computational advantage with quantum computers to the characterization of new phases of matter with topological order. Obtaining a deeper understanding of entanglement is a goal we pursue in Chapter 3, where we reveal a universal constraint that many-body entanglement must satisfy in non-critical phases of interacting quantum systems, and in Chapter 2 where we develop a statistical theory of testing the amount of entanglement in quantum states.

One implication of the tensor product structure of the Hilbert space is that the number of bits needed to accurately describe an entangled many-body quantum system can grow exponentially with the number of particles, compared to the linear scaling for classical systems. To store and manipulate many-body quantum states, classical computers need to cope with this exponential complexity due to entanglement. This motivates studying the computational power of quantum states which admit low-complexity descriptions that can be dealt with using classical computers. This question is visited in Chapter 6 where we consider quantum states generated with low-depth quantum circuits, and Chapter 2 and Chapter 3 where we examine matrix product states (MPS) and projected entangled-pair states (PEPS), both examples of tensor network states which are efficient representations of quantum states. Finally, another consequence of the exponentially-large dimension of the Hilbert space is that even in the absence of entanglement, some of the problems commonly encountered in quantum mechanics, such as finding the ground state or free energy of many-body systems, may become computational intractable. We study conditions under which this intractability can be avoided in Chapter 5 and Chapter 6.

Non-commuting interactions. Unlike classical interactions, in quantum many-body systems, the interactions are described in terms of operators that do not necessarily commute with each other. This is one of the biggest hurdles that we actively need to deal with in our analysis throughout this thesis. For example, we will see in Chapter 4 and Chapter 5 how non-commutative interactions cause certain classical constraints on the thermal correlations to be violated in the quantum case. To get around this non-commutativity, various tools have been developed which are explained in more detail in Section 1.7.

Destructive measurements. When learning or testing a property of quantum systems or finding a classical description of quantum states, we need to perform a measurement. One crux of quantum mechanics is that measurements can alter the state of the system and affect the outcome of future measurements on the same state. This plus the fact that we often deal with a probabilistic mixture of different quantum states involving additional sources of randomness—for instance due to thermal interactions—means that we generally need many identical copies of a quantum state for a reliable estimation of its properties. At the same time, one of our contradictory goals in statistical problems is to reduce the number of such copies, given by the sample complexity. The resolution may involve reusing some of the copies after they are measured, accepting some level of disturbance caused by the measurements. Balancing this trade-off has been a topic of study in the past and also appears as a major barrier in Chapter 2 and comes with a quick fix in Chapter 4.

1.7 Locality and what makes quantum mechanics easy

One of the main ways we cope with the challenges presented in Section 1.6 is by relying on the *locality of the interactions*. This allows us to either recover some of the structures found in the classical setting or characterize new constraints on quantum systems that can be exploited in our analysis.

The idea behind the locality assumption is that the set of interactions in quantum many-body systems of interest can often be accurately modeled such that (a) each interaction only involves a few particles, (b) each particle is only included in a few such interactions, and (c) only particles within a close spatial neighborhood interact with each other. Not all of these three conditions (a)-(c) are strictly required for the results presented in this thesis. To distinguish between them, we refer to interactions that are only required to satisfy condition (a) *local*, those that satisfy (a) and (b) *bounded-degree local*, and those that meet (a),(b), and (c) *geometrically local*.

Most interactions that we naturally encounter in physics are geometrically local, which are also the most-structured among local interactions (a)-(c) exhibiting features that we crucially use in future chapters. Why do we also consider the more relaxed notions of bounded-degree or, more generally, local interactions? One reason is that these interactions can be engineered in certain experimental setups or simulated with quantum computers. Quantum many-body systems with local and bounded-degree interactions also appear in modelling quantum chaos, investigating holographic duality in connections to quantum gravity, studying disordered dynamics, and developing quantum error correcting codes allowing coherent quantum computation.

Having introduced different notions of locality in interacting quantum systems, in what follows we discuss some of its implications. Our objects of study are local Hamiltonians $H = \sum_{j=1}^{m} h_j$ where each Hermitian operator h_j describes one of the local interactions among particles. Each local Hamiltonian has a series of eigenvalues E_j and eigenvectors $|E_j\rangle$. We will be interested in both dynamic and static (or equilibrium) properties of interacting quantum systems. When in thermal equilibrium at an inverse temperature β, the system is in the energy eigenstate $|E_j\rangle$ with probability proportional to $e^{-\beta E_j}$. This is more concisely given by the Gibbs (thermal) state $e^{-\beta H}/Z_\beta$ where $Z_\beta = \mathrm{tr}(e^{-\beta H})$ is the partition function of the system. The low-temperature properties are described by the ground state which corresponds to the lowest energy eigenstate $|E_0\rangle$ — here and in some later introductory parts, we neglect the possible degeneracy of the lowest energy state. The dynamics of interacting quantum systems are given by the time evolution operator e^{-itH}. We will now see how these states and the time evolution of quantum systems are constrained by the locality of the Hamiltonian.

Lieb-Robinson bound. Consider a collection of qubits (or finite-dimensional quantum particles) that are initially in a product (unentangled) state. That is, there is no correlation between any pairs of qubits. Suppose the state of one of the qubits (called A) is changed and the qubits are then left to evolve in time according to the geometrically-local interactions among them. It turns out that there is a finite speed at which the effect of changing the state of qubit A is propagated to the other qubits. This means at any time, there is a region around qubit A—called its effective lightcone—such that measuring qubits outside this region, up to exponentially small corrections, does not reveal any information about the initial change on qubit A. This behavior, known as the Lieb-Robinson bound, has been first formalized in [LR72] and later improved in various works starting with [Has04, NS09]. Many of the tools we use in the next chapters (especially our results in Chapter 4) rely on the

Lieb-Robinson bound or its variants.

Limited leakage to high-energy states. Suppose a quantum many-body system is in a superposition of energy eigenstates with energy at most E. If we apply a transformation on the qubits inside a region A, we can potentially create a superposition of different energy eigenstates, some with energy greater than E. It is shown in [AKL16] that for bounded-degree interactions where each qubit is involved in a constant number of interactions, up to exponentially small corrections, all the eignestates that appear in the new state have energies bounded by $E + O(|A|)$. This statement trivially holds in the classical setting where after flipping the state of $|A|$ classical spins, no superposition is created and the new state is an eigenstate with energy at most $E + O(|A|)$. We will see a direct application of this bound in Chapter 4 where we study the variance of quasi-local observables.

Cluster expansions. In thermal equilibrium and at an inverse temperature β, the partition function of a many-body system with a local Hamiltonian H is given by $Z_\beta = \mathrm{tr}(e^{-\beta H})$. When studying the finite-temperature properties of interacting systems or evaluating many thermal properties of interest such as the free energy, we need to deal with the partition function $Z_\beta(H)$ and its logarithm $\log Z_\beta(H)$. Being a complicated mixture of different eigenstates of the Hamiltonian, in general, it is quite challenging to theoretically investigate or computationally evaluate the features of partition functions. However, for geometrically-local Hamiltonians at high enough temperatures (which often means when $\beta \leq \beta_0$ for some constant inverse temperature β_0 that depends on the geometric details of the Hamiltonian), a powerful technique, known as the cluster expansion, can be used to analyze quantum and classical partition functions. In cluster expansions, the partition function $Z_\beta(H)$ is expanded into the trace of terms involving a sum of the product of local terms in the Hamiltonian. This, for instance, lets us separate the contribution of the interaction terms acting on a specific particle in the overall free energy (or log-partition function). Using cluster expansions, we can rigorously show that at high temperatures, the correlations between distant observables in the Gibbs state decay exponentially or that the free energy of the system is analytic. In Chapter 5, we introduce and apply cluster expansions in connection to efficiently estimating quantum partition functions.

Quantum belief propagation When studying the Gibbs state of interacting quantum systems, we often need to know how a local change in the Hamiltonian affects the Gibbs state. In classical interactions, if we consider a modified Hamiltonian $H + V$, the (unnormalized) Gibbs state $e^{-\beta H}$ simply changes to $e^{-\beta H}e^{-\beta V}$ or equivalently $e^{-\beta V/2}e^{-\beta H}e^{-\beta V/2}$. It is shown in [Has07b, Kim12] that for geometrically-local quantum interactions, a similar result holds if V is additionally local. That is, $e^{-\beta(H+V)} = \eta e^{-\beta H}\eta^\dagger$ where η is a quasi-local operator, known as the quantum belief propagation operator, which aside from exponentially decaying tails, acts on the same region as V and has a similar norm as the operator $e^{-\beta V/2}$. We will see various consequences of this in Chapter 4 and Chapter 5.

Low-degree ground-state projectors. At low temperatures, the state of a quantum many-body system is given by the ground state which is the lowest energy eigenstate $|E_0\rangle$ of the Hamiltonian. A Hamiltonian H is called gapped if there is a constant energy difference $\gamma > 0$ between the ground state and the higher energy states. The ground state of a gapped local Hamiltonian inherits many

local features from its Hamiltonian. To roughly see why, suppose the ground-state energy $E_0 = 0$ and denote the maximum energy eigenvalue by $\|H\|$. Consider the operator $\mathbb{1} - H/\|H\|$. When acted on energy eigenstates $|E_j\rangle$, this operator leaves the ground state $|E_0\rangle$ unchanged, but shrinks the remaining eigensates $|E_j\rangle$ with $j \geq 1$ by multiplying them with a factor $\leq 1 - \gamma/\|H\|$, which can be amplified by further applications of $\mathbb{1} - H/\|H\|$ on the state. This implies that for a sufficiently large k, the operator $p(H) = (\mathbb{1} - H/\|H\|)^k$, which is a degree-$k$ polynomial in terms of the local terms in the Hamiltonian, is an approximate projector onto the ground state. While this provides a simple polynomial approximation of the ground state, there are more sophisticated construction of $p(H)$ that achieve improved properties with a substantially lower degree. Such low-degree approximate ground-state projectors (AGSP) imply various local features in the ground state of gapped systems. When the local Hamiltonian H is on a 1D chain, AGSPs can be used to show that the entanglement entropy of a region A in the ground state $S(A)$ is upper bounded by the size of the boundary $|\partial A|$ rather than $|A|$ as in generic many-body quantum systems, and there are efficient classical algorithms for finding the ground state in terms of matrix product states (MPS). These features are the subject of our study in next chapters with a brief overview given in the next section.

1.8 Synopsis of this thesis

1.8.1 Testing and learning quantum many-body systems

The destructive nature of measurements in quantum mechanics along with the exponential complexity of quantum systems and the presence of multipartite entanglement often complicates the statistical analysis of the data that arise in quantum physics. Such data, for instance, may be obtained when probing the intricate microscopic properties of quantum systems or when measuring the output of quantum computers. Can we devise algorithms for learning and testing the properties of quantum systems that consume modest computational resources and get around the apparent hurdles posed by quantum mechanics? These efficient algorithms will be crucial for validating the performance of upcoming quantum devices and examining the outcome of future quantum experiments that access the high-complexity regime of quantum mechanics. This thesis brings tools from statistical learning theory and combines them with new structural results about quantum many-body systems to develop such provably efficient algorithms.

A statistical theory of testing entanglement. Entanglement is a crucial resource for quantum computing and quantum information processing, and its study has shed light on various phases of matter and quantum many-body phenomena. Starting with the widely-known Bell test, different schemes have been considered for testing the amount of entanglement in a quantum state with applications ranging from benchmarking quantum devices to quantum cryptography. Here, we study this question from the statistical viewpoint. We devise an algorithm that when provided with multiple copies of a quantum state, correctly decides with high probability whether the amount of entanglement in that state is bounded by a certain value or not. Our main objective in Chapter 2 is to determine:

What is the fewest number of copies of a many-body quantum state needed for reliably testing the amount of entanglement it possesses?

In order to measure the amount of entanglement in a quantum state, we use *matrix product states*. A quantum state comprised of n qudits is said to be a matrix product state of bond dimension r

if the reduced density matrix $\psi_{1,\dots,k}$ has rank r for each $k \in \{1,\dots,n\}$. We denote the family of such states with $\mathsf{MPS}(r)$. The bond dimension r limits the amount of entanglement in the state: Matrix product states with $r = 1$ correspond to the set of product states with no entanglement, and as r grows, $\mathsf{MPS}(r)$ includes more quantum states. Matrix product states of small bond dimension are known to well approximate some physically-relevant quantum states such as those related to interacting quantum systems with a 1D geometry.

Using matrix product states as a way to quantify many-body entanglement, our entanglement tester checks if a quantum state is in $\mathsf{MPS}(r)$ or if it is far from states in $\mathsf{MPS}(r)$. In the case of $r = 1$, i.e. testing if a state is a product state, our results give a simple and improved analysis of "the product test," previously studied by [HM13]. Besides its relevance in the statistical analysis of quantum data, the product testing is known to be related to the computational hardness of various problems both in and out of quantum information theory related to entanglement and tensor optimization. For the case of $r \geq 2$, we give an efficient algorithm for testing whether $|\psi\rangle$ is an MPS of bond dimension r using $m = O(nr^2)$ copies, independent of the dimensions of the qudits, and we show that $\Omega(n^{1/2})$ copies are necessary for this task. This chapter is based on:

[SW22] Mehdi Soleimanifar and John Wright. Testing matrix product states. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1679–1701, 2022.

Testing bipartite states and structure of ground states entanglement. As mentioned in Section 1.7, the ground state of gapped local Hamiltonians inherits the locality of the Hamiltonian. A long-known manifestation of locality in these states is the decay of correlations; That is, correlations between two subsystems decay exponentially in their distance which is measured according to the geometry induced by the Hamiltonian. Another intriguing yet much more challenging problem has been to uncover local features that bipartite entanglement exhibits in the ground state of such systems. This problem has been the focus of many previous studies resulting in a number of prominent conjectures, in particular, the entanglement entropy "area law" which states that the entropy of a region grows proportional to the size of its boundary. Previous works have also investigated the information available in the ground-state entanglement beyond what can be inferred from entanglement entropy, predicting that the entanglement spectrum can be described in terms of the spectrum of a boundary (modular) Hamiltonian in one fewer dimension. These conjectures have yet to be rigorously established for general geometries. For instance, the entanglement entropy area law is proven for 1-D systems, for 2-D systems subject to an additional assumption (being "frustration-free"), and is known to fail for general geometries. In Chapter 3 we ask:

What are the features of the ground-state entanglement in gapped local Hamiltonians that provably hold for arbitrary geometries?

Our work addresses this question by connecting it to a seemingly unrelated topic: the communication complexity of testing bipartite entangled states. To explain this connection, suppose Alice and Bob want to test if a bipartite state shared between them is a target state $|\psi\rangle_{AB}$ or not. The communication complexity of this protocol (i.e. the minimum number of qubits the parties need to exchange) can be shown to be related to a feature of the target state $|\psi\rangle_{AB}$ known as "entanglement spread." This quantum information quantity measures the difference between the Rényi entanglement entropies and in a sense is a measure of how far a state is from both maximally entangled and product states. We then develop a new communication protocol for the case that the target state $|\psi\rangle_{AB}$ is the unique ground state of a gapped local Hamiltonian. By analyzing the communication

complexity of this protocol, we establish an area law for entanglement spread. This result is not restricted to low-dimensional lattices and holds for any gapped local Hamiltonian with a general geometry. For the special case of lattices, we improve our bound on the entanglement spread to a sub-area law scaling. We also show how this feature has interesting algorithmic implications and connects to other conjectures regarding the locality of modular Hamiltonians. This chapter is based on:

[AHS20] Anurag Anshu, Aram W Harrow, and Mehdi Soleimanifar. From communication complexity to an entanglement spread area law in the ground state of gapped local Hamiltonians. *To appear in Nature Physics. Preprint available at arXiv:2004.15009*, 2020.

Learning quantum interactions. The interactions among particles in a quantum system are described by the Hamiltonian. Learning the Hamiltonian from measurement data is an important task in experiments that probe quantum many-body physics and the verification of quantum technologies. The main question we study in Chapter 4 is:

How can we devise a quantum Hamiltonian learning algorithm with efficient performance guarantees?

The classical analog of this problem, known as learning Boltzmann machines, is a well-studied problem in machine learning which admits efficient algorithms and corresponds to learning the Hamiltonian of the classical Ising model and its generalizations. For the case of quantum interactions, although there have been various proposals for learning quantum Hamiltonians, their performance either has not been rigorously analyzed or they require a number of measurements that scales exponentially with the number of particles. This lack of efficient algorithms for quantum Hamiltonian learning is not a mere technicality and points to a fundamental distinction between quantum and classical interactions: At finite temperatures, quantum systems can violate the Markov property, a feature that allows for machine learning techniques to be successfully applied to inferring classical systems such as Ising models. We introduce several new ideas to obtain an unconditional result that avoids relying on the Markov property of quantum systems. The key to our findings is proving that the absolute value of the finite-temperature free energy of quantum many-body systems is strongly convex with respect to the interaction coefficients. Using this, we prove that only a polynomial number of simple local measurements on the thermal state of a quantum system are necessary and sufficient for accurately learning its Hamiltonian. This chapter is based on:

[AAKS21] Anurag Anshu, Srinivasan Arunachalam, Tomotaka Kuwahara, and Mehdi Soleimanifar. *Sample-efficient learning of interacting quantum systems. Nature Physics, 17(8):931–935*, 2021. Also in *Proceedings of IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS), pages 685–691,* 2020.

1.8.2 Simulating quantum many-body systems

The exponential complexity of quantum many-body systems limits our ability to simulate their properties from first principles of quantum mechanics. One way to sidestep this difficulty is to intelligently exploit the special features that certain families of quantum systems exhibit. An integral part of this thesis is finding such properties. Chapter 5 studies the roles of entanglement and thermal phase transitions in the computational hardness of simulating quantum systems. In Chapter 6, we

investigate the extent to which quantum computers can be useful for simulating the low-temperature properties of quantum many-body systems.

Thermal phase transition and computational complexity. As the temperature of quantum many-body systems is lowered, we see a transition in the computational hardness of estimating their thermal properties, changing from trivial at infinite temperatures to computationally intractable at very low temperatures. But we also know that lowering the temperature causes the phase of the system to undergo a sharp transition at a critical temperature. There are two seemingly different characterizations for the high-temperature phase residing above this critical point. One is that the correlations between two observables decay exponentially with their distance. The other is the analyticity of the free energy of the system, equivalently given by the absence of complex zeros of the partition function near real temperatures. In Chapter 5 we study:

How are the different characterizations of the thermal phase transition related to the computational complexity of estimating the finite-temperature properties?

We show that the thermal transition in the phase of a quantum many-body system can be accompanied by a transition in the computational hardness of approximating its properties. More precisely, we show that at temperatures above the thermal phase transition point, there is a (quasi)polynomial-time algorithm for estimating quantum partition functions. The same problem is known to be computationally intractable (NP-hard) in the worst case below the phase transition point. We establish this result using the characterization of thermal phase transitions in terms of the complex zeros of the partition function. This allows us to estimate the free energy using convergent Taylor expansions which can be efficiently computed, resulting in our algorithm. Building on [DS85], we also study the connection between the complex zeros of quantum partition functions and the exponential decay of correlations in quantum systems. We prove that at temperatures above the phase transition point, where the complex zeros of the partition function are far from the real temperature axis, the correlations between two observables with sufficiently large distance decay exponentially. This chapter is based on:

[HMS20] Aram W. Harrow, Saeed Mehraban, and Mehdi Soleimanifar. Classical algorithms, correlation decay, and complex zeros of partition functions of quantum many-body systems. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC), pages 378–386*, 2020.

Power of near-term quantum computers. While it is known that quantum computers can efficiently simulate the dynamics of quantum many-body systems, the extent to which they are useful for simulating the equilibrium (or static) properties, especially at very low temperatures, is not well understood. Many of the known classical techniques for this purpose—most famously, those based on mean-field approximations—only produce product (unentangled) states. Given that the ground state of quantum systems may be highly entangled, it is natural to ask what advantage can quantum computers provide for estimating the low-temperature properties of quantum systems. Due to the fragile nature of quantum computers, their applications in the near future will be restricted to algorithms that can be run on small quantum computers with low-depth (shallow) circuits. Hence, we ask:

How much improvement can low-depth quantum circuits provide over methods based on product states in quantum simulations?

We prove that shallow quantum circuits can providing an extensive improvement in estimating the ground state energy upon product state approximations. More formally, we consider bounded-degree local Hamiltonians H involving m local terms where each qubit participates in at most d interactions. Given any product state $|v\rangle$ with mean energy $e_0 = \langle v|H|v\rangle$ and energy variance $\mathrm{Var}_v(H) = \langle v|(H-e_0)^2|v\rangle$, we design a low-depth quantum circuit that when acted on $|v\rangle$ improves its energy by an amount at least proportional to $\mathrm{Var}_v(H)^2/(d^2m)$. For typical product states $|v\rangle$ this translates into an extensive energy improvement proportional to n, the number of qubits. This chapter is based on:

[AGMKS21] Anurag Anshu, David Gosset, Karen J. Morenz Korol, and Mehdi Soleimanifar. Improved approximation algorithms for bounded-degree local Hamiltonians. *Phys. Rev. Lett., 127:250502, Dec 2021*

Funding Acknowledgment: The upcoming chapters of this thesis were supported by NSF Grant CCF-1729369.

Chapter 2

Testing matrix product states

Chapter summary: Matrix product states (MPS) are a class of physically-relevant quantum states which arise in the study of quantum many-body systems. A quantum state $|\psi_{1,\ldots,n}\rangle \in \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n}$ comprised of n qudits is said to be an MPS of bond dimension r if the reduced density matrix $\psi_{1,\ldots,k}$ has rank r for each $k \in \{1,\ldots,n\}$. When $r = 1$, this corresponds to the set of product states, i.e. states of the form $|\psi_1\rangle \otimes \cdots \otimes |\psi_n\rangle$, which possess no entanglement. For larger values of r, this yields a more expressive class of quantum states, which are allowed to possess limited amounts of entanglement.

Devising schemes for testing the amount of entanglement in quantum systems has played a crucial role in quantum computing and information theory. In this chapter, we study the problem of testing whether an unknown state $|\psi\rangle$ is an MPS in the property testing model. In this model, one is given m identical copies of $|\psi\rangle$, and the goal is to determine whether $|\psi\rangle$ is an MPS of bond dimension r or whether $|\psi\rangle$ is far from all such states. For the case of product states, we study the product test, a simple two-copy test previously analyzed by Harrow and Montanaro [HM13], and a key ingredient in their proof that $\mathsf{QMA}(2) = \mathsf{QMA}(k)$ for $k \geq 2$. We give a new and simpler analysis of the product test which achieves an optimal bound for a wide range of parameters, answering open problems in [HM13] and [MdW13]. For the case of $r \geq 2$, we give an efficient algorithm for testing whether $|\psi\rangle$ is an MPS of bond dimension r using $m = O(nr^2)$ copies, independent of the dimensions of the qudits, and we show that $\Omega(n^{1/2})$ copies are necessary for this task. This lower bound shows that a dependence on the number of qudits n is necessary, in sharp contrast to the case of product states where a constant number of copies suffices. This chapter is based on:

[SW22] Mehdi Soleimanifar and John Wright. Testing matrix product states. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1679–1701, 2022.

2.1 Introduction

This chapter is about matrix product states (MPS).

Definition 1 (Matrix product states)**.** *A quantum state $|\psi\rangle \in \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n}$ consisting of n qudits is a* matrix product state with bond dimension r *if it can be written as*

$$|\psi_{1,\ldots,n}\rangle = \sum_{i_1 \in [d_1],\ldots,i_n \in [d_n]} \mathrm{tr}[A_{i_1}^{(1)} \cdots A_{i_n}^{(n)}] \cdot |i_1 \cdots i_n\rangle,$$

where each matrix $A_j^{(i)}$ is an $r \times r$ complex matrix, for $i \in [n]$ and $j \in [d_i]$. We write $\mathsf{MPS}_n(r)$ for the set of such states, or more simply $\mathsf{MPS}(r)$ when the dependency on n is clear from the context.

The parameter r controls the amount of entanglement $|\psi\rangle$ is allowed to possess, and as it increases, the set of MPS grows larger and more expressive. On one extreme, when $r = 1$ this corresponds to the set of product states, i.e. state of the form $|\psi_1\rangle \otimes \cdots \otimes |\psi_n\rangle$, which possess no entanglement between different qudits. On the other extreme, every state $|\psi\rangle$, even a highly entangled one, is an MPS of bond dimension $r = d_1 \cdots d_n$. Between these two extremes, MPS allow for nonzero though still limited entanglement, which grows with r. This can be seen more readily in the following alternative characterization of MPS, which states that $|\psi_{1,\ldots,n}\rangle$ is an MPS of bond dimension r if and only if $\psi_{1,\ldots,k}$ has rank r for each $1 \leq k \leq n$, where $\psi_{1,\ldots,k}$ is the reduced density matrix on the first k qudits. Here, we say that a Hermitian matrix has rank r if it has at most r nonzero eigenvalues. This implies, for example, that the entanglement entropy between the first k and the last $n - k$ qudits is always at most $\log(r)$, for each k. We will prefer this alternative characterization in this chapter.

MPS feature prominently in the study of quantum many-body physics, with a particular emphasis on one-dimensional quantum systems. In a typical one-dimensional quantum system, n qudits are arranged on a line, and their interactions are governed by a local Hamiltonian H which only contains local terms between neighboring qudits, i.e. terms of the form $H_{i,i+1}$. The *one-dimensional area law* of Hastings [Has07a], as well as further refinements in [ALV12, AKLV13, LVV15], implies that if H is a gapped Hamiltonian, then its ground state $|\psi_{1,\ldots,n}\rangle$, is well-approximated by an MPS of "small" bond dimension. One-dimensional quantum systems are an important class of physically-motivated systems, and this characterization in terms of MPS means they are tractable to analyze with computers. For example, [ALVV17] have developed rigorous algorithms for approximating the ground state of a one-dimensional gapped Hamiltonian. And [CPF+10] have suggested using MPS tomography to efficiently learn the state of a one-dimensional system using a small number of copies, motivated by the fact that an MPS only has $(d_1 + \cdots + d_n)r^2$ parameters to "learn", exponentially fewer than the $d_1 \cdots d_n$ parameters of a general quantum state. The classical tractability of matrix product states has also resulted in their widespread application as a computational method in the classical simulation of quantum circuits, both in one and higher dimensions. This includes the simulation of shallow quantum circuits [NLPD+19, BGM21, CC22], slightly entangled quantum circuits [Vid03], and noisy quantum circuits [ZSW20].

In this chapter, we study the problem of "testing" whether an unknown state $|\psi\rangle$ is an MPS. We will study this in the model of property testing. In this model, an algorithm is given access to multiple copies of $|\psi\rangle$ which it is allowed to measure; its goal is to determine if $|\psi\rangle$ is an MPS using as few copies as possible. This problem has been previously studied for the $r = 1$ case of product states by Harrow and Montanaro [HM13], and studying the case of general r was suggested as an open direction by Montanaro and de Wolf [MdW13]. To define this model, we begin by formally defining what it means for a state to be "far" from being an MPS.

Definition 2 (Distance to $\mathsf{MPS}(r)$)**.** *Given $n \geq 1$ and a state $|\psi\rangle \in \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n}$, the distance of $|\psi\rangle$ to the set $\mathsf{MPS}(r)$ is defined as*

$$\mathrm{Dist}_r(|\psi\rangle) = \min_{|\phi\rangle \in \mathsf{MPS}(r)} \mathrm{D}_{\mathrm{tr}}(\psi, \phi) = \min_{|\phi\rangle \in \mathsf{MPS}(r)} \sqrt{1 - |\langle\psi|\phi\rangle|^2},$$

where $\mathrm{D}_{\mathrm{tr}}(\cdot, \cdot)$ denotes the standard trace distance*, and ψ and ϕ denote the mixed states corresponding to $|\psi\rangle$ and $|\phi\rangle$, respectively. Sometimes we will prefer to work with the maximum squared overlap*

of $|\psi\rangle$ *with* $\mathsf{MPS}(r)$*, defined as*

$$\mathrm{Overlap}_r(|\psi\rangle) = \max_{|\phi\rangle \in \mathsf{MPS}(r)} |\langle\psi|\phi\rangle|^2.$$

When referring to the distance, we will typically use the variable name $\delta = \mathrm{Dist}_r(|\psi\rangle)$*, and when referring to the overlap, we will typically use* $\omega = \mathrm{Overlap}_r(|\psi\rangle)$ *or, alternatively,* $1 - \epsilon = \omega$*. Note that*

$$\delta = \sqrt{1-\omega} = \sqrt{\epsilon}.$$

Now we define the problem we consider, that of property testing MPS.

Definition 3 ($\mathsf{MPS}(r)$ tester)**.** *An algorithm* $\mathcal{A}$ *is a* property tester for $\mathsf{MPS}(r)$ using $m = m(n, r, \delta)$ copies *if, given* $\delta > 0$ *and* m *copies of* $|\psi\rangle \in \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n}$*, it acts as follows.*

- *(Completeness): If* $|\psi\rangle \in \mathsf{MPS}(r)$*, then*

  $$\Pr[\mathcal{A} \text{ accepts given } |\psi\rangle^{\otimes m}] \geq \tfrac{2}{3}.$$

  If instead it accepts with probability exactly 1 *in this case, we say that it has* perfect completeness.

- *(Soundness): If* $\mathrm{Dist}_r(|\psi\rangle) \geq \delta$*, then*

  $$\Pr[\mathcal{A} \text{ accepts given } |\psi\rangle^{\otimes m}] \leq \tfrac{1}{3}.$$

All property testers considered in this chapter have perfect completeness, whereas our lower bounds will apply to property testers even with imperfect completeness.

Previous works have considered testing a variety of properties of quantum states. Perhaps the most relevant is that of O'Donnell and Wright [OW15], which considered testing properties of a mixed state ρ's spectrum, such as testing whether its rank is at most r—we will revisit this later. Another relevant work is that of Harrow, Montanaro, and Lin [HLM17], which considers the problem of testing whether $|\psi_{1,\ldots,n}\rangle$ is a product state across *some* cut, meaning there exists an $S \subseteq \{1, \ldots, n\}$ such that $|\psi_{1,\ldots,n}\rangle = |\psi_S\rangle \otimes |\psi_{\overline{S}}\rangle$. If not, they say that $|\psi_{1,\ldots,n}\rangle$ possesses "genuine n-partite entanglement". (In contrast, in $r = 1$ case of product testing, we want to verify that $|\psi_{1,\ldots,n}\rangle$ is a product state across *every* cut S.) They give a tester for this problem which uses $m = O(n/\epsilon^2)$ copies of the state. For more on quantum property testing, see the survey of Montanaro and de Wolf [MdW13].

More broadly, testing and characterizing the entanglement of quantum systems has been an important theme running throughout quantum computation, even outside the model of property testing. This includes the study of nonlocal games, where the CHSH game [CHSH69] allows one to verify that two parties share an EPR state, with applications in delegation of quantum computation [Mah18, CGJV19, RUV13], device-independent quantum cryptography [VV19], and interactive proof systems [JNV+20]. Moreover, the communication complexity of two-party protocols for testing shared entangled states, including EPR states, has been used to reveal the properties of entanglement in ground states of local Hamiltonians [AHL+14, AHS20].

We emphasize that we are specifically considering property testing of *pure* states. In particular, we assume that the state the algorithm $\mathcal{A}$ is given m copies of is pure, not mixed. There are,

however, problems related to ours in the property testing of mixed states, although we do not cover these in this chapter. One example is the question of testing whether a *mixed* state ρ_{AB} on two d-dimensional subsystems is separable (i.e. not entangled), which is both fascinating and still very much open. The best known algorithm for this problem is the trivial one: simply use $O(d^4)$ copies to "learn" ρ_{AB} and classically compute whether it is entangled. On the other hand, the best known lower bound is $\Omega(d^2)$. Another example is the problem of testing whether ρ_{AB} is a tensor product, i.e. whether $\rho_{\mathrm{AB}} = \rho_A \otimes \rho_B$. For this problem, we *do* know the optimal bound: $\Theta(d^2)$ copies, given by the algorithm of [Yu21]. One convenience of pure states is that these two problems coincide for this case, since a pure state is a product state if and only if it is unentangled. For mixed states, this is not true.

While in this chapter, we focus primarily on MPS. We note that these states are a special example of the more general class of tensor network states. Devising learning and testing algorithms for these states is an interesting future direction to explore.

2.1.1 The product test

We begin with the simplest case of MPS testing, when the bond dimension $r = 1$, which corresponds to testing whether $|\psi\rangle \in \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n}$ is a product state. We study a simple two-copy property tester for this problem known as the *product test* which was introduced by Mintert, Kuś, and Buchleitner [MKB05] and later studied by Harrow and Montanaro [HM13]. The product test is itself built out of a simpler subroutine known as the *SWAP test* due to Buhrman, Cleve, Watrous, and de Wolf [BCWdW01], which measures the similarity between two qudit states $|a\rangle, |b\rangle \in \mathbb{C}^d$.

Definition 4 (The SWAP test)**.** *Given two qudit states $|a\rangle, |b\rangle \in \mathbb{C}^d$, the SWAP test applies the two-outcome projective measurement $\{\Pi_{\mathrm{SWAP}}, \mathbb{1} - \Pi_{\mathrm{SWAP}}\}$ to $|a\rangle \otimes |b\rangle$, where $\Pi_{\mathrm{SWAP}} = (\mathbb{1} + \mathsf{SWAP})/2$. Here, SWAP is the two-qudit swap operator, defined as*

$$\mathsf{SWAP}|i\rangle \otimes |j\rangle = |j\rangle \otimes |i\rangle$$

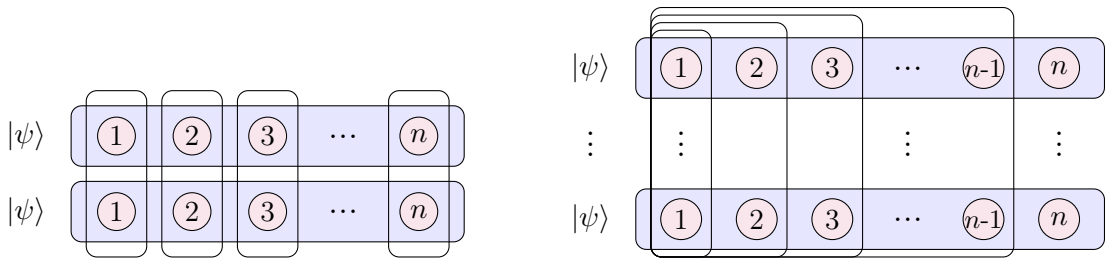
for all $i, j \in [d]$. The test accepts if it observes the first outcome, and it rejects otherwise.

It can be checked that the SWAP test succeeds with probability $\frac{1}{2} + \frac{1}{2}|\langle a|b\rangle|^2$. In particular, it succeeds with probability 1 if and only if, modulo a phase factor, $|a\rangle = |b\rangle$. Having defined the SWAP test, we can now define the product test.

Definition 5 (The product test)**.** *Given two copies of a state $|\psi\rangle \in \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n}$, the product test performs the SWAP test on the i-th qudit in each copy of $|\psi\rangle$, simultaneously over all $i \in [n]$, and accepts if they all accept. Equivalently, it performs the two-outcome projective measurement $\{\Pi_{\mathrm{Prod}}, \mathbb{1} - \Pi_{\mathrm{Prod}}\}$, where $\Pi_{\mathrm{Prod}} = \Pi_{\mathrm{SWAP}}^{\otimes n}$. and the i-th Π_{SWAP} applies to the i-th qudits in both copies of $|\psi\rangle$. We include an illustration of the product test in Figure 2-1a.*

In the case when $|\psi\rangle$ is a product state, i.e. $|\psi\rangle = |\psi_1\rangle \otimes \cdots \otimes |\psi_n\rangle$, the product test passes with probability 1, because for each $i \in [n]$ the i-th SWAP test is applied to $|\psi_i\rangle \otimes |\psi_i\rangle$, and so it always succeeds. This property of always accepting product states is known as *perfect completeness*. In fact, Harrow and Montanaro [HM13, Section 5] show that the product test is the optimal two-copy test for product states with perfect completeness, in the sense that any other two-copy test with perfect completeness will reject any non-product state $|\psi\rangle$ with at most the probability the product test rejects it.

We are interested in the maximum probability a state passes the product test, defined as follows.



(a) The product test performs a SWAP test on each of the n pairs of subsystems of the two copies of $|\psi\rangle$. Figure taken from [HM13].

(b) The MPS tester simultaneously performs the rank tester on each of the $n-1$ contiguous cuts across the multiple copies of $|\psi\rangle$.

Figure 2-1: The product tester and the MPS tester.

Definition 6. *Let $n \geq 1$ and $\omega \in [0,1]$. Given a state $|\psi\rangle \in \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n}$ we define $\mathrm{PT}_n(|\psi\rangle)$ to be the probability the product test succeeds on $|\psi\rangle$. In addition, we define $\mathrm{PT}_n(\omega)$ to be the supremum of $\mathrm{PT}_n(|\psi\rangle)$ over all n-partite states $|\psi\rangle$ such that $\mathrm{Overlap}_1(|\psi\rangle) = \omega$.*

The main result of Harrow and Montanaro [HM13] is the following upper-bound on $\mathrm{PT}_n(\omega)$. It will be more convenient to parameterize their result by ϵ, where $1 - \epsilon = \omega$.

Theorem 7 ([HM13, Theorem 1])**.** *For all $n \geq 1$ and $0 < \epsilon < 1$,*

$$\mathrm{PT}_n(1-\epsilon) \leq \min\{1 - \epsilon + \epsilon^2 + \epsilon^{3/2}, 1 - \tfrac{11}{512}\epsilon\}.$$

Equivalently, we may write

$$\mathrm{PT}_n(1-\epsilon) \leq \begin{cases} 1 - \epsilon + \epsilon^2 + \epsilon^{3/2} & \text{if } \epsilon \leq \epsilon_0, \\ 1 - \frac{11}{512}\epsilon & \text{if } \epsilon \geq \epsilon_0, \end{cases}$$

where $\epsilon_0 = \frac{1}{512}(757 - 16\sqrt{1258}) \approx 0.37$. We include a plot of this upper-bound in Figure 2-2.

The most important regime of parameters is when ϵ is a constant, in which case this result states that the product test rejects with constant probability. This implies that two copies are sufficient to test if $|\psi\rangle$ is constantly far from being product. Theorem 7 is a key ingredient in Harrow and Montanaro's proof that $\mathsf{QMA}(2) = \mathsf{QMA}(k)$ for $k \geq 2$ [HM13]. Here, $\mathsf{QMA}(k)$ refers to *Quantum Merlin Arthur with multiple certificates*, the complexity class which contains all problems solvable by a quantum polynomial-time verifier with the help of k unentangled proofs. Their result shows that a verifier can use two unentangled copies of a proof $|\psi\rangle$ to simulate k unentangled proofs by running the product test to enforce that it is of the form $|\psi_1\rangle \otimes \cdots \otimes |\psi_k\rangle$. As further applications of Theorem 7, they are able to derive hardness results for numerous (19, in fact!) problems both in and out of quantum information theory related to entanglement, tensor optimization, and other topics. For example, one of their applications is to the problem of detecting separability, in which the goal is to compute whether a mixed state ρ on two subsystems of dimension d (described by a $d^2 \times d^2$ complex matrix) is separable or entangled. They show that there exists a constant $\delta > 0$ such that if K is a convex set in which every element has trace distance δ to a separable state, then there is no polynomial time algorithm for computing whether $\rho \in K$ unless 3-SAT $\in \mathsf{DTIME}(\exp(\sqrt{n}\log^{O(1)}(n)))$. See [HM13, Section 4.2] for further details and descriptions of the 18 other applications.

Our first result is a new and simpler analysis of the product test which yields an improved bound. We show the following.

Theorem 8 (Product test upper-bound)**.** *For all $n \geq 1$,*

$$\mathrm{PT}_n(\omega) \leq \begin{cases} \omega^2 - \omega + 1 & \text{if } \omega \geq \frac{1}{2}, \\ \frac{1}{3}\omega^2 + \frac{2}{3} & \text{otherwise.} \end{cases}$$

We include a plot of this upper-bound in Figure 2-2.

To compare this with Theorem 7, if we set $1 - \epsilon = \omega$ then we can rewrite this bound as

$$\mathrm{PT}_n(1-\epsilon) \leq \begin{cases} 1 - \epsilon + \epsilon^2 & \text{if } \epsilon \leq \frac{1}{2}, \\ 1 - \frac{2}{3}\epsilon + \frac{1}{3}\epsilon^2 & \text{otherwise.} \end{cases}$$

This improves upon Theorem 7 for all choices of $\epsilon > 0$, i.e. all $\omega < 1$, which answers open problem no. 2 from [HM13] and question no. 5 from [MdW13]. In addition, the bound we achieve when $\omega \geq \frac{1}{2}$ is optimal, as the following well-known example shows (cf. [HM13, Page 31]).

Proposition 9 (Product test lower-bound)**.** *For $n = 2$ and $\omega \geq \frac{1}{2}$, consider the state $|\psi\rangle = \sqrt{\omega}|11\rangle + \sqrt{1-\omega}|22\rangle$. Then $\mathrm{Overlap}_1(|\psi\rangle) = \omega$ and*

$$\mathrm{PT}_2(|\psi\rangle) = \omega^2 - \omega + 1.$$

In addition, for $n > 2$, consider $|\psi\rangle \otimes |\phi\rangle$, where $|\phi\rangle$ is any product state in $\mathbb{C}^{d_3} \otimes \cdots \otimes \mathbb{C}^{d_n}$. Then this has the same overlap and probability of success as $|\psi\rangle$.

The proof of Proposition 9 is standard and we include it in Section 2.2.2. Combining Theorem 8 and Proposition 9 allows us to exactly compute $\mathrm{PT}(\omega)$ for $\omega \geq \frac{1}{2}$.

Corollary 10 (Product test, tight bound)**.** *For all $n \geq 2$ and $\omega \geq \frac{1}{2}$, $\mathrm{PT}_n(\omega) = \omega^2 - \omega + 1$.*

This settles the performance of the product test when $\omega \geq \frac{1}{2}$. The regime of $\omega < \frac{1}{2}$ remains open, however. As [HM13] points out, this regime "is generally somewhat mysterious", and getting a better understanding of this case is part of open problem no. 2 in their work. One possible starting point is to understand the behavior of $\mathrm{PT}_n(\omega)$ as $\omega \to 0$. For example, as [HM13] show on page 32, the d-dimensional maximally entangled state $|\psi\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |ii\rangle$ has

$$\omega = 1/d \quad \text{and} \quad \mathrm{PT}_2(|\psi\rangle) = \tfrac{1}{2}(1 + \tfrac{1}{d}).$$

This suggests the following question: does $\mathrm{PT}_n(\omega) \to \frac{1}{2}$ as $\omega \to 0$?

Our proof of Theorem 8 is a simple inductive argument. Decomposing the product test measurement as $\Pi_{\mathrm{Prod}} = (\mathbb{1} \otimes \Pi_{\mathrm{SWAP}}^{\otimes n-1}) \cdot (\Pi_{\mathrm{SWAP}} \otimes \mathbb{1})$, we can view it as first performing the SWAP test on the first qudit register of $|\psi\rangle$ and then, if it succeeds, performing the $(n-1)$-qudit product test on the remaining qudit registers. Supposing that $|\psi\rangle$ is far from being a product state, either the first qudit of $|\psi\rangle$ is highly entangled with the remaining qudits, or the other qudits are far from being a product state (even conditioned on the first SWAP test succeeding). In the first case, the SWAP test rejects with good probability, and in the second case, the $(n-1)$-qubit products test rejects with good probability, by induction. Balancing between these two cases gives our bound.

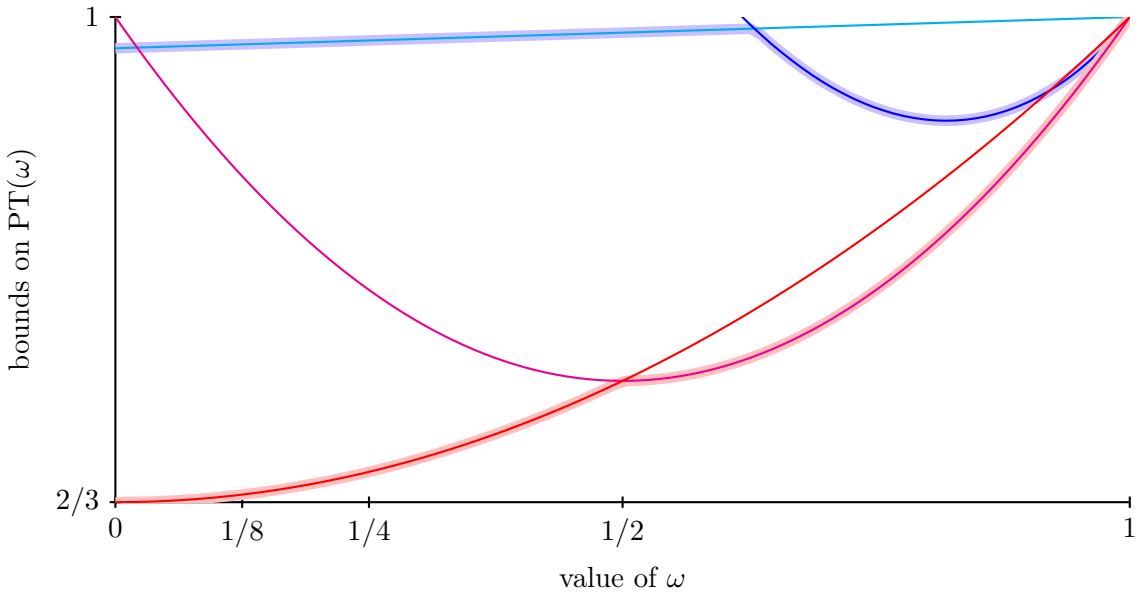


Figure 2-2: Upper bounds on $\mathrm{PT}(\omega)$ as a function of $\omega = 1 - \epsilon$.

- The red line is the function $\frac{1}{3}\omega^2 + \frac{2}{3}$ and the magenta line is the function $\omega^2 - \omega + 1$. The thick pink line is the minimum of the two. This is the upper bound we prove.
- The blue line is the function $1 - \epsilon + \epsilon^2 + \epsilon^{3/2}$ and the cyan line is the function $1 - \frac{11}{512}\epsilon$. The thick light blue line is the minimum of the two. This is the upper bound of Harrow and Montanaro [HM13].

The proof of the bound $\mathrm{PT}_n(\omega) \leq \frac{1}{3}\omega^2 + \frac{2}{3}$ is especially simple and fits in a page. Though weaker than our general bound when $\omega \geq \frac{1}{2}$, this bound is still sufficient to recover all the applications of the product test in [HM13], including the proof that $\mathsf{QMA}(2) = \mathsf{QMA}(k)$ for $k \geq 2$. We include it as a separate argument in Section 2.2.1. The proof of the general bound from Theorem 8 is contained in Section 2.2.2.

So far we have considered the case of product testing where the number of copies m is exactly two, but the property testing model requires us to take m sufficiently large to detect non-product states with constant probability. For even m, a simple strategy is to run $m/2$ parallel copies of the product test and reject if any of them rejects. If $\mathrm{Overlap}_1(|\psi\rangle) = 1 - \epsilon$, then this will accept with probability at most $(1 - \frac{2}{3}\epsilon + \frac{1}{3}\epsilon^2)^{m/2}$. Making this probability smaller than $\frac{1}{3}$ as required by Definition 3 entails setting $m = O(1/\epsilon)$. Using the distance $\delta = \sqrt{\epsilon}$, this can be stated as follows.

Proposition 11 (Copy complexity of testing product states)**.** *Following the language of Definition 3, testing whether a state $|\psi\rangle \in \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n}$ is a product state can be done using $m = O(1/\delta^2)$ copies and with prefect completeness.*

This is optimal, as $\Omega(1/\delta^2)$ copies are always required to distinguish between two states which are δ-far from each other in trace distance. We note that the same copy complexity follows from Theorem 7, the bound given by Harrow and Montanaro [HM13].

2.1.2 Testing matrix product states

Having already considered the case of MPS testing with bond dimension $r = 1$, we now consider the case of bond dimension $r > 1$. To our knowledge, there is no prior work on this problem.

One idea for testing MPS is to use the general "test-by-learning" framework from property testing. In our case, given a state $|\psi\rangle$, this entails performing MPS tomography on $|\psi\rangle$ to learn an $\mathsf{MPS}(r)$ approximation $|\phi\rangle$ and then applying the SWAP test on $|\psi\rangle$ and $|\phi\rangle$. If $|\psi\rangle$ is in $\mathsf{MPS}(r)$, then $|\phi\rangle$ will be a good approximation, and so the SWAP test will usually succeed, but if $|\psi\rangle$ is far from $\mathsf{MPS}(r)$, then $|\phi\rangle$ will be a bad approximation, and so the SWAP test will usually fail. Various algorithms for MPS tomography have been proposed in the literature, for example those in the works [CPF+10, LMH+17]. One would expect that since states in $\mathsf{MPS}(r)$ can be described using ndr^2 parameters, where d is the largest subsystem dimension, the optimal algorithm for $\mathsf{MPS}(r)$ tomography should use $O(ndr^2/\delta^2)$ copies, though this precise bound is not yet known to our knowledge. We propose and analyze a more direct MPS testing algorithm that improves on this "test-by-learning" method by a factor of $O(d)$.

We begin by designing an algorithm for this problem which we call the *MPS tester*. The MPS tester is motivated by the fact that $|\psi_{1,\dots,n}\rangle$ is in $\mathsf{MPS}(r)$ if and only if $\psi_{1,\dots,k}$ has rank r for each $1 \leq k \leq n$. This relates the problem of MPS testing to the problem of *rank testing*, i.e. of testing whether a mixed state ρ has rank r, which was previously considered in the work of O'Donnell and Wright [OW15]. They designed an algorithm called the *rank tester* which can test whether ρ is rank r using $m = \Theta(r^2/\delta)$ copies of ρ. When the $r = 1$ rank tester is run with $m = 2$ copies of ρ, it is equivalent to the SWAP test, and for larger values of r and m it uses a generalization of the SWAP measurement known as *weak Schur sampling*. It has perfect completeness, meaning that it always accepts states of rank r, and in fact it is the optimal test for states of rank r with perfect completeness, as shown in [OW15, Proposition 6.1].

With the rank tester in hand, we define the MPS tester to be the algorithm which simultaneously performs a separate instance of the rank tester on $\psi_{1,\dots,k}$ for each $1 \leq k \leq n$ and accepts if each instance of the rank tester accepts. We include an illustration of the MPS tester in Figure 2-1b. We show that this test has perfect completeness, meaning that it accepts every state in $\mathsf{MPS}(r)$ with probability 1, although we are not sure if it is the optimal algorithm with perfect completeness; we view this as an interesting open direction. We show the following bound on its copy complexity.

Theorem 12 (Copy complexity of the MPS tester)**.** *Given $m = O(nr^2/\delta^2)$ copies of a state $|\psi\rangle \in \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n}$, the MPS tester tests whether $|\psi\rangle$ is in $\mathsf{MPS}(r)$ with perfect completeness.*

To prove this result, we first show that if $|\psi\rangle$ is δ-far from the set $\mathsf{MPS}(r)$, then there exists $1 \leq k \leq n$ such that $\psi_{1,\dots,k}$ is $\delta' = (\delta^2/2n)$-far from being rank-r. Then the probability that the MPS tester accepts $|\psi\rangle$ is at most the probability that the rank tester accepts $\psi_{1,\dots,k}$, and this is at most $1/3$ given that we are using $O(r^2/\delta') = O(nr^2/\delta^2)$ copies of $|\psi\rangle$. One minor technicality that arises is checking that the MPS tester does indeed perform a valid measurement, which entails showing that the rank testers for each $\psi_{1,\dots,k}$ can all be simultaneously measured.

Remark 13 (Time complexity of the MPS tester)**.** *The m-copy rank tester of [OW15] can be performed efficiently with a quantum circuit of size* $\mathrm{poly}\,(m, \log(d))$ *using the algorithm of [Kro19] or [Har05, Page 160] that implements weak Schur sampling. Here d is the dimension of the state ρ whose m copies $\rho^{\otimes m}$ are input to the rank tester. The MPS tester performs the rank tester on $m = nr^2/\delta^2$ copies of the reduced states $\psi_{1,\dots,k}$ for $1 \leq k \leq n$. The maximum dimension of these*

reduced states is less than $\log(d_1 \dots d_n) \leq n \log(d)$ *where* $d = \max_{i \in [n]} d_i$. *Hence, the MPS tester can be implemented with a quantum circuit of size* $\mathrm{poly}\,(n, r, 1/\delta, \log(d))$.

We believe that the bound in Theorem 12 is not tight, and that an inductive argument similar to our analysis of the product tester should be able to improve it. As an example, consider the "bunny state"

$$|b_n\rangle = \tfrac{1}{\sqrt{n-1}}(|110\cdots0\rangle + |011\cdots0\rangle + \cdots + |0\cdots011\rangle).$$

We can show that this state, which is in $\mathsf{MPS}(3)$, has $\mathrm{Overlap}_2(|b_n\rangle) \leq \frac{2}{3}$. But does the $r = 2$ MPS tester detect this? The above analysis suggests we should find the reduced density matrix $(b_n)_{1,\dots,i}$ which is farthest from being rank 2. It can be checked that $(b_n)_1$ and $(b_n)_{1,\dots,n-1}$ are both rank 2. Otherwise, for $i \in \{2, \dots, n-2\}$, the Schmidt decomposition of $|b_n\rangle$ into subsystems $\{1, \dots, i\}$ and $\{i+1, \dots, n\}$ is

$$|b_n\rangle = \sqrt{\tfrac{i-1}{n-1}}|b_i\rangle \otimes |0\cdots0\rangle + \tfrac{1}{\sqrt{n-1}}|0\cdots01\rangle \otimes |10\cdots0\rangle + \sqrt{\tfrac{n-i-1}{n-1}}|b_{n-i}\rangle \otimes |0\cdots0\rangle$$

Hence, $(b_n)_{1,\dots,i}$ has eigenvalues $\frac{i-1}{n-1}$, $\frac{1}{n-1}$, $\frac{n-i-1}{n-1}$, and so it is distance $\frac{1}{n-1}$ from rank-2. As a result, the rank tester needs $O(n)$ copies of $(b_n)_{1,\dots,i}$ to detect this, and therefore the MPS tester needs $O(n)$ copies of $|b_n\rangle$ if we use our above analysis. However, we have done a more careful analysis of the bunny state in line with the inductive argument for the product test, and we can show that the MPS tester only needs 3 copies of $|b_n\rangle$ to detect that it is not in $\mathsf{MPS}(2)$. In particular, the MPS tester rejects $|b_n\rangle^{\otimes 3}$ with probability at least $\frac{1}{6}$. This means that the above analysis is too pessimistic, at least for the bunny state.

Unfortunately, we were unable to carry this proof strategy out in general. One difficulty is that we are not even sure what upper bound to conjecture for this problem. Originally, we had guessed that the MPS tester only needed $m = O(r^2/\delta^2)$ copies, or perhaps some other copy complexity which is independent of n, but we now know this is false, due to the following lower bound.

Theorem 14 (MPS testing lower bound)**.** *For* $r \geq 2$ *and* $\delta \leq 1/\sqrt{2}$, *testing whether a state* $|\psi\rangle \in \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n}$, *is in* $\mathsf{MPS}(r)$ *requires* $\Omega(n^{1/2}/\delta^2)$ *copies of* $|\psi\rangle$.

Theorem 14 shows that a polynomial dependence on n, as in Theorem 12, *is* required, even for the case of bond dimension $r = 2$. This in sharp contrast to the $r = 1$ case of product testing, in which a constant number of copies suffice, independent of n. This leaves open the following question: what is the optimal copy complexity for $\mathsf{MPS}(r)$ testing, for $r \geq 2$?

The proof of the lower bound consists of two parts. We consider a quantum state $|\Phi_n\rangle = |\varphi\rangle^{\otimes \frac{n}{2}}$ where $|\varphi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ for some $d \geq 2r - 1$ and $\mathrm{Dist}_r\,(|\varphi\rangle) = \Omega(\delta/\sqrt{n})$. In the first step, we use an inductive argument to prove that $\mathrm{Dist}_r\,(|\Phi_n\rangle) \geq \delta$. In the second step, we consider the density matrix corresponding to the ensemble of states obtained by applying random local unitaries to the subsystems of $|\Phi_n\rangle$. Since all the states in this ensemble are δ-far from $\mathsf{MPS}(r)$, a tester should reject this density matrix with probability at least $2/3$. We show that without sufficiently large number of copies, no $\mathsf{MPS}(r)$ tester that accepts states in $\mathsf{MPS}(r)$ with probability $\geq 2/3$ can also reject this density matrix with probability $\geq 2/3$.

We prove our MPS tester upper bound (Theorem 12) in Section 2.4 and our MPS testing lower bound (Theorem 14) in Section 2.5.

2.2 The product test

2.2.1 A simple analysis of the product test

We begin with a simple analysis of the product test which shows that it rejects with constant probability if $|\psi\rangle$ is a constant distance from the set of product states. This is sufficient to show $\mathsf{QMA}(2) = \mathsf{QMA}(k)$ via the proof of [HM13].

Theorem 15 (Product test, simple bound)**.** *For all $n \geq 1$, $\mathrm{PT}_n(\omega) \leq \frac{1}{3}\omega^2 + \frac{2}{3}$.*

Proof. By induction, the $n = 1$ case being trivial. For the inductive step, let us assume Theorem 15 holds for $(n-1)$-partite states. Let $|\psi\rangle \in \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n}$ be a state with $\mathrm{Overlap}(|\psi\rangle) = \omega$. For shorthand we write $d := d_1$. Note that the product test measurement can be written as $\Pi_{\mathrm{SWAP}}^{\otimes n} = \left(\mathbb{1} \otimes \Pi_{\mathrm{SWAP}}^{\otimes n-1}\right) \cdot (\Pi_{\mathrm{SWAP}} \otimes \mathbb{1})$. We can therefore view the test as first applying Π_{SWAP} to the first subsystem, and, if it succeeds, then applying $\Pi_{\mathrm{SWAP}}^{\otimes n-1}$ (i.e. the product test) to the resulting reduced state on the last $n-1$ subsystems. The probability this succeeds we bound by induction.

We begin by taking the Schmidt decomposition of $|\psi\rangle$ into subsystems $\{1\}$ and $\{2, \ldots, n\}$:

$$|\psi\rangle = \sqrt{\lambda_1}\,|a_1\rangle\,|b_1\rangle + \cdots + \sqrt{\lambda_d}\,|a_d\rangle\,|b_d\rangle\,,$$

where $\lambda_1 \geq \cdots \geq \lambda_d$, $|a_i\rangle \in \mathbb{C}^d$, and $|b_i\rangle \in \mathbb{C}^{d_2} \otimes \cdots \otimes \mathbb{C}^{d_n}$. As a result,

$$|\psi\rangle^{\otimes 2} = \sum_{i\in[d]} \lambda_i\,|a_i\rangle^{\otimes 2}\,|b_i\rangle^{\otimes 2} + \sum_{i<j} \sqrt{\lambda_i\lambda_j}\,(|a_ia_j\rangle\,|b_ib_j\rangle + |a_ja_i\rangle\,|b_jb_i\rangle)\,.$$

The result of applying the first projector to $|\psi\rangle^{\otimes 2}$ is therefore

$$\Pi_{\mathrm{SWAP}} \otimes I \cdot |\psi\rangle^{\otimes 2} = \sum_{i\in[d]} \lambda_i\,|a_i\rangle^{\otimes 2}\,|b_i\rangle^{\otimes 2} + \sum_{i<j} \sqrt{\lambda_i\lambda_j}\left(\frac{|a_ia_j\rangle + |a_ja_i\rangle}{\sqrt{2}}\right)\left(\frac{|b_ib_j\rangle + |b_jb_i\rangle}{\sqrt{2}}\right). \quad (2.1)$$

We note that this vector's two-norm, and hence the probability the test passes in the first step, is $\mu := \sum_i \lambda_i^2 + \sum_{i<j} \lambda_i\lambda_j$. Conditioned on this, the mixed state of subsystems $2, \ldots, n$ is

$$|b_i\rangle^{\otimes 2} \text{ with prob. } \frac{\lambda_i^2}{\mu}, \quad \frac{1}{\sqrt{2}}\,(|b_ib_j\rangle + |b_jb_i\rangle) \text{ with prob. } \frac{\lambda_i\lambda_j}{\mu}.$$

This is by (2.1) and the fact that the $|a_i\rangle^{\otimes 2}$'s and the $(|a_ia_j\rangle + |a_ja_i\rangle)$'s are orthogonal. We must now bound the probability that the product test on $n-1$ subsystems succeeds in each of these cases. In the first case this is $\mathrm{PT}_{n-1}(|b_1\rangle)$, and in the rest of the cases we will charitably bound the probability by 1. This gives us:

$$\mathrm{PT}_n(|\psi\rangle) \leq \mu \cdot \left(\frac{\lambda_1^2}{\mu} \cdot \mathrm{PT}_{n-1}(|b_1\rangle) + \sum_{i>1} \frac{\lambda_i^2}{\mu} + \sum_{i<j} \frac{\lambda_i\lambda_j}{\mu}\right) = \lambda_1^2 \cdot \mathrm{PT}_{n-1}(|b_1\rangle) + \sum_{i>1} \lambda_i^2 + \sum_{i<j} \lambda_i\lambda_j. \quad (2.2)$$

Writing $\phi = \text{Overlap}(|b_1\rangle)$, the inductive hypothesis gives us

$$\begin{aligned}\text{PT}_n(|\psi\rangle) &\leq \lambda_1^2 \cdot \left(\frac{1}{3}\phi^2 + \frac{2}{3}\right) + \sum_{i>1} \lambda_i^2 + \sum_{i<j} \lambda_i \lambda_j \\ &= \left(\frac{1}{3}(\lambda_1\phi)^2 + \frac{2}{3}\right) - \frac{1}{3}\sum_{i>1}(\lambda_1 - \lambda_i)\lambda_i - \frac{1}{3}\sum_{1<i<j}\lambda_i\lambda_j \leq \frac{1}{3}(\lambda_1\phi)^2 + \frac{2}{3}, \end{aligned} \tag{2.3}$$

where the last inequality follows because $\lambda_1 \geq \cdots \geq \lambda_d$. Now, by definition of ϕ, there exists a product state $|v\rangle \in \mathbb{C}^{d_2} \otimes \cdots \otimes \mathbb{C}^{d_n}$ such that $\phi = |\langle b_1|v\rangle|^2$. But then $|a_1\rangle \otimes |v\rangle$, also a product state, has squared-inner-product $\lambda_1\phi$ with $|\psi\rangle$, meaning $\lambda_1\phi \leq \text{Overlap}(\psi) = \omega$, and so by (2.3) the test succeeds with probability at most $\frac{1}{3}\omega^2 + \frac{2}{3}$. □

2.2.2 A tight analysis of the product test for $\omega \geq \frac{1}{2}$

Next, we sharpen our upper-bound from Theorem 15 in the $\omega \geq \frac{1}{2}$ case.

Theorem 16 (Product test, sharpened bound; Theorem 8 restated)**.** *For all $n \geq 1$,*

$$\text{PT}_n(\omega) \leq \begin{cases} \omega^2 - \omega + 1 & \text{if } \omega \geq \frac{1}{2}, \\ \frac{1}{3}\omega^2 + \frac{2}{3} & \text{otherwise.} \end{cases}$$

We begin by showing that Theorem 16 is tight for $\omega \geq \frac{1}{2}$ using a simple example from [HM13, Page 31]).

Proposition 17 (Product test lower-bound; Proposition 9 restated)**.** *For $n = 2$ and $\omega \geq \frac{1}{2}$, consider the state $|\psi\rangle = \sqrt{\omega}|11\rangle + \sqrt{1-\omega}|22\rangle$. Then $\text{Overlap}_1(|\psi\rangle) = \omega$ and*

$$\text{PT}_2(|\psi\rangle) = \omega^2 - \omega + 1.$$

In addition, for $n > 2$, consider $|\psi\rangle \otimes |\phi\rangle$, where $|\phi\rangle$ is any product state in $\mathbb{C}^{d_3} \otimes \cdots \otimes \mathbb{C}^{d_n}$. Then this has the same overlap and probability of success as $|\psi\rangle$.

Proof. First, we show $\text{Overlap}_1(|\psi\rangle) = \omega$. This is because if $|a\rangle = \sum_{i=1}^{d_1} \alpha_i|i\rangle$ and $|b\rangle = \sum_{i=1}^{d_2} \beta_i|i\rangle$,

$$|\langle\psi|ab\rangle|^2 = |\sqrt{\omega}\cdot\alpha_1\beta_1 + \sqrt{1-\omega}\cdot\alpha_2\beta_2|^2 \leq \omega\cdot|\alpha_1\beta_1|^2 + (1-\omega)\cdot|\alpha_2\beta_2|^2.$$

This is maximized by taking $\alpha_1 = \beta_1 = 1$, in which case it equals ω. Next, the probability of success is $\text{PT}_2(|\psi\rangle) = \|\Pi_{\text{SWAP}}^{\otimes 2}|\psi\rangle^{\otimes 2}\|^2$, and so we first compute $\Pi_{\text{SWAP}}^{\otimes 2}|\psi\rangle^{\otimes 2}$:

$$\begin{aligned} &\left(\frac{\mathbb{1} + \text{SWAP}}{2}\right)^{\otimes 2} \cdot (\sqrt{\omega}|11\rangle + \sqrt{1-\omega}|22\rangle)^{\otimes 2} \\ &= \left(\frac{\mathbb{1} + \text{SWAP}}{2}\right)^{\otimes 2} \cdot (\omega|11\rangle|11\rangle + \sqrt{\omega(1-\omega)}(|11\rangle|22\rangle + |22\rangle|11\rangle) + (1-\omega)|22\rangle|22\rangle) \\ &= \omega|11\rangle|11\rangle + \sqrt{\omega(1-\omega)}\left(\frac{|12\rangle + |21\rangle}{\sqrt{2}}\right) \otimes \left(\frac{|12\rangle + |21\rangle}{\sqrt{2}}\right) + (1-\omega)|22\rangle|22\rangle. \end{aligned}$$

The squared length of this is $\omega^2 + \omega(1-\omega) + (1-\omega)^2 = \omega^2 - \omega + 1$, and so this equals $\text{PT}_2(|\psi\rangle)$.

The $n > 2$ case is an immediate consequence of the $n = 2$ case. □

Now we prove Theorem 16.

Proof of Theorem 16. By induction, where again the $n = 1$ case is trivial. For the inductive step, let $f_0(\omega) = \frac{1}{3}\omega^2 + \frac{2}{3}$ and $I_0 = [0, \frac{1}{2}]$. Let $f_1(\omega) = \omega^2 - \omega + 1$ and $I_1 = [\frac{1}{2}, 1]$. The upper bound on $\mathrm{PT}(\omega)$ we are trying to show is

$$\mathsf{UB}(\omega) = \begin{cases} f_0(\omega) & \text{if } \omega \in I_0, \\ f_1(\omega) & \text{if } \omega \in I_1. \end{cases}$$

Note that $\mathsf{UB}(\omega)$ is a non-decreasing function of ω, and that $\mathsf{UB}(\omega) = \min\{f_0(\omega), f_1(\omega)\}$ because $f_0(\omega) \leq f_1(\omega)$ for $\omega \in I_0$ and $f_1(\omega) \leq f_0(\omega)$ for $\omega \in I_1$. Recalling the proof of Theorem 15, we showed in (2.2) that

$$\mathrm{PT}_n(|\psi\rangle) \leq \lambda_1^2 \cdot \mathrm{PT}_{n-1}(|b_1\rangle) + \sum_{i>1} \lambda_i^2 + \sum_{i<j} \lambda_i \lambda_j.$$

Write $\phi = \mathrm{Overlap}(|b_1\rangle)$, and suppose that $\phi \in I_\alpha$, for $\alpha \in \{0, 1\}$. Then the inductive hypothesis gives us

$$\mathrm{PT}_n(|\psi\rangle) \leq \lambda_1^2 \cdot f_\alpha(\phi) + \sum_{i>1} \lambda_i^2 + \sum_{i<j} \lambda_i \lambda_j. \tag{2.4}$$

Recall also that $\lambda_1 \phi \leq \mathrm{Overlap}(|\psi\rangle) = \omega$, and suppose that $\lambda_1 \phi \in I_\beta$, for $\beta \in \{0, 1\}$. Our goal will be to show the inequality (2.4) $\leq f_\beta(\lambda_1 \phi)$. Then because $\lambda_1 \phi \in I_\beta$ we have $f_\beta(\lambda_1 \phi) = \mathsf{UB}(\lambda_1 \phi)$, and because $\lambda_1 \phi \leq \omega$ and $\mathsf{UB}(\cdot)$ is a nondecreasing function, we have $\mathsf{UB}(\lambda_1 \phi) \leq \mathsf{UB}(\omega)$, completing the inductive step. Note that we only have to show (2.4) $\leq f_\beta(\lambda_1 \phi)$ in the case that $\beta \leq \alpha$, as $\lambda_1 \phi \leq \phi$, so we will never have $\beta > \alpha$. In particular, we need not consider the case $\alpha = 0, \beta = 1$.

Case 1: $\alpha = 1, \beta = 1$. This case can be shown as follows.

$$\begin{aligned}
&\lambda_1^2 \cdot f_1(\phi) + \sum_{i>1} \lambda_i^2 + \sum_{i<j} \lambda_i \lambda_j \\
&= \lambda_1^2 \cdot (\phi^2 - \phi + 1) + \sum_{i>1} \lambda_i^2 + \sum_{i<j} \lambda_i \lambda_j \\
&= ((\lambda_1 \phi)^2 - \lambda_1 \phi + 1) - 1 + \lambda_1(1 - \lambda_1)\phi + \sum_{i} \lambda_i^2 + \sum_{i<j} \lambda_i \lambda_j \\
&\leq f_1(\lambda_1 \phi) - 1 + \lambda_1(1 - \lambda_1) + \sum_{i} \lambda_i^2 + \sum_{i<j} \lambda_i \lambda_j && \text{(because } \phi \leq 1\text{)} \\
&= f_1(\lambda_1 \phi) - 1 + \sum_{1<j} \lambda_1 \lambda_j + \sum_{i} \lambda_i^2 + \sum_{i<j} \lambda_i \lambda_j \\
&\leq f_1(\lambda_1 \phi) - 1 + \sum_{i} \lambda_i^2 + 2 \sum_{i<j} \lambda_i \lambda_j \\
&= f_1(\lambda_1 \phi) - 1 + \Big(\sum_{i} \lambda_i \Big)^2 \\
&= f_1(\lambda_1 \phi). && \text{(because } \lambda_1 + \cdots + \lambda_d = 1\text{)}
\end{aligned}$$

This completes the proof.

Case 2: $\alpha = 1, \beta = 0$. This case can be shown as follows.

$$\begin{aligned}
&\lambda_1^2 \cdot f_1(\phi) + \sum_{i>1} \lambda_i^2 + \sum_{i<j} \lambda_i \lambda_j \\
&= \lambda_1^2 \cdot (\phi^2 - \phi + 1) + \sum_{i>1} \lambda_i^2 + \sum_{i<j} \lambda_i \lambda_j \\
&= (\tfrac{1}{3}(\lambda_1 \phi)^2 + \tfrac{2}{3}) - \tfrac{2}{3} + \lambda_1^2(\tfrac{2}{3}\phi^2 - \phi + \tfrac{1}{3}) + \tfrac{2}{3}\lambda_1^2 + \sum_{i>1} \lambda_i^2 + \sum_{i<j} \lambda_i \lambda_j \\
&\leq f_0(\lambda_1 \phi) - \tfrac{2}{3} + \tfrac{2}{3}\lambda_1^2 + \sum_{i>1} \lambda_i^2 + \sum_{i<j} \lambda_i \lambda_j && \text{(because } \tfrac{1}{2} \leq \phi \leq 1\text{)} \\
&\leq f_0(\lambda_1 \phi) - \tfrac{2}{3} + \tfrac{2}{3} \sum_{i} \lambda_i^2 + \tfrac{1}{3} \sum_{1<j} \lambda_1 \lambda_j + \sum_{i<j} \lambda_i \lambda_j && \text{(because } \lambda_j \leq \lambda_1\text{)} \\
&\leq f_0(\lambda_1 \phi) - \tfrac{2}{3} + \tfrac{2}{3} \sum_{i} \lambda_i^2 + \tfrac{4}{3} \sum_{i<j} \lambda_i \lambda_j \\
&= f_0(\lambda_1 \phi) - \tfrac{2}{3} + \tfrac{2}{3} \Big(\sum_i \lambda_i \Big)^2 \\
&= f_0(\lambda_1 \phi). && \text{(because } \lambda_1 + \cdots + \lambda_d = 1\text{)}
\end{aligned}$$

This completes the proof.

Case 3: $\alpha = 0, \beta = 0$. In this case, we aim to show that

$$\lambda_1^2 \cdot f_0(\phi) + \sum_{i>1} \lambda_i^2 + \sum_{i<j} \lambda_i \lambda_j \leq f_0(\lambda_1 \phi).$$

This was already shown in Equation (2.3) in the proof of Theorem 15 for all ϕ and λ_1. This concludes case 3 and the proof of the theorem. □

2.3 Preliminaries for MPS testing

2.3.1 Low rank approximation to MPS

Lemma 18 (Young-Eckart Theorem [EY36]). *Consider a bipartite state* $|\psi\rangle \in \mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}$ *with* $d_1 \geq d_2$*, and let*

$$|\psi\rangle = \sum_{i=1}^{d_2} \sqrt{\lambda_i} |a_i\rangle |b_i\rangle$$

be its Schmidt decomposition, where $\lambda_1 \geq \cdots \geq \lambda_{d_2}$*. Then the maximum overlap of* $|\psi\rangle$ *with a state in* $\mathsf{MPS}(r)$ *is* $\mathrm{Overlap}_r(|\psi\rangle) = \sum_{i=1}^r \lambda_i$*, and it is achieved by the state*

$$|\phi\rangle = \frac{1}{\sqrt{\sum_{i=1}^r \lambda_i}} \sum_{i=1}^r \sqrt{\lambda_i} |a_i\rangle |b_i\rangle.$$

While Lemma 18 gives the closest $\mathsf{MPS}(r)$ approximation to a bipartite state, a general lower bound on the overlap between an n-partite state and $\mathsf{MPS}(r)$ can be also derived. This is stated in the following lemma.

Lemma 19 (Low-rank Approximation, Lemma 1 of [VC06])**.** *Consider an n-partite state $|\psi\rangle \in \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n}$. For each $i \in \{1, \ldots, n-1\}$, write the Schmidt decomposition of $|\psi\rangle$ across the subsystems $\{1, \ldots, i\}$ and $\{i+1, \ldots, n\}$ as*

$$|\psi\rangle = \sum_{j=1}^{D_i} \sqrt{\lambda_j^{(i)}} |a_j^{(i)}\rangle |b_j^{(i)}\rangle,$$

where $D_i = \min\{d_1 \cdots d_i, d_{i+1} \cdots d_n\}$ and $\lambda_1^{(i)} \geq \cdots \geq \lambda_{D_i}^{(i)}$. Then there exists a state $|\phi\rangle \in \mathsf{MPS}(r)$ with (unsquared) overlap

$$|\langle \phi | \psi \rangle| \geq 1 - \sum_{i=1}^{n-1} \sum_{j=r+1}^{D_i} \lambda_j^{(i)}.$$

2.3.2 Representation theory and weak Schur sampling

A common scenario involves having i.i.d. copies of a state ρ, which in this chapter could be the state $|\psi_{1,\ldots,n}\rangle$ or one of its marginals. These copies are invariant under permutation. The spectrum of the state ρ, including its rank, is also invariant under the action of any unitary operator U that maps ρ to $U\rho U^\dagger$. In this section, we study these symmetries and discuss how we can exploit them in the analysis of our MPS tester. The content of this section is already covered in detail in previous works (see [Wri16, Section 2.5], [Har05, Section 5.3], or [CHW07, Section 2] for example). Here we briefly review these topics.

Definition 20 (Partitions)**.** *A partition of m, denoted by $\mu \vdash m$, is a list of nonnegative integers $\mu = (\mu_1, \mu_2, \ldots, \mu_k)$ that satisfy $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_k$ and $\mu_1 + \mu_2 + \cdots + \mu_k = m$. We call the number of nonzero elements μ_i in μ the length of the partition and denote it by $\ell(\mu)$.*

The group of all permutations of $\{1, \ldots, m\}$ is known as the *symmetric group*, and we denote it by S_m. In addition, we denote the group of $d \times d$ unitary operators by U_d. Two natural representations of the groups S_m and U_d over the space $(\mathbb{C}^d)^{\otimes m}$ are given as follows.

$$\begin{aligned}
\mathcal{P}(\pi)\ |a_1\rangle \otimes |a_2\rangle \otimes \ldots \otimes |a_m\rangle &= |a_{\pi^{-1}(1)}\rangle \otimes |a_{\pi^{-1}(2)}\rangle \otimes \ldots \otimes |a_{\pi^{-1}(m)}\rangle, \\
\mathcal{Q}(U)\ |a_1\rangle \otimes |a_2\rangle \otimes \ldots \otimes |a_m\rangle &= \ (U|a_1\rangle) \otimes (U|a_2\rangle) \otimes \ldots \otimes (U|a_m\rangle),
\end{aligned}$$

where $\{\bigotimes_{i=1}^m |a_i\rangle\}$ with $a_i \in [d]$ is a basis for $(\mathbb{C}^d)^{\otimes m}$, and $\pi \in \mathsf{S}_m$, $U \in \mathsf{U}_d$. The irreducible representations (*irreps*) of the symmetric group, denoted $\mathcal{P}_\mu$, are indexed by partitions $\mu \vdash m$. Similarly, the polynomial irreps of the unitary group, denoted $\mathcal{Q}_\mu^d$, are indexed by partitions μ with $\ell(\mu) \leq d$. The dimension of the symmetric group irrep $\mathcal{P}_\mu$ is denoted $\dim(\mu)$, and its corresponding character χ_μ is given by $\chi_\mu(\pi) = \mathrm{tr}[\mathcal{P}(\pi)]$.

The representations $\mathcal{P}(\pi)$ and $\mathcal{Q}(U)$ commute, meaning that $\mathcal{P}(\pi)\mathcal{Q}(U) = \mathcal{Q}(U)\mathcal{P}(\pi)$. Hence, we can consider $\mathcal{P}(\pi)\mathcal{Q}(U)$ as a representation of the direct product group $\mathsf{S}_m \times \mathsf{U}_d$. *Schur-Weyl duality*, stated as follows, establishes a strong connection between these representations.

Theorem 21 (Schur–Weyl duality)**.** *The space* $(\mathbb{C}^d)^{\otimes m}$ *decomposes as*

$$\mathcal{P}\mathcal{Q} \overset{\mathsf{S}_m \times \mathsf{U}_d}{\cong} \bigoplus_{\substack{\mu \vdash m \\ \ell(\mu) \leq d}} \mathcal{P}_\mu \otimes \mathcal{Q}^d_\mu.$$

In other words, there exist a unitary $U_{\mathrm{Schur}} \in \mathsf{U}_{d^m}$ *such that for all* $\pi \in \mathsf{S}_m$ *and* $U \in \mathsf{U}_d$*,*

$$U_{\mathrm{Schur}} \mathcal{P}(\pi) \mathcal{Q}(U) U_{\mathrm{Schur}}^\dagger = \sum_{\substack{\mu \vdash m \\ \ell(\mu) \leq d}} |\mu\rangle\langle\mu| \otimes \mathcal{P}_\lambda(\pi) \otimes \mathcal{Q}^d_\lambda(U). \tag{2.5}$$

The unitary operator U_{Schur} transforms the standard basis into a basis that called the *Schur basis* and label by $|\mu\rangle|q_\mu\rangle|p_\mu\rangle$. In this basis,

$$\begin{aligned} \mathcal{Q}(U) \cdot |\mu\rangle|q_\mu\rangle|p_\mu\rangle &= |\mu\rangle(\mathcal{Q}^d_\mu(U)|q_\mu\rangle)|p_\mu\rangle, \\ \mathcal{P}(\pi) \cdot |\mu\rangle|q_\mu\rangle|p_\mu\rangle &= |\mu\rangle|q_\mu\rangle(\mathcal{P}_\mu(\pi)|p_\mu\rangle). \end{aligned}$$

Because $\mathcal{Q}^d_\mu$ is a polynomial irrep, it is well-defined for any $d \times d$ matrix. For example, when applied to invertible matrices it gives the μ-irrep of the general linear group GL_d. We can also apply it to (possibly) non-invertible matrices, like the state ρ. In this case, if we set $\pi = e$ in (2.5), where e is the identity permutation, we see that the operator $\mathcal{Q}(\rho) = \rho^{\otimes m}$ is block-diagonalized in the Schur basis.

Corollary 22. *Given a* $d \times d$ *density operator* ρ*,*

$$U_{\mathrm{Schur}} \rho^{\otimes m} U_{\mathrm{Schur}} = \sum_{\substack{\mu \vdash m \\ \ell(\mu) \leq d}} |\mu\rangle\langle\mu| \otimes \mathbb{1}_{\dim(\mu)} \otimes \mathcal{Q}^d_\mu(\rho). \tag{2.6}$$

The equality (2.6) shows that there is a unitary U_{Schur} *independent* of the state ρ which puts this state in the block-diagonal form. We can therefore interpret the density matrix $U_{\mathrm{Schur}} \rho^{\otimes m} U_{\mathrm{Schur}}$ as corresponding to a mixed state with one element in the mixture for each block μ. In this case, measuring the block μ can be done without loss of generality, as it does not perturb the state. This gives rise to the following measurement.

Definition 23 (Weak Schur sampling)**.** Weak Schur sampling (WSS) *refers to the projective measurement* $\{\Pi_\mu\}_{\mu \vdash m, \ell(\mu) \leq d}$ *in which* Π_μ *projects onto the subspace specified by the partition* μ *in the Schur basis.*

The distribution of μ measured by WSS only depends on the spectrum of the state ρ. In fact, if one is only interested learning some property of ρ's spectrum, it can be shown that WSS is the optimal measurement, and that further measuring within the μ-irrep (e.g. measuring $\mathcal{Q}^d_\mu(\rho)$) yields no additional information about ρ's spectrum.

In the analysis of the MPS tester in Section 2.4, we will use the following expression for the projector Π_μ using the characters χ_μ.

Theorem 24 (Weak Schur sampling projector, cf. [CHW07, Equation 7])**.** *The weak Schur sampling*

projectors Π_μ *can be expressed as*

$$\Pi_\mu = \dim(\mu) \cdot \mathop{\mathbf{E}}_{\boldsymbol{\pi} \in \mathsf{S}_m} [\chi_\mu(\boldsymbol{\pi}) \mathcal{P}(\boldsymbol{\pi})] \,. \tag{2.7}$$

2.4 An algorithm for testing matrix product states

In this section we introduce the MPS tester. To begin, we introduce the rank tester of O'Donnell and Wright [OW15], which is meant to test whether a mixed state ρ is rank r. This refers to the following problem.

Definition 25 (Rank testing)**.** *Given a mixed stated state* $\rho \in \mathbb{C}^{d \times d}$*, let* $\rho = \sum_{i=1}^{d} \alpha_i \cdot |u_i\rangle\langle u_i|$ *be its eigendecomposition, where* $\alpha_1 \geq \cdots \geq \alpha_d$*. Then* ρ *is* δ-far from rank r *if* $\lambda_{r+1} + \cdots + \lambda_d \geq \delta$*.*

An algorithm $\mathcal{A}$ *is a* property tester for rank r matrices using $m = m(r, \delta)$ copies *if, given* $\delta > 0$ *and* m *copies of* $\rho \in \mathbb{C}^{d \times d}$*, it acts as follows. If* ρ *is rank-*r*, then it accepts with probability at least* $\frac{2}{3}$*. (If instead it accepts with probability* 1 *in this case, we say that it has* perfect completeness*.) And if* ρ *is* δ*-far from rank-*r*, then it accepts with probability at most* $\frac{1}{3}$*.*

The rank tester of [OW15] is motivated by the fact that if ρ is indeed rank r, then weak Schur sampling (as in Definition 23) always returns a Young diagram $\boldsymbol{\mu}$ with $\ell(\boldsymbol{\mu}) \leq r$.

Definition 26 (The rank tester)**.** *Let* $r \geq 1$*. Given* $\rho^{\otimes n}$*, the* rank tester *performs weak Schur sampling and receives a random* $\boldsymbol{\mu}$*. The rank tester accepts if* $\ell(\boldsymbol{\mu}) \leq r$ *and rejects otherwise. Equivalently, it performs the two-outcome projective measurement* $\{\Pi_{\leq r}, \mathbb{1} - \Pi_{\leq r}\}$*, where* $\Pi_{\leq r} = \sum_{\mu : \ell(\mu) \leq r} \Pi_\mu$*, and accepts if it observes the first outcome.*

The next theorem states the copy complexity of the rank tester.

Theorem 27 ([OW15, Lemma 6.2])**.** *The rank tester tests whether* ρ *has rank* r *with* $O(r^2/\delta)$ *copies.*

O'Donnell and Wright also show that the rank tester requires $\Omega(r^2/\delta)$ copies [OW15, Lemma 6.2]. The rank tester has perfect completeness, and in fact it is the optimal algorithm for rank testing with perfect completeness [OW15, Proposition 6.1]. However, among algorithms with imperfect completeness, the best known lower bound states that $\Omega(r/\delta)$ are necessary [OW15, Theorem 1.11]. It remains an open question whether the rank tester is indeed the optimal algorithm for this task, or whether it can be improved upon.

Now we state the MPS tester. It is motivated by the fact that $|\psi_{1,\ldots,n}\rangle$ is in $\mathsf{MPS}(r)$ if and only if $\psi_{1,\ldots,i}$ has rank r for each $i \in [n]$.

Definition 28 (The MPS tester)**.** *Given* m *copies of the state* $|\psi_{1,\ldots,n}\rangle \in \mathbb{C}^{d_1} \otimes \ldots \otimes \mathbb{C}^{d_n}$*, the* MPS tester *acts as follows. For all* $i \in [n]$*, it runs the rank tester on* $\psi_{1,\ldots,i}^{\otimes m}$*. It accepts if each of them accepts, and rejects otherwise.*

Equivalently, for each $i \in [n]$*, let* $\mathcal{H}_{1,\ldots i}$ *be the Hilbert space* $\mathcal{H}_{1,\ldots i} = \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_i}$*, and define* $\mathcal{H}_{i+1,\ldots,n}$ *analogously. Let* $\{\Pi_{\leq r,1,\ldots,i}, \mathbb{1} - \Pi_{\leq r,1,\ldots,i}\}$ *be the rank tester's measurement when performed on* $\mathcal{H}_{1,\ldots,i}^{\otimes m}$*. Then the MPS tester performs the two-outcome projective measurement* $\{\Pi_{\mathrm{MPS}}, \mathbb{1} - \Pi_{\mathrm{MPS}}\}$*, where*

$$\Pi_{\mathrm{MPS}} = \prod_{i=1}^{n} (\Pi_{\leq r,1,\ldots,i} \otimes \mathbb{1}_{\mathcal{H}_{i+1,\ldots,n}^{\otimes m}}),$$

and accepts if it observes the first outcome.

Before analyzing the copy complexity of the MPS tester, we first show that it is well-defined. In particular, we will show that the different rank tester measurements commute with each other, which implies that they can be simultaneously measured and that $\{\Pi_{\mathrm{MPS}}, \mathbb{1} - \Pi_{\mathrm{MPS}}\}$ is indeed a two-outcome projective measurement, as claimed in Definition 28. We first prove the following lemma, which shows that two overlapping weak Schur sampling measurements commute.

Lemma 29 (Overlapping weak Schur sampling commutes)**.** *Consider a bipartite system with Hilbert space $\mathcal{H}_L \otimes \mathcal{H}_R$, where $\mathcal{H}_L = \mathbb{C}^{d_L}, \mathcal{H}_R = \mathbb{C}^{d_R}$. Let $\{\Pi_{\lambda,L}\}$ and $\{\Pi_{\mu,LR}\}$ denote the weak Schur sampling measurements when applied to $\mathcal{H}_L^{\otimes m}$ and $\mathcal{H}_{LR}^{\otimes m}$, respectively. Then these two measurements commute, meaning that for any two partitions λ and μ,*

$$(\Pi_{\lambda,L} \otimes \mathbb{1}_{\mathcal{H}_R^{\otimes m}}) \cdot \Pi_{\mu,LR} = \Pi_{\mu,LR} \cdot (\Pi_{\lambda,L} \otimes \mathbb{1}_{\mathcal{H}_R^{\otimes m}}).$$

Proof. Throughout this proof, we will omit the "$\mathbb{1}_{\mathcal{H}_R^{\otimes m}}$" when writing $\Pi_{\lambda,L} \otimes \mathbb{1}_{\mathcal{H}_R^{\otimes m}}$ or $\mathcal{P}_L \otimes \mathbb{1}_{\mathcal{H}_R^{\otimes m}}$, for simplicity.

First, we note that $\mathcal{P}_L(\pi) \cdot \mathcal{P}_{LR}(\sigma) = \mathcal{P}_{LR}(\sigma)\mathcal{P}_L(\sigma^{-1}\pi\sigma)$. To show this, let $|\ell_1\rangle, \ldots, |\ell_m\rangle$ be m standard basis vectors in $\mathcal{H}_L$, and let $|r_1\rangle, \ldots, |r_m\rangle$ be m standard basis vectors in $\mathcal{H}_R$. Then each $|\ell_i\rangle \otimes |r_i\rangle$ is a standard basis vector in $\mathcal{H}_L \otimes \mathcal{H}_R$, and so

$$\begin{aligned}
&\mathcal{P}_L(\pi) \cdot \mathcal{P}_{LR}(\sigma) \cdot |\ell_1 \cdots \ell_m\rangle \otimes |r_1 \cdots r_m\rangle \\
&= \mathcal{P}_L(\pi) \cdot |\ell_{\sigma^{-1}(1)} \cdots \ell_{\sigma^{-1}(m)}\rangle \otimes |r_{\sigma^{-1}(1)} \cdots r_{\sigma^{-1}(m)}\rangle \\
&= |\ell_{\sigma^{-1}(\pi^{-1}(1))} \cdots \ell_{\sigma^{-1}(\pi^{-1}(m))}\rangle \otimes |r_{\sigma^{-1}(1)} \cdots r_{\sigma^{-1}(m)}\rangle \\
&= \mathcal{P}_{LR}(\sigma) \cdot |\ell_{\sigma^{-1}(\pi^{-1}(\sigma(1)))} \cdots \ell_{\sigma^{-1}(\pi^{-1}(\sigma(m)))}\rangle \otimes |r_1 \cdots r_m\rangle \\
&= \mathcal{P}_{LR}(\sigma) \cdot \mathcal{P}_L(\sigma^{-1}\pi\sigma) \cdot |\ell_1 \cdots \ell_m\rangle \otimes |r_1 \cdots r_m\rangle.
\end{aligned}$$

Extending this to all of $\mathcal{H}_L \otimes \mathcal{H}_R$ via linearity proves the equality. Next, we note that $\chi_\lambda(\sigma^{-1}\pi\sigma) = \chi_\lambda(\pi)$ because $\chi_\lambda(\cdot)$ is a class function. Putting these together, we have

$$\begin{aligned}
\Pi_{\lambda,L}\Pi_{\mu,LR} &= \dim(\lambda)\dim(\mu) \cdot \mathop{\mathbf{E}}_{\boldsymbol{\pi},\boldsymbol{\sigma} \sim \mathsf{S}_m} [\chi_\lambda(\boldsymbol{\pi})\chi_\mu(\boldsymbol{\sigma})\mathcal{P}_L(\boldsymbol{\pi})\mathcal{P}_{LR}(\boldsymbol{\sigma})] \\
&= \dim(\lambda)\dim(\mu) \cdot \mathop{\mathbf{E}}_{\boldsymbol{\pi},\boldsymbol{\sigma} \sim \mathsf{S}_m} \left[\chi_\lambda(\boldsymbol{\sigma}^{-1}\boldsymbol{\pi}\boldsymbol{\sigma})\chi_\mu(\boldsymbol{\sigma})\mathcal{P}_{LR}(\boldsymbol{\sigma})\mathcal{P}_L(\boldsymbol{\sigma}^{-1}\boldsymbol{\pi}\boldsymbol{\sigma})\right] \\
&= \dim(\lambda)\dim(\mu) \cdot \mathop{\mathbf{E}}_{\boldsymbol{\pi},\boldsymbol{\sigma} \sim \mathsf{S}_m} [\chi_\lambda(\boldsymbol{\pi})\chi_\mu(\boldsymbol{\sigma})\mathcal{P}_{LR}(\boldsymbol{\sigma})\mathcal{P}_L(\boldsymbol{\pi})] \\
&= \Pi_{\mu,LR}\Pi_{\lambda,L},
\end{aligned}$$

where the third line uses the fact that $\boldsymbol{\sigma}^{-1}\boldsymbol{\pi}\boldsymbol{\sigma}$ is distributed as a uniformly random element of S_m, even conditioned on the value of $\boldsymbol{\sigma}$. This completes the proof. □

As an immediate corollary, we get that the MPS tester is well-defined.

Proposition 30 (The MPS tester is well-defined)**.** *The matrices*

$$\Pi_{\leq r,1,\ldots,i} \otimes \mathbb{1}_{\mathcal{H}_{i+1,\ldots,n}^{\otimes m}}$$

commute for all $i \in [n]$. As a result, the MPS tester measurement $\{\Pi_{\mathrm{MPS}}, \mathbb{1} - \Pi_{\mathrm{MPS}}\}$ is a two-outcome projective measurement.

Now we analyze the copy complexity of the MPS tester. Because it runs a separate rank tester on each cut of $|\psi_{1,\dots,n}\rangle$ simultaneously, the outcome of one rank tester can affect the rank of the remaining cuts, and therefore the outcomes of the remaining rank testers. This complicates the analysis of this collective set of measurements. Instead, we will do a pessimistic analysis and just show that the MPS tester does well on at least one cut. This analysis uses the following proposition.

Proposition 31 (Far from MPS implies a cut is far from low-rank)**.** *Suppose $|\psi_{1,\dots,n}\rangle$ is δ-far from $\mathsf{MPS}(r)$. Then there exists an $i \in [n-1]$ such that $\psi_{1,\dots,i}$ is $(\delta^2/2n)$-far from rank-r.*

Proof. $\mathrm{Dist}_r(|\psi_{1,\dots,n}\rangle) \geq \delta$ implies $\mathrm{Overlap}_r(|\psi_{1,\dots,n}\rangle) \leq 1 - \delta^2$. By Lemma 19, there exists a state $|\phi\rangle \in \mathsf{MPS}(r)$ such that

$$|\langle\phi|\psi\rangle| \geq 1 - \sum_{i=1}^{n-1} \sum_{j=r+1}^{D_i} \lambda_j^{(i)}.$$

Then $|\langle\phi|\psi\rangle| \leq \sqrt{1-\delta^2} \leq 1 - \delta^2/2$. Rearranging, we have

$$\delta^2/2 \leq 1 - \sqrt{1-\delta^2} \leq 1 - |\langle\phi|\psi\rangle| \leq \sum_{i=1}^{n-1} \sum_{j=r+1}^{D_i} \lambda_j^{(i)} \leq n \cdot \max_{i\in[n-1]} \left\{ \sum_{j=r+1}^{D_i} \lambda_j^{(i)} \right\}.$$

Letting i be the maximizing coordinate, this implies that $\psi_{1,\dots,i}$ is $(\delta^2/2n)$-far from rank-r, which completes the proof. □

There are two ways that Proposition 31 "loses" in going from $|\psi_{1,\dots,n}\rangle$ being δ-far to $\psi_{1,\dots,i}$ being $(\delta^2/2n)$-far. The first is the factor of $1/n$ which is unavoidable since we are ignoring all but one cut. The second "loss" is the fact that δ is squared in the conclusion. However, this turns out to just be a quirk in the different ways we measure distance to MPS and distance to rank-r. For example, even for a bipartite state $|\psi_{1,2}\rangle$, Lemma 18 tells us that $|\psi_{1,2}\rangle$ is δ-far from $\mathsf{MPS}(r)$ if and only if ψ_1 is δ^2-far from rank-r.

Now we prove the main theorem of this section.

Theorem 32. *Given $m = O(nr^2/\delta^2)$ copies of a state $|\psi_{1,\dots,n}\rangle \in \mathbb{C}^{d_1} \otimes \dots \otimes \mathbb{C}^{d_n}$, the MPS tester tests whether $|\psi\rangle$ is in $\mathsf{MPS}(r)$ with perfect completeness.*

Proof. If $|\psi_{1,\dots,n}\rangle$ is in $\mathsf{MPS}(r)$, then $\psi_{1,\dots,i}$ is rank-r for each $i \in [n]$. As a result, the rank tester applied to each cut always accepts because the rank tester has perfect completeness, and so the MPS tester always accepts. On the other hand, if ρ is δ-far from $\mathsf{MPS}(r)$, then Proposition 31 implies there exists an $i \in [n-1]$ such that $\psi_{1,\dots,i}$ is $\delta' = (\delta^2/2n)$-far from rank r. The probability the MPS tester accepts $|\psi_{1,\dots,n}\rangle$ is at most the probability the rank tester accepts $\psi_{1,\dots,i}$, and since we are using $O(nr^2/\delta^2) = O(r^2/\delta')$ copies, the rank tester will accept with probability at most $\frac{1}{3}$. Thus, the MPS tester tests whether $|\psi_{1,\dots,n}\rangle$ is in $\mathsf{MPS}(r)$, and this completes the proof. □

2.5 A lower bound for testing matrix product states

We now derive a lower bound on the sample complexity of testing whether a state is in $\mathsf{MPS}(r)$, for $r \geq 2$. Let $d \geq 0$ satisfy $d-1 \geq 2 \cdot (r-1)$, and consider the bipartite state $|\varphi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ defined as

$$|\varphi\rangle = \sqrt{1-\theta} \cdot |1\rangle|1\rangle + \sum_{i=2}^{d} \sqrt{\frac{\theta}{d-1}} \cdot |i\rangle|i\rangle,$$

where $0 \leq \theta \leq 1$ is a parameter that we set later. By the Young-Eckart Theorem (Lemma 18),

$$\mathrm{Overlap}_r(|\varphi\rangle) = (1-\theta) + (r-1) \cdot \frac{\theta}{d-1}.$$

Let n be an even integer, and define

$$|\Phi_n\rangle = |\varphi\rangle^{\otimes \frac{n}{2}}.$$

To compute the overlap of $|\Phi_n\rangle$ with $\mathsf{MPS}(r)$, we use the following proposition.

Proposition 33 (Overlap of tensor products)**.** *Let $|\varphi\rangle \in \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_k}$ be a k-partite state with $\mathrm{Overlap}_r(|\varphi\rangle) = \omega$. Then for each $\ell \geq 1$, $\mathrm{Overlap}_r(|\varphi\rangle^{\otimes \ell}) = \omega^\ell$.*

Proof. Let $|\psi\rangle$ be the state in $\mathsf{MPS}_k(r)$ with $|\langle\psi|\varphi\rangle|^2 = \omega$ guaranteed by the assumption. Then $|\psi\rangle^{\otimes \ell}$ is in $\mathsf{MPS}_{\ell k}(r)$ and has $|\langle\psi|^{\otimes \ell} \cdot |\varphi\rangle^{\otimes \ell}|^2 = |\langle\psi|\varphi\rangle|^{2\ell} = \omega^\ell$. This proves the lower-bound $\mathrm{Overlap}_{\ell k}(|\psi\rangle^{\otimes \ell}) \geq \omega^\ell$,

As for the upper-bound, the proof is by induction on ℓ, the base case being trivial. For the inductive step, write $|\varphi^\ell\rangle$ as shorthand for $|\varphi\rangle^{\otimes \ell}$, and suppose that the inductive hypothesis holds for $|\varphi^\ell\rangle$, i.e. that $\mathrm{Overlap}_r(|\varphi^\ell\rangle) \leq \omega^\ell$. Then we show that

$$\forall |\beta\rangle \in \mathsf{MPS}_{(\ell+1)k}(r), \quad |\langle\varphi^{\ell+1}|\beta\rangle|^2 \leq \omega^{\ell+1}.$$

Since $|\varphi^{\ell+1}\rangle = |\varphi^\ell\rangle \otimes |\varphi\rangle$, this is equivalent to proving that for all $|\beta\rangle \in \mathsf{MPS}_{(\ell+1)k}(r)$,

$$\begin{aligned} \langle\varphi^{\ell+1}|\beta\rangle\langle\beta|\varphi^{\ell+1}\rangle &= \langle\varphi| \cdot (\langle\varphi^\ell| \otimes \mathbb{1}) \cdot |\beta\rangle\langle\beta| \cdot (|\varphi^\ell\rangle \otimes \mathbb{1}) \cdot |\varphi\rangle \\ &= \||\Gamma\rangle\|^2 \cdot |\langle\varphi|\widetilde{\Gamma}\rangle|^2 \leq \omega^{\ell+1}, \end{aligned} \tag{2.8}$$

where we define

$$|\Gamma\rangle = (\langle\varphi^\ell| \otimes \mathbb{1}) \cdot |\beta\rangle, \quad |\widetilde{\Gamma}\rangle = \frac{|\Gamma\rangle}{\||\Gamma\rangle\|}. \tag{2.9}$$

From here, our proof of (2.8) breaks into two steps: step 1, showing that $\||\Gamma\rangle\|^2 \leq \omega^\ell$, and step 2, showing that $|\langle\varphi|\widetilde{\Gamma}\rangle|^2 \leq \omega$. We begin with the former.

Step 1: bounding $\||\Gamma\rangle\|^2$. Let the Schmidt decomposition of the state $|\beta\rangle \in \mathsf{MPS}_{(\ell+1)k}(r)$ across the subsystems $\{1, \ldots, \ell k\}$ and $\{\ell k + 1, \cdots, (\ell+1)k\}$ be $|\beta\rangle = \sum_{i=1}^r \sqrt{\mu_i}|e_i\rangle|f_i\rangle$. Then $|\Gamma\rangle = \sum_{i=1}^r \sqrt{\mu_i}\langle\varphi^\ell|e_i\rangle|f_i\rangle$, and

$$\||\Gamma\rangle\|^2 = \sum_{i=1}^{r} \mu_i |\langle\varphi^\ell|e_i\rangle|^2.$$

Suppose it holds that $|e_i\rangle$ is in $\mathsf{MPS}_{\ell k}(r)$, for each i. Then the inductive hypothesis implies that

$$\||\Gamma\rangle\|^2 \leq \sum_{i=1}^{r} \mu_i \omega^\ell = \omega^\ell,$$

which is the desired bound on $\||\Gamma\rangle\|^2$. It remains to show that $|e_i\rangle \in \mathsf{MPS}_{\ell k}(r)$. Consider partitioning the $(\ell+1)k$ subsystems into $\{1,\ldots,q\}$ and $\{q+1,\ldots,(\ell+1)k\}$ for any integer $1 \leq q \leq \ell k$. We have

$$\begin{aligned}
\mathrm{tr}_{q+1,\ldots,(\ell+1)k} |\beta\rangle\langle\beta| &= \mathrm{tr}_{q+1,\ldots,(\ell+1)k} \left[\sum_{i,j=1}^{r} \sqrt{\mu_i \mu_j} |e_i\rangle\langle e_j| \otimes |f_i\rangle\langle f_j| \right] \\
&= \sum_{i,j=1}^{r} \sqrt{\mu_i \mu_j} \cdot \mathrm{tr}_{q+1,\ldots,(k+1)\ell} \left[|e_i\rangle\langle e_j| \otimes |f_i\rangle\langle f_j| \right] \\
&= \sum_{i=1}^{r} \mu_i \cdot \mathrm{tr}_{q+1,\ldots,k\ell} |e_i\rangle\langle e_i|. \qquad (2.10)
\end{aligned}$$

Equation (2.10) is a sum of PSD operators and rank does not decrease by adding such operators. Since $|\beta\rangle \in \mathsf{MPS}_{(\ell+1)k}(r)$, we have $\mathrm{tr}_{q+1,\ldots,(\ell+1)k} |\beta\rangle\langle\beta| \leq r$. For this to happen, the rank of each of the terms in Equation (2.10) must also be $\leq r$. This implies $|e_i\rangle \in \mathsf{MPS}_{\ell k}(r)$, as we claimed.

Step 2: bounding $|\langle\varphi|\widetilde{\Gamma}\rangle|^2$. By the base case of the induction, $\mathrm{Overlap}_r(|\varphi\rangle) \leq \omega$. Thus, to complete step 2, it is sufficient to show that $|\widetilde{\Gamma}\rangle$ is in $\mathsf{MPS}_k(r)$. Let $D = (d_1 \cdots d_k)^\ell$, and let $|A_1\rangle, |A_2\rangle, \ldots, |A_D\rangle$ be an orthonormal basis for the first ℓk qudits such that $|A_1\rangle = |\varphi^\ell\rangle$. In addition, let $i \in \{1,\ldots,k\}$. By tracing out the first $\ell k + i$ qudits in the state $|\beta\rangle \in \mathsf{MPS}_{(\ell+1)k}(r)$ we get

$$\mathrm{tr}_{1,\ldots,\ell k+i} |\beta\rangle\langle\beta| = \mathrm{tr}_{\ell k+1,\ldots,\ell k+i} \left[\sum_{i \in [D]} (\langle A_i| \otimes \mathbb{1}) \cdot |\beta\rangle\langle\beta| \cdot (|A_i\rangle \otimes \mathbb{1}) \right]. \qquad (2.11)$$

By the definition of $|\Gamma\rangle$ in (2.9) and our choice of the state $|A_1\rangle$, the $i = 1$ part of this sum is

$$\mathrm{tr}_{\ell k+1,\ldots,\ell k+i} \left[(\langle\varphi^\ell| \otimes \mathbb{1}) \cdot |\beta\rangle\langle\beta| \cdot (|\varphi^\ell\rangle \otimes \mathbb{1}) \right] = \mathrm{tr}_{1,\ldots,i} |\Gamma\rangle\langle\Gamma|.$$

Since $|\beta\rangle$ is in $\mathsf{MPS}_{(\ell+1)k}(r)$, $\mathrm{tr}_{1,\ldots,\ell k+i} |\beta\rangle\langle\beta|$ has rank at most r. But Equation (2.11) is a sum of PSD operators, so in order to have $\mathrm{tr}_{\ell k+i} |\beta\rangle\langle\beta|$ be rank $\leq r$, the $i = 1$ part of the sum in Equation (2.11) must also be rank $\leq r$. This implies that $\mathrm{tr}_{1,\ldots,i} |\Gamma\rangle\langle\Gamma|$, and therefore also $\mathrm{tr}_{1,\ldots,i} |\widetilde{\Gamma}\rangle\langle\widetilde{\Gamma}|$, is rank r. As a result, $|\widetilde{\Gamma}\rangle \in \mathsf{MPS}_k(r)$, which concludes the proof. □

Applying Proposition 33 to $|\Phi_n\rangle$, we can compute its overlap as

$$\mathrm{Overlap}_r(|\Phi_n\rangle) = \left((1-\theta) + (r-1) \cdot \frac{\theta}{d-1} \right)^{n/2} \leq \left(1 - \frac{\theta}{2} \right)^{n/2},$$

where the first inequality uses $d - 1 \geq 2 \cdot (r-1)$. Now if we pick θ to be

$$\theta = \frac{8\delta^2}{n}, \tag{2.12}$$

we get

$$\left(1 - \frac{4\delta^2}{n}\right)^{n/2} \leq 1 - \delta^2,$$

for $\delta \leq \frac{1}{\sqrt{2}}$, where we have used the inequality $(1-x)^n \leq 1 - \frac{1}{2}xn$ for $x \leq \frac{1}{n}$. As a result, the distance of $|\Phi_n\rangle$ to $\mathsf{MPS}(r)$ is

$$\mathrm{Dist}_r(|\Phi_n\rangle) = \sqrt{1 - \mathrm{Overlap}_r(|\Phi_n\rangle)} \geq \delta.$$

Therefore, $|\Phi_n\rangle$ is far from $\mathsf{MPS}(r)$, and any $\mathsf{MPS}(r)$ testing algorithm should reject it with probability at least $\frac{2}{3}$. (Note that because $0 \leq \theta \leq 1$, we must have $8\delta^2/n \leq 1$, which is satisfied if $n \geq 4$.)

Our hard family of states which are far from $\mathsf{MPS}(r)$ will consist of $|\Phi_n\rangle$ and any state which can be computed from $|\Phi_n\rangle$ by a local unitary. To make this formal, consider the ensemble of pure states in which a random element is sampled as follows: first, sample $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_{n/2}, \boldsymbol{V}_1, \ldots, \boldsymbol{V}_{n/2} \sim \mathsf{U}_d$, i.e. n Haar random $d \times d$ unitary matrices, and output

$$(\boldsymbol{U}_1 \otimes \boldsymbol{V}_1) \otimes \cdots \otimes (\boldsymbol{U}_{n/2} \otimes \boldsymbol{V}_{n/2}) \cdot |\Phi_n\rangle.$$

Local unitaries do not affect the distance to $\mathsf{MPS}(r)$, and so each state in this ensemble is distance δ from $\mathsf{MPS}(r)$. Thus, if a tester is given m copies of any of these states, it should reject with probability at least $\frac{2}{3}$. As a result, it should also reject with probability at least $\frac{2}{3}$ if given the density matrix

$$\rho_{\mathrm{far}} = \mathbf{E}\left((\boldsymbol{U}_1 \otimes \boldsymbol{V}_1) \otimes \cdots \otimes (\boldsymbol{U}_{n/2} \otimes \boldsymbol{V}_{n/2}) \cdot |\Phi_n\rangle\langle\Phi_n| \cdot (\boldsymbol{U}_1^\dagger \otimes \boldsymbol{V}_1^\dagger) \otimes \cdots \otimes (\boldsymbol{U}_{n/2}^\dagger \otimes \boldsymbol{V}_{n/2}^\dagger)\right)^{\otimes m}$$

corresponding to m copies of a random state drawn from this ensemble. We will show that this is difficult for an $\mathsf{MPS}(r)$ tester unless m is sufficiently large. To do this, we will show that there exists another density matrix ρ_{MPS} corresponding to a mixture over states in $\mathsf{MPS}(r)$ such that the trace distance between ρ_{MPS} and ρ_{far} is small unless m is sufficiently large. To define ρ_{MPS}, let us first define the state

$$|\gamma\rangle = \sqrt{1-\theta} \cdot |1\rangle|1\rangle + \sum_{i=2}^{r} \sqrt{\frac{\theta}{r-1}} \cdot |i\rangle|i\rangle,$$

and the state $|\Gamma_n\rangle = |\gamma\rangle^{\otimes n/2}$. The state $|\gamma\rangle$ is an element of $\mathsf{MPS}(r)$, and therefore so is $|\Gamma_n\rangle$. Then we define

$$\rho_{\mathrm{MPS}} = \mathbf{E}\left((\boldsymbol{U}_1 \otimes \boldsymbol{V}_1) \otimes \cdots \otimes (\boldsymbol{U}_{n/2} \otimes \boldsymbol{V}_{n/2}) \cdot |\Gamma_n\rangle\langle\Gamma_n| \cdot (\boldsymbol{U}_1^\dagger \otimes \boldsymbol{V}_1^\dagger) \otimes \cdots \otimes (\boldsymbol{U}_{n/2}^\dagger \otimes \boldsymbol{V}_{n/2}^\dagger)\right)^{\otimes m}.$$

Each state in this ensemble is in $\mathsf{MPS}(r)$, and so if a tester is given this density matrix, it should accept with probability at least $\frac{2}{3}$. Our main result is as follows.

Theorem 34. *(Lower bound on copy complexity of MPS testing) Suppose there is an algorithm*

that accepts ρ_{MPS} *with probability at least* $\frac{2}{3}$ *and accepts* ρ_{far} *with probability at most* $\frac{1}{3}$*. Then* $m = \Omega(\sqrt{n}/\delta^2)$*.*

As a result, $\Omega(\sqrt{n}/\delta^2)$ *copies are necessary to test whether a state is in* $\mathsf{MPS}(r)$ *for* $\delta \leq \frac{1}{\sqrt{2}}$*.*

Proof. Our goal is to bound $\mathrm{D}_{\mathrm{tr}}(\rho_{\mathrm{far}}, \rho_{\mathrm{MPS}})$. To do so, it is convenient to also work with the fidelity of these states. Recall that the fidelity $F(\alpha, \beta)$ of two mixed states α, β is defined by $F(\alpha, \beta) = \|\sqrt{\alpha}\sqrt{\beta}\|_1$. One useful property of this measure is that it is multiplicative with respect to tensor products, i.e. $F(\alpha_1 \otimes \alpha_2, \beta_1 \otimes \beta_2) = F(\alpha_1, \beta_1)F(\alpha_2, \beta_2)$. Another is the bound

$$1 - F(\alpha, \beta) \leq \mathrm{D}_{\mathrm{tr}}(\alpha, \beta) \leq \sqrt{1 - F(\alpha, \beta)^2} \tag{2.13}$$

between the trace distance $\mathrm{D}_{\mathrm{tr}}(\alpha, \beta)$ and the fidelity $F(\alpha, \beta)$, which we can use to switch back and forth between these two measures.

We begin by applying the upper-bound in (2.13) to switch to fidelity:

$$\mathrm{D}_{\mathrm{tr}}(\rho_{\mathrm{far}}, \rho_{\mathrm{MPS}}) \leq \sqrt{1 - F(\rho_{\mathrm{far}}, \rho_{\mathrm{MPS}})^2}.$$

Hence, to upper-bound the trace distance between two states, it is sufficient to lower-bound their fidelity. We note that since $|\Phi_n\rangle = |\varphi\rangle^{\otimes n/2}$, we can rewrite the state ρ_{far} as

$$\rho_{\mathrm{far}} = \left(\mathop{\mathbf{E}}_{\boldsymbol{U}, \boldsymbol{V} \sim \mathsf{U}_d} (\boldsymbol{U} \otimes \boldsymbol{V} \cdot |\varphi\rangle\langle\varphi| \cdot \boldsymbol{U}^\dagger \otimes \boldsymbol{V}^\dagger)^{\otimes m} \right)^{\otimes n/2} =: \sigma_{\mathrm{far}}^{\otimes n/2}.$$

By similar reasoning, we can rewrite ρ_{MPS} as

$$\rho_{\mathrm{MPS}} = \left(\mathop{\mathbf{E}}_{\boldsymbol{U}, \boldsymbol{V} \sim \mathsf{U}_d} (\boldsymbol{U} \otimes \boldsymbol{V} \cdot |\gamma\rangle\langle\gamma| \cdot \boldsymbol{U}^\dagger \otimes \boldsymbol{V}^\dagger)^{\otimes m} \right)^{\otimes n/2} =: \sigma_{\mathrm{MPS}}^{\otimes n/2}$$

Hence, by the multiplicativity of fidelity, we have

$$F(\rho_{\mathrm{far}}, \rho_{\mathrm{MPS}}) = F(\sigma_{\mathrm{far}}, \sigma_{\mathrm{MPS}})^{n/2}.$$

Now, by applying (2.13) again to switch back to trace distance, we have

$$F(\sigma_{\mathrm{far}}, \sigma_{\mathrm{MPS}}) \geq 1 - \mathrm{D}_{\mathrm{tr}}(\sigma_{\mathrm{far}}, \sigma_{\mathrm{MPS}}).$$

As a result, we would like to upper-bound the trace distance of σ_{far} and σ_{MPS}.

Consider an algorithm trying to distinguish these two states. For $i \in \{1, \ldots, d\}$, let $|\boldsymbol{a}_i\rangle = \boldsymbol{U}|i\rangle$ and let $|\boldsymbol{b}_i\rangle = \boldsymbol{V}|i\rangle$. Then when the algorithm is given σ_{far}, we can equivalently view it as the algorithm being given m copies of the random sample

$$\sqrt{1-\theta} \cdot |\boldsymbol{a}_1\rangle|\boldsymbol{b}_1\rangle + \sum_{i=2}^{d} \sqrt{\frac{\theta}{d-1}} \cdot |\boldsymbol{a}_i\rangle|\boldsymbol{b}_i\rangle,$$

and when it is given σ_{MPS}, we can equivalently view it as being given m copies of the random sample

$$\sqrt{1-\theta} \cdot |\boldsymbol{a}_1\rangle|\boldsymbol{b}_1\rangle + \sum_{i=2}^{r} \sqrt{\frac{\theta}{r-1}} \cdot |\boldsymbol{a}_i\rangle|\boldsymbol{b}_i\rangle.$$

The only difference between these two mixtures is whether the state has Schmidt coefficients $1-\theta, \theta/(d-1), \ldots, \theta/(d-1)$ or Schmidt coefficients $1-\theta, \theta/(r-1), \ldots, \theta/(r-1)$. As we show in Theorem 35, this means that the algorithm learns everything it needs to learn about which case it is in simply by measuring the m $|\boldsymbol{a}_i\rangle$ registers, and it can ignore the m $|\boldsymbol{b}_i\rangle$ registers. In other words, if we set

$$\tau_{\text{far}} = \operatorname{tr}_2 |\varphi\rangle\langle\varphi| = (1-\theta)\cdot|1\rangle\langle 1| + \theta \cdot \sum_{i=2}^{d} \frac{1}{d-1}\cdot|i\rangle\langle i|$$

and

$$\tau_{\text{MPS}} = \operatorname{tr}_2 |\gamma\rangle\langle\gamma| = (1-\theta)\cdot|1\rangle\langle 1| + \theta \cdot \sum_{i=2}^{r} \frac{1}{r-1}\cdot|i\rangle\langle i|,$$

then

$$\mathrm{D}_{\text{tr}}(\sigma_{\text{far}}, \sigma_{\text{MPS}}) = \mathrm{D}_{\text{tr}}\left(\mathop{\mathbf{E}}_{\boldsymbol{U}\sim \mathsf{U}_d} (\boldsymbol{U}\tau_{\text{far}}\boldsymbol{U}^\dagger)^{\otimes m}, \mathop{\mathbf{E}}_{\boldsymbol{U}\sim \mathsf{U}_d} (\boldsymbol{U}\tau_{\text{MPS}}\boldsymbol{U}^\dagger)^{\otimes m}\right).$$

The density matrix $\mathbf{E}_{\boldsymbol{U}\sim \mathsf{U}_d}(\boldsymbol{U}\tau_{\text{far}}\boldsymbol{U}^\dagger)^{\otimes m}$ can be described by the following mixture. Let $|\boldsymbol{a}_1\rangle, \ldots, |\boldsymbol{a}_d\rangle$ be a random orthonormal basis for $\mathbb{C}^d$ as above. Draw m samples as follows.

(i) With probability $1-\theta$, output $|\boldsymbol{a}_1\rangle$.

(ii) With probability θ, output one of the states $|\boldsymbol{a}_2\rangle, \ldots, |\boldsymbol{a}_d\rangle$ uniformly at random.

The state $\mathbf{E}_{\boldsymbol{U}\sim \mathsf{U}_d}(\boldsymbol{U}\tau_{\text{MPS}}\boldsymbol{U}^\dagger)^{\otimes m}$ can be described by a similar mixture except that now in step (ii), with probability θ, the output is one of the states $|\boldsymbol{a}_2\rangle, \ldots, |\boldsymbol{a}_r\rangle$ chosen uniformly at random. Consider the event that either all the m draws are from step (i) or $m-1$ draws are from step (i) and the remaining sample is from step (ii). The probability of this event occurring is simply $(1-\theta)^m + m\,\theta(1-\theta)^{m-1}$. In both of these cases, it is not possible to distinguish the two states. In all the other cases, where more than one sample is drawn according to step (ii), we loosely upper bound the distance between the states by 1. This gives us the following overall upper bound on the distance between the random ensembles:

$$\begin{aligned}
\mathrm{D}_{\text{tr}}\left(\mathop{\mathbf{E}}_{\boldsymbol{U}\sim \mathsf{U}_d} (\boldsymbol{U}\tau_{\text{far}}\boldsymbol{U}^\dagger)^{\otimes m}, \mathop{\mathbf{E}}_{\boldsymbol{U}\sim \mathsf{U}_d} (\boldsymbol{U}\tau_{\text{MPS}}\boldsymbol{U}^\dagger)^{\otimes m}\right) &\leq 1-(1-\theta)^m - m\theta(1-\theta)^{m-1} \\
&\leq 1-(1-m\theta) - m\theta\,(1-(m-1)\theta) \\
&= m(m-1)\theta^2.
\end{aligned}$$

As a result, this implies that ρ_{far} and ρ_{MPS} have distance

$$\begin{aligned}
\mathrm{D}_{\text{tr}}(\rho_{\text{far}}, \rho_{\text{MPS}}) &\leq \left(1-\left(1-m(m-1)\theta^2\right)^n\right)^{1/2} \\
&\leq \left(1-\left(1-n\cdot m(m-1)\theta^2\right)\right)^{1/2} \\
&\leq \sqrt{n}m\theta.
\end{aligned}$$

By our choice of $\theta = 8\delta^2/n$ in Equation (2.12), this is at most $4m\delta^2/\sqrt{n}$. For an algorithm to accept ρ_{MPS} with probability at least $\frac{2}{3}$ and ρ_{far} with probability at most $\frac{1}{3}$, this trace distance must be at least $\frac{1}{3}$. This implies that m must be at least $\frac{1}{24}\sqrt{n}/\delta^2$, which completes the proof. □

Here we prove the claim in the proof of Theorem 34 that it suffices for any algorithm that tries to distinguish between the states σ_{MPS} and σ_{far} to only measure their m $|\boldsymbol{a}_i\rangle$ registers. The proof is standard and based on repeated applications of Schur's Lemma.

Theorem 35. *Let $|\varphi\rangle_{AB}$ and $|\gamma\rangle_{AB}$ be two bipartite states on subsystems A and B. Any algorithm for distinguishing between the two mixed states*

$$\mathop{\mathbf{E}}_{\boldsymbol{U}_A\sim\mathsf{U}_d,\boldsymbol{V}_B\sim\mathsf{U}_d}\left(\boldsymbol{U}_A\otimes\boldsymbol{V}_B\cdot|\varphi\rangle\langle\varphi|\cdot\boldsymbol{U}_A^\dagger\otimes\boldsymbol{V}_B^\dagger\right)^{\otimes m} \tag{2.14}$$

and

$$\mathop{\mathbf{E}}_{\boldsymbol{U}_A\sim\mathsf{U}_d,\boldsymbol{V}_B\sim\mathsf{U}_d}\left(\boldsymbol{U}_A\otimes\boldsymbol{V}_B\cdot|\gamma\rangle\langle\gamma|\cdot\boldsymbol{U}_A^\dagger\otimes\boldsymbol{V}_B^\dagger\right)^{\otimes m} \tag{2.15}$$

can without loss of generality leave out the m B registers and only measure the m A registers.

We denote the Schur-Weyl basis for $(\mathbb{C}^d)^{\otimes m}$ (see Theorem 21) by $|\mu\rangle|q\rangle|p\rangle$, where q is a basis vector for the μ-irrep $\mathcal{Q}_\mu^d$ of U_d and p is a basis vector for the μ-irrep $\mathcal{P}_\mu$ of S_d. One technicality is that although any orthonormal basis $\{|q\rangle\}_q$ of $\mathcal{Q}_\mu^d$ will suffice for our purposes, we will need to pick a basis $\{|p\rangle\}_p$ of $\mathcal{P}_\mu$ such that the matrix entries of $\mathcal{P}_\mu(\pi)$ are real-valued for each $\pi\in\mathsf{S}_d$. (This is used to establish Equation (2.20) below.) One basis that satisfies this property is known as the *Gelfand-Tsetlin basis*, and the resulting matrices $\{\mathcal{P}_\mu(\pi)\}_{\pi\in\mathsf{S}_d}$ give rise to *Young's orthogonal representation*. In this basis, the matrix elements $\mathcal{P}_\mu(\pi)_{p,p'}:=\langle p|\mathcal{P}_\mu(\pi)|p'\rangle$ are real-valued, and so each matrix $\mathcal{P}_\mu(\pi)$ is an orthogonal matrix. For an introduction to the Gelfand-Tsetlin basis, see [HGG09, Appendix B] and the citations contained therein.

Before proving Theorem 35, we show some helper lemmas.

Lemma 36. *Let $\mathcal{H}=(\mathbb{C}^d)^{\otimes m}$, and let $\mathcal{H}'$ be another Hilbert space. Consider a matrix N acting on $\mathcal{H}\otimes\mathcal{H}'$ of the form*

$$N=\sum_{\substack{\mu,\mu'\\q,q'}}|\mu\rangle\langle\mu'|\otimes|q\rangle\langle q'|\otimes N_{\mu,\mu',q,q'}, \tag{2.16}$$

where $N_{\mu,\mu',q,q'}$ is an operator acting on $\mathcal{H}'$ and the $|p\rangle$ register of $\mathcal{H}$. Then it holds that

$$\mathop{\mathbf{E}}_{\boldsymbol{U}\sim\mathsf{U}_d}\left(\boldsymbol{U}^{\otimes m}\cdot N\cdot(\boldsymbol{U}^\dagger)^{\otimes m}\right)=\sum_\mu|\mu\rangle\langle\mu|\otimes\mathbb{1}\otimes N_\mu, \tag{2.17}$$

where N_μ is an operator acting on $\mathcal{H}'$ and the $|p\rangle$ register of $\mathcal{H}$.

Proof. To begin, we calculate

$$\mathop{\mathbf{E}}_{\boldsymbol{U}\sim\mathsf{U}_d}\left(\boldsymbol{U}^{\otimes m}\cdot N\cdot(\boldsymbol{U}^\dagger)^{\otimes m}\right)=\sum_{\substack{\mu,\mu'\\q,q'}}|\mu\rangle\langle\mu'|\otimes\left(\mathop{\mathbf{E}}_{\boldsymbol{U}\sim\mathsf{U}_d}\mathcal{Q}_\mu^d(\boldsymbol{U})\cdot|q\rangle\langle q'|\cdot\mathcal{Q}_{\mu'}^d(\boldsymbol{U})^\dagger\right)\otimes N_{\mu,\mu',q,q'}. \tag{2.18}$$

For each μ,μ',q,q', the matrix

$$T_{\mu,\mu',q,q'}=\mathop{\mathbf{E}}_{\boldsymbol{U}\sim\mathsf{U}_d}[\mathcal{Q}_\mu^d(\boldsymbol{U})\cdot|q\rangle\langle q'|\cdot\mathcal{Q}_{\mu'}^d(\boldsymbol{U})^\dagger]$$

is an intertwining operator operator for $\mathcal{Q}^d_\mu$ and $\mathcal{Q}^d_{\mu'}$, because for each $V \in \mathsf{U}_d$,

$$\begin{aligned}
\mathcal{Q}^d_\mu(V) \cdot T_{\mu,\mu',q,q'} &= \mathcal{Q}^d_\mu(V) \cdot \mathop{\mathbf{E}}_{\boldsymbol{U}\sim\mathsf{U}_d}[\mathcal{Q}^d_\mu(\boldsymbol{U}) \cdot |q\rangle\langle q'| \cdot \mathcal{Q}^d_{\mu'}(\boldsymbol{U})^\dagger] \\
&= \mathop{\mathbf{E}}_{\boldsymbol{U}\sim\mathsf{U}_d}[\mathcal{Q}^d_\mu(V\boldsymbol{U}) \cdot |q\rangle\langle q'| \cdot \mathcal{Q}^d_{\mu'}(\boldsymbol{U})^\dagger] \\
&= \mathop{\mathbf{E}}_{\boldsymbol{W}\sim\mathsf{U}_d}[\mathcal{Q}^d_\mu(\boldsymbol{W}) \cdot |q\rangle\langle q'| \cdot \mathcal{Q}^d_{\mu'}(V^\dagger\boldsymbol{W})^\dagger] \\
&= \mathop{\mathbf{E}}_{\boldsymbol{W}\sim\mathsf{U}_d}[\mathcal{Q}^d_\mu(\boldsymbol{W}) \cdot |q\rangle\langle q'| \cdot \mathcal{Q}^d_{\mu'}(\boldsymbol{W})^\dagger] \cdot \mathcal{Q}^d_{\mu'}(V) \\
&= T_{\mu,\mu',q,q'} \cdot \mathcal{Q}^d_{\mu'}(V).
\end{aligned}$$

As a result, Schur's lemma states that $T_{\mu,\mu',q,q'}$ is zero when $\mu \neq \mu'$, and a multiple of the identity $c_{\mu,q,q'} \cdot \mathbb{1}$ when $\mu = \mu'$. Indeed, we may compute $c_{\mu,q,q'}$ exactly as

$$\begin{aligned}
c_{\mu,q,q'} = \frac{1}{\dim(\mathcal{Q}^d_\mu)} \cdot \operatorname{tr}[T_{\mu,\mu,q,q'}] &= \frac{1}{\dim(\mathcal{Q}^d_\mu)} \cdot \mathop{\mathbf{E}}_{\boldsymbol{U}\sim\mathsf{U}_d}[\operatorname{tr}[\mathcal{Q}^d_\mu(\boldsymbol{U}) \cdot |q\rangle\langle q'| \cdot \mathcal{Q}^d_\mu(\boldsymbol{U})^\dagger]] \\
&= \frac{1}{\dim(\mathcal{Q}^d_\mu)} \cdot \mathop{\mathbf{E}}_{\boldsymbol{U}\sim\mathsf{U}_d}[\langle q'| \cdot \mathcal{Q}^d_\mu(\boldsymbol{U})^\dagger \mathcal{Q}^d_\mu(\boldsymbol{U}) \cdot |q\rangle] \\
&= \frac{1}{\dim(\mathcal{Q}^d_\mu)} \cdot \mathop{\mathbf{E}}_{\boldsymbol{U}\sim\mathsf{U}_d}[\langle q'|q\rangle] = \begin{cases} 1/\dim(\mathcal{Q}^d_\mu) & \text{if } q = q', \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}$$

Overall, then, $T_{\mu,\mu',q,q'}$ is $(1/\dim(\mathcal{Q}^d_\mu)) \cdot \mathbb{1}$ if $\mu = \mu'$ and $q = q'$ and zero otherwise. Thus,

$$(2.18) = \sum_{\mu,q} |\mu\rangle\langle\mu| \otimes \Big(\frac{1}{\dim(\mathcal{Q}^d_\mu)} \cdot \mathbb{1}\Big) \otimes N_{\mu,\mu,q,q} = \sum_\mu |\mu\rangle\langle\mu| \otimes \mathbb{1} \otimes \Big(\frac{1}{\dim(\mathcal{Q}^d_\mu)} \cdot \sum_q N_{\mu,\mu,q,q}\Big).$$

The lemma follows by taking $N_\mu = (1/\dim(\mathcal{Q}^d_\mu)) \cdot \sum_q N_{\mu,\mu,q,q}$. □

Lemma 37 (EPR state in an irrep)**.** *Given $\mu \vdash m$, we define the EPR state corresponding to the permutation irrep $\mathcal{P}_\mu$ as*

$$|\mathrm{EPR}_\mu\rangle = \frac{1}{\sqrt{\dim(\mu)}} \cdot \sum_p |p\rangle \otimes |p\rangle,$$

where the sum ranges over basis vectors of $\mathcal{P}_\mu$. Then

$$\mathop{\mathbf{E}}_{\boldsymbol{\pi}\sim\mathsf{S}_d}[\mathcal{P}_A(\boldsymbol{\pi}) \otimes \mathcal{P}_B(\boldsymbol{\pi})] = \sum_\mu |\mu\rangle\langle\mu|_A \otimes |\mu\rangle\langle\mu|_B \otimes \mathbb{1}_A \otimes \mathbb{1}_B \otimes |\mathrm{EPR}_\mu\rangle\langle\mathrm{EPR}_\mu|_{A,B},$$

where the two identity matrices act on the $|q\rangle$ registers of Hilbert spaces A and B.

Proof. We begin by calculating

$$\mathop{\mathbf{E}}_{\boldsymbol{\pi}\sim\mathsf{S}_d}[\mathcal{P}_{\mu_A}(\boldsymbol{\pi}) \otimes \mathcal{P}_{\mu_B}(\boldsymbol{\pi})] = \sum_{\substack{p_A,p'_A \\ p_B,p'_B}} |p_A\rangle\langle p'_A| \otimes |p_B\rangle\langle p'_B| \cdot \mathop{\mathbf{E}}_{\boldsymbol{\pi}\sim\mathsf{S}_d}[\mathcal{P}_{\mu_A}(\boldsymbol{\pi})_{p_A,p'_A} \cdot \mathcal{P}_{\mu_B}(\boldsymbol{\pi})_{p_B,p'_B}]. \tag{2.19}$$

The Schur orthogonality relations state that

$$\mathop{\mathbf{E}}_{\boldsymbol{\pi}\sim\mathsf{S}_d}[\mathcal{P}_{\mu_A}(\boldsymbol{\pi})^\dagger_{p_A,p'_A}\cdot\mathcal{P}_{\mu_B}(\boldsymbol{\pi})_{p_B,p'_B}]=\begin{cases}1/\dim(\mu_A) & \text{if } \mu_A=\mu_B,\ p_A=p_B, \text{ and } p'_A=p'_B,\\ 0 & \text{otherwise.}\end{cases}$$

Recall that we have chosen our basis of $\mathcal{P}_\mu$ so that $\mathcal{P}_\mu(\pi)$ is a real-valued (orthogonal) matrix for each $\pi\in\mathsf{S}_d$. Then

$$\mathcal{P}_{\mu_A}(\pi)^\dagger_{p_A,p'_A}=\mathcal{P}_{\mu_A}(\pi)_{p_A,p'_A},$$

and so

$$\mathop{\mathbf{E}}_{\boldsymbol{\pi}\sim\mathsf{S}_d}[\mathcal{P}_{\mu_A}(\boldsymbol{\pi})_{p_A,p'_A}\cdot\mathcal{P}_{\mu_B}(\boldsymbol{\pi})_{p_B,p'_B}]=\begin{cases}1/\dim(\mu_A) & \text{if } \mu_A=\mu_B,\ p_A=p_B, \text{ and } p'_A=p'_B,\\ 0 & \text{otherwise.}\end{cases}\tag{2.20}$$

As a result, (2.19) is zero if $\mu_A\neq\mu_B$, and

$$(2.19)=\frac{1}{\dim(\mu_A)}\cdot\sum_{p,p'}|p\rangle\langle p'|\otimes|p\rangle\langle p'|=|\mathrm{EPR}_{\mu_A}\rangle\langle\mathrm{EPR}_{\mu_A}|.$$

if $\mu_A=\mu_B$. This allows us to express

$$\begin{aligned}\mathop{\mathbf{E}}_{\boldsymbol{\pi}\sim\mathsf{S}_d}[\mathcal{P}_A(\boldsymbol{\pi})\otimes\mathcal{P}_B(\boldsymbol{\pi})]&=\sum_{\mu_A,\mu_B}|\mu_A\rangle\langle\mu_A|\otimes|\mu_B\rangle\langle\mu_B|\otimes\mathbb{1}\otimes\mathbb{1}\otimes\mathop{\mathbf{E}}_{\boldsymbol{\pi}\sim\mathsf{S}_d}[\mathcal{P}_{\mu_A}(\boldsymbol{\pi})\otimes\mathcal{P}_{\mu_B}(\boldsymbol{\pi})]\\&=\sum_{\mu}|\mu\rangle\langle\mu|\otimes|\mu\rangle\langle\mu|\otimes\mathbb{1}\otimes\mathbb{1}\otimes|\mathrm{EPR}_{\mu}\rangle\langle\mathrm{EPR}_{\mu}|.\end{aligned}$$

This completes the proof. □

Next, we have the following immediate corollary of Lemma 37.

Corollary 38. *Consider an operator of the form*

$$O=\sum_{\mu_A,\mu_B}|\mu_A\rangle\langle\mu_A|\otimes|\mu_B\rangle\langle\mu_B|\otimes\mathbb{1}_A\otimes\mathbb{1}_B\otimes O_{\mu_A,\mu_B},$$

where the two identity matrices act on the $|q\rangle$ registers of Hilbert spaces A and B, and the O_{μ_A,μ_B} matrix acts on the $|p\rangle$ registers of A and B, Next, let Z be the matrix

$$Z=\mathop{\mathbf{E}}_{\boldsymbol{\pi}\sim\mathsf{S}_d}[\mathcal{P}_A(\boldsymbol{\pi})\otimes\mathcal{P}_B(\boldsymbol{\pi})].$$

Then

$$Z\cdot O\cdot Z=\sum_{\mu}c_\mu\cdot|\mu\rangle\langle\mu|_A\otimes|\mu\rangle\langle\mu|_B\otimes\mathbb{1}_A\otimes\mathbb{1}_B\otimes|\mathrm{EPR}_{\mu}\rangle\langle\mathrm{EPR}_{\mu}|,$$

for some constants c_μ.

Proof. By Lemma 37,

$$\begin{aligned} Z \cdot O \cdot Z &= \sum_{\mu} |\mu\rangle\langle\mu|_A \otimes |\mu\rangle\langle\mu|_B \otimes \mathbb{1}_A \otimes \mathbb{1}_B \otimes (|\mathrm{EPR}_\mu\rangle\langle\mathrm{EPR}_\mu| \cdot O_{\mu,\mu} \cdot |\mathrm{EPR}_\mu\rangle\langle\mathrm{EPR}_\mu|) \\ &= \sum_{\mu} c_\mu \cdot |\mu\rangle\langle\mu|_A \otimes |\mu\rangle\langle\mu|_B \otimes \mathbb{1}_A \otimes \mathbb{1}_B \otimes |\mathrm{EPR}_\mu\rangle\langle\mathrm{EPR}_\mu|, \end{aligned}$$

where $c_\mu = \langle\mathrm{EPR}_\mu| \cdot O_{\mu,\mu} \cdot |\mathrm{EPR}_\mu\rangle$. This completes the proof. □

Now we prove Theorem 35.

Proof of Theorem 35. Given $\psi \in \{\varphi, \gamma\}$, consider the state M_ψ defined as

$$M_\psi := \mathop{\mathbf{E}}_{\boldsymbol{U}_A \sim \mathsf{U}_d, \boldsymbol{V}_B \sim \mathsf{U}_d} \left(\boldsymbol{U}_A \otimes \boldsymbol{V}_B \cdot |\psi\rangle\langle\psi| \cdot \boldsymbol{U}_A^\dagger \otimes \boldsymbol{V}_B^\dagger\right)^{\otimes m}.$$

Using the left and right invariance property of the Haar measure and the commutation between $\mathcal{P}_A$ and $\mathcal{Q}_A$ (and likewise for $\mathcal{P}_B$ and $\mathcal{Q}_B$), we can see that the mixed state M_ψ remains invariant under the following permutations and unitary rotations:

1. $\mathbf{E}_{\boldsymbol{U} \sim \mathsf{U}_d} \left(\boldsymbol{U}_A^{\otimes m} \cdot M_\psi \cdot (\boldsymbol{U}_A^\dagger)^{\otimes m}\right) = M_\psi,$
2. $\mathbf{E}_{\boldsymbol{V} \sim \mathsf{U}_d} \left(\boldsymbol{V}_B^{\otimes m} \cdot M_\psi \cdot (\boldsymbol{V}_B^\dagger)^{\otimes m}\right) = M_\psi,$
3. $\mathbf{E}_{\boldsymbol{\pi} \sim \mathsf{S}_d} \left(\mathcal{P}_A(\boldsymbol{\pi}) \otimes \mathcal{P}_B(\boldsymbol{\pi}) \cdot M_\psi\right) = M_\psi,$
4. $\mathbf{E}_{\boldsymbol{\pi} \sim \mathsf{S}_d} \left(M_\psi \cdot \mathcal{P}_A(\boldsymbol{\pi}) \otimes \mathcal{P}_B(\boldsymbol{\pi})\right) = M_\psi.$

We can now apply the results of Lemma 36 and Corollary 38 to put the mixed state M_ψ in the following form

$$M_\psi = \sum_{\mu} c_{\psi,\mu} \cdot |\mu\rangle\langle\mu|_A \otimes |\mu\rangle\langle\mu|_B \otimes \mathbb{1}_A \otimes \mathbb{1}_B \otimes |\mathrm{EPR}_\mu\rangle\langle\mathrm{EPR}_\mu|_{AB}. \tag{2.21}$$

We can therefore interpret the density matrix M_ψ as corresponding to a mixed state with one element in the mixture for each block μ. In this case, measuring the block μ can be done without loss of generality, as it does not perturb the state. That can be done entirely on the A subsystem by performing the projective measurement $\{|\mu\rangle\langle\mu|_A\}$. Having done this measurement and received outcome $\boldsymbol{\mu}$, the state is equal to

$$|\boldsymbol{\mu}\rangle\langle\boldsymbol{\mu}|_A \otimes |\boldsymbol{\mu}\rangle\langle\boldsymbol{\mu}|_B \otimes \mathbb{1}_A \otimes \mathbb{1}_B \otimes |\mathrm{EPR}_{\boldsymbol{\mu}}\rangle\langle\mathrm{EPR}_{\boldsymbol{\mu}}|_{AB},$$

regardless of whether $\psi = \varphi$ or γ. Hence, no further information can be learned about ψ by performing any further measurements, and this implies that measuring only the A subsystem is without loss of generality. □

Chapter 3

Entanglement spread area law in gapped ground states

Chapter summary: In this chapter, we make a connection between two seemingly different problems. The first problem involves characterizing the properties of entanglement in the ground state of gapped local Hamiltonians, which is a central topic in quantum many-body physics. The second problem is on the quantum communication complexity of testing bipartite states, a well-known question in quantum information theory. We begin with constructing a communication protocol between two parties that allows them to test if they share a specific bipartite state. We then use the communication complexity of this protocol to reveal a new structural property for the ground state entanglement. This property is captured in a quantity known as the entanglement spread, which measures the difference between the Rényi entanglement entropies. Our main result shows that gapped ground states possess limited entanglement spread across any cut, exhibiting an "area law" behavior. Our result applies to any interaction graph with an improved bound for the special case of lattices. This entanglement spread area law includes interaction graphs constructed in [AHL+14] that violate a generalized area law for the entanglement *entropy*.

On the technical side, we use recent advances in Hamiltonian simulation algorithms along with the quantum phase estimation to give a new construction for an approximate ground space projector (AGSP) over arbitrary interaction graphs, which might be of independent interest. This chapter is based on:

[AHS20] Anurag Anshu, Aram W Harrow, and Mehdi Soleimanifar. From communication complexity to an entanglement spread area law in the ground state of gapped local Hamiltonians. *To appear in Nature Physics. Preprint available at arXiv:2004.15009*, 2020.

3.1 Introduction

3.1.1 Background on area law and entanglement spectra

The ground states of local Hamiltonians are examples of quantum many-body states with central significance in condensed matter physics and quantum chemistry. A crucial distinction between these states and their classical counterparts – the satisfying assignments in constraint satisfaction problems – is the presence of multipartite entanglement. This leads to novel phenomena in these systems such as exotic phases of matter, but also complicates the theoretical and numerical study of their properties.

There is a successful line of research that applies the tools developed in quantum information theory and computer science to study various features of entanglement in the ground states. An important problem that has been the focus of many such studies is proving a conjecture known as the "area law" for the entanglement entropy in the ground state of gapped local Hamiltonians. We can more precisely state this by considering the interaction (hyper)graph where each vertex represents a qudit and the edges correspond to the interaction terms in the Hamiltonian (see Figure 3-1). Suppose we fix a partition of the qudits into two parts A and B. We denote the ground state by $|\Omega\rangle_{AB}$. In general, the qudits in part A will be entangled with those in part B. The area law asserts that the amount of entanglement – measured by the entropy of either of the reduced states Ω_A or Ω_B – is at most proportional to the number of interaction terms that cross the cut ∂A. This behavior is drastically different from the generic situation where the entanglement across the cut ∂A scales with the size of the smaller partition $|A|$ rather than $|\partial A|$. Thus, loosely speaking, the area law implies that the ground state entanglement is local and limited to the boundary. This conjecture has been rigorously proven when the interaction graph is a 1D chain [Has07a, ALV12, AKLV13] and there has been recent progress on trees [Abr19] and 2D lattices [AAG21].

A generalization of this conjecture asks if the area law holds for arbitrary interaction graphs beyond lattices. It turns out that this generalized conjecture is false. Using quantum expanders, an interaction graph is constructed in [AHL+14] which admits a partition into two parts A and B such that the size of the cut is $|\partial A| = 1$, but the amount of entanglement across the cut is proportional to $|A|$.

Thus far, these results study the ground state entanglement in terms of the *entropy* of the reduced state Ω_A on partition A. One can go beyond this and consider other features of the eigenvalues of the reduced state Ω_A besides its entropy. We denote the non-zero eigenvalues of Ω_A by $\lambda_1, \lambda_2, , \ldots, \lambda_r$ which are often called the Schmidt coefficients. A common way of describing the distribution of these eigenvalues is by using the *entanglement spectrum* which is defined as the log of the inverse of the non-zero Schmidt coefficients $\{\log \frac{1}{\lambda_1}, \log \frac{1}{\lambda_2}, \ldots, \log \frac{1}{\lambda_r}\}$. More intuitively, the entanglement spectrum of a reduced state Ω_A is the set of eigenvalues of a Hamiltonian H_{mod}, known as the modular Hamiltonian, such that the reduced state Ω_A corresponds to the Gibbs (thermal) state of H_{mod}. That is, $\Omega_A = e^{-H_{\text{mod}}}$. Various works, initiated by a result of Li and Haldane [LH08], relate the entanglement spectrum to different characteristics of a given phase (see e.g., [PTBO10, SPCPG13, CPSV11, KBa19a, DVZ18]). These studies suggest that in 2D gapped systems, the entanglement spectrum often has features of the spectrum of a 1D local Hamiltonian, although its features are not always isomorphic to the original Hamiltonian [CKS14]. These studies motivate investigating different features of the entanglement spectrum beyond what is evident from its entropy.

Inspired by these results, we prove a new structural property for the entanglement of gapped ground states. The key to our findings is a connection to the field of quantum communication complexity. A basic question there is when two parties want to *test* whether they share a specific entangled state $|\psi\rangle$ by exchanging as few messages as possible. In other words, they want to perform the measurement $\{|\psi\rangle\langle\psi|, \mathbb{1} - |\psi\rangle\langle\psi|\}$. Building on [CH19], we resolve the communication complexity of this problem and relate it to the details of the entanglement in $|\psi\rangle$. We can therefore learn about the nature of the ground state entanglement by allowing $|\psi\rangle$ to be be the ground state $|\Omega\rangle$ of a local Hamiltonian and design a testing protocol for $|\Omega\rangle$. We devise such a measurement protocol tailored for the ground state of *gapped* local Hamiltonians by combining recent Hamiltonian simulation techniques with the quantum phase estimation algorithm, which might be of independent

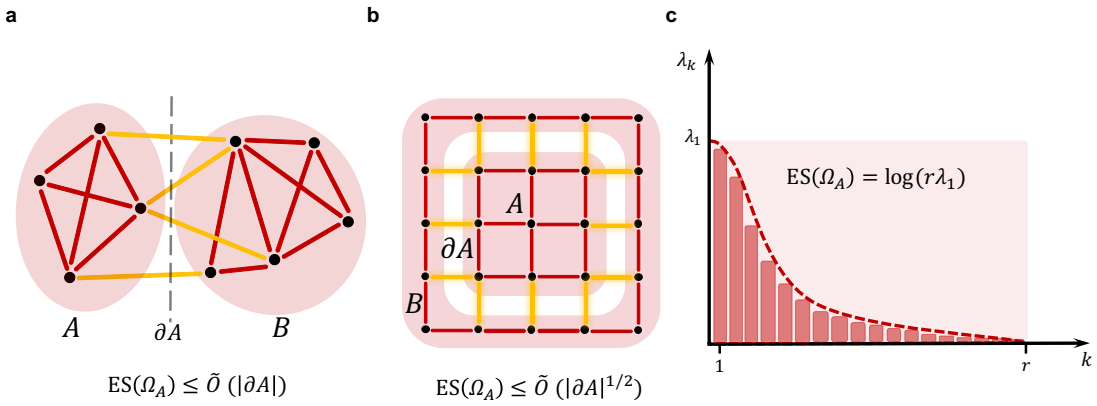


Figure 3-1: **a**, Interacting quantum systems with a general interaction graph partitioned into two parts A and B with boundary ∂A. **b**, A similar partition for a system on a lattice. **c**, The profile of the eigenvalues (aka Schmidt coefficients) $\lambda_1, \lambda_2, \ldots, \lambda_r$ of the reduced ground state in region A. This distribution fully encodes the information about the bipartite entanglement between particles in region A and B . Modulo some smoothing that we explain in the main text, the entanglement *spread* across the cut ∂A is defined as $\log(r\lambda_1) \approx \log(\lambda_1/\lambda_r)$. We show that the entanglement spread scales as $O(|\partial A|)$. We improve this to $O(\sqrt{|\partial A|})$ for lattice Hamiltonians. In comparison the area law conjecture for the *entanglement entropy* asserts that the entropy of the distribution of the Schmidt coefficients is bounded by the size of the cut $|\partial A|$, i.e. $S(\Omega_A) = O(|\partial A|)$.

interest.

The property that we study is known as the *entanglement spread*, which roughly measures the log of the ratio between the largest and the smallest eigenvalue of the reduced state Ω_A, giving an estimate of how spread out their distribution is (see Figure 3-1). Our results apply generally to *any* interaction graph, with some improved statement for the special case of lattices. We show that as long as the Hamiltonian is gapped, its ground state possesses limited entanglement spread on general interaction graphs, exhibiting an "area law" behavior. On lattices, we prove a sub-area scaling for this quantity. We connect our result to conjectures regarding the locality of modular Hamiltonians and also efficient algorithms for estimating the expected value of local observables in the ground state. We also show that both states that satisfy the entropy area law and those in the counter-example construction in [AHL+14] fit into our framework. In the next sections, we provide a more detailed overview of our setup and results.

3.1.2 Entanglement spread and communication complexity of non-local measurements

Consider a bipartite state $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, where Alice and Bob own registers, $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively. Suppose the parties engage in a communication protocol whose goal is to test if they share the state $|\psi\rangle$. That is, they would like to implement the reflection operator $\mathrm{Ref}(\psi) = 2|\psi\rangle\langle\psi| - \mathbb{1}$ or similarly, perform the two-outcome measurement $\{|\psi\rangle\langle\psi|, \mathbb{1} - |\psi\rangle\langle\psi|\}$. In our setup, these tasks are locally interchangeable. Namely, the ability to perform controlled reflections will give us the ability

to do measurements and vice versa [1]. Since in general, the state $|\psi\rangle$ is an entangled state, Alice and Bob need to exchange qubits to perform this operation. For instance, Alice can send her register $\mathcal{H}_A$ to Bob who then performs the joint operation on $\mathcal{H}_A \otimes \mathcal{H}_B$ and sends back Alice's register. As we will see later, they can often do much better. The communication complexity (or cost) of such a protocol $\mathrm{C}_\Delta(\psi)$ is defined as the *minimum* number of qubits that the parties need to exchange to perform this task with error at most Δ. We refer to $\mathrm{C}_\Delta(\psi)$ as the communication complexity of an *EPR-assisted protocol*, if we allow Alice and Bob to also share unlimited EPR pairs during the communication protocol. Here, we are interested in the case where $|\psi\rangle$ is the ground state of a gapped local Hamiltonian. In other words, we ask

> *What is the communication cost $C_\Delta(\Omega)$ of approximately measuring or reflecting about the ground state of a gapped local Hamiltonian which is shared between Alice and Bob?*

Before specializing to the ground state, it is insightful to consider a few general instances.

1) $|\psi\rangle = |0\rangle^{\otimes n}|0\rangle^{\otimes n}$: This is equivalent to a CZ gate, or equivalently a CNOT gate, which has communication cost 1.

2) $|\psi\rangle = \frac{1}{\sqrt{p}} \sum_{j=1}^{p} |j\rangle|j\rangle$: Initially, one might think that reflecting about the maximally entangled state $|\Phi_p\rangle = \frac{1}{\sqrt{p}} \sum_{j=1}^{p} |j\rangle|j\rangle$ requires exchanging a large number of qubits, but it turns out that by using quantum expanders [AHL+14], one can perform such a reflection up to error Δ by exchanging $\mathrm{C}_\Delta(\psi) = O(\log(1/\Delta))$ qubits, which is independent of the dimension p.

3) $|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle^{\otimes n} + |\Phi_2\rangle^{\otimes n})$: This is a superposition of the last two cases with $|\Phi_2\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ being the EPR pair. We claim that $\mathrm{C}_\Delta(\psi) = \Theta(n)$. This can be verified by noticing that $\| \mathrm{Ref}(\psi)|00\rangle^{\otimes n} - |\Phi_2\rangle^{\otimes n}\| \leq 2^{-(n-1)/2}$, but it is well-known that creating n EPR pairs $|\Phi_2\rangle^{\otimes n}$ from product states requires $\Omega(n)$ qubits of communication; see also [HL11].

In general, by applying local unitaries, a bipartite state $|\psi\rangle$ can be always transformed to a standard form $|\psi\rangle = \sum_{j=1}^{r} \sqrt{\lambda_j}|j\rangle|j\rangle$ known as the Schmidt decomposition. Thus, we expect $\mathrm{C}_\Delta(\psi)$ to depend only on the Schmidt coefficients λ_j. We assume these coefficients are arranged in the descending order $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r$, where r is called the Schmidt rank of $|\psi\rangle$. A closer look at the above examples reveals a pattern. In the first two instances, where $\mathrm{C}_\Delta(\psi)$ is small, the Schmidt coefficients of $|\psi\rangle$ are all equal (which we refer to as the *concentrated* case). In the third example, which has high communication cost, the Schmidt coefficients are *spread out* between two different values $\frac{1}{2}(1 + \frac{1}{\sqrt{2^n}})^2 \approx 1/2$ and $\frac{1}{2^{n+1}}$ with almost equal weights.

This motivates a more general lower bound on the communication complexity in terms of the *entanglement spread* of the state $|\psi\rangle$, which is a measure of how spread out the Schmidt coefficients are across a cut [HW03]. In its simplest form, the entanglement spread, denoted by $\mathrm{ES}(\psi)$, is defined by

$$\mathrm{ES}(\psi) = \log(r\lambda_1) = S_{\max}(\psi_A) - S_{\min}(\psi_A), \tag{3.1}$$

[1] To see this, just locally initialize a qubit in state $|+\rangle$, perform a controlled reflection, and then locally measure in the Hadamard basis. The ability to do a coherent measurement also gives the power of reflection: We can add a -1 phase if the outcome of the measurement is $\mathbb{1} - |\psi\rangle\langle\psi|$.

where $\psi_A = \mathrm{tr}_B |\psi\rangle\langle\psi|$ and

$$S_{\max}(\psi_A) = \log r, \quad S_{\min}(\psi_A) = -\log \lambda_1 \tag{3.2}$$

are the max- and min-entropies given in terms of the Schmidt rank r and the maximum Schmidt coefficient λ_1 (see Figure 3-1). Indeed, one can verify that in the first two examples, $\mathrm{ES}(\psi) = 0$ while in the third case, $\mathrm{ES}(\psi) = \Theta(n)$. Thus, $\mathrm{ES}(\psi)$ distinguishes between the concentrated versus spread out cases in the examples above.

The definition of the entanglement spread given in equation (3.1) is however not robust to small perturbation of the spectrum of the eigenvalues. For instance, addition of a series of small eigenvalues to the tail of the distribution significantly affects the Schmidt rank r but alters the state negligibly in trace distance. In our results, we need a more robust version of the entanglement spread (3.1) that applies to protocols which only *approximately* implement the two-outcome ground state measurement, i.e. when the error $\Delta > 0$. We denote this version by $\mathrm{ES}_\delta(\psi)$ and following [HW03], we define:

Definition 39 (Entanglement spread)**.** *Let $\delta \in [0, 1)$. The (δ-smooth) entanglement spread of a bipartite state $|\psi\rangle \in \mathcal{H}_{AB}$ is defined by*

$$\mathrm{ES}_\delta(\psi) = S^\delta_{\max}(\psi_A) - S^\delta_{\min}(\psi_A),$$

where $S^\delta_{\min}(\psi_A)$ and $S^\delta_{\max}(\psi_A)$ are the smooth min- and max- entropies defined similar to (3.2) after removing up to a mass δ from the Schmidt distribution of $|\psi\rangle$ (see Definition 48 in the body for details).

We now look more closely at why, besides the above examples, the entanglement spread provides a lower bound on the communication cost of measuring (testing) a *general* state $|\psi\rangle$? In the exact case $\Delta = 0$, this can be seen by observing that for each qubit exchanged between Alice and Bob, the rank r and the largest Schmidt coefficient λ_1 change at most by a factor of 2 and hence, after c rounds of communication, $\log(r\lambda_1)$ is at most $2c$ (refer to [HW03, Theorem 1] for the more detailed proof of the $\Delta = 0$ case). This shows that (modulo a constant) $\mathrm{ES}(\psi)$ provides a lower bound on the exact communication complexity. In the approximate regime $\Delta > 0$, similar lower bounds in terms of $\mathrm{ES}_\delta(\psi)$ have been proved before [HW03, CH19]. We derive an analogous lower bound tailored for when the state of interest is the ground state of a gapped local Hamiltonian, which is explained in detail in the next section.

A very useful property of the entanglement spread is that it remains unchanged when the state is supplemented by arbitrary numbers of EPR pairs. That is, $\mathrm{ES}_\delta(\psi \otimes \Phi_p) = \mathrm{ES}_\delta(\psi)$ for any maximally entangled state of arbitrary size p. One way of seeing this is that adding $|\Phi_p\rangle$ multiplies r by 2^p and divides λ_1 by 2^p, leaving the entanglement spread unchanged. One interesting implication of this equality is that the lower bound on the communication complexity in terms of $\mathrm{ES}_\delta(\psi)$ continues to hold even in protocols where Alice and Bob share an arbitrary number of EPR pairs during their communication. As we will see later, this improves and simplifies our analysis. We denote the shared EPR pairs collectively by $|\Phi\rangle$ and call such protocols *EPR-assisted*.

3.2 Our results

3.2.1 Lower bound on communication complexity from entanglement spread

As described in the previous section, the goal of Alice and Bob is to implement a measurement $\{K, \mathbb{1}-K\}$ that acts jointly on an input state and the shared EPR pairs $|\Phi\rangle$, where K is an operator approximately projects the input state onto the ground state $|\Omega\rangle$ while leaving $|\Phi\rangle$ untouched. More precisely, we define:

Definition 40 (EPR-assisted AGSP)**.** *An EPR-assisted Approximate Ground State Projector (EPR-assisted AGSP for short) associated with the ground state $|\Omega\rangle$ of a local Hamiltonian is an operator K that for some error $\Delta \in (0,1)$ satisfies*

$$\|(K - \mathbb{1} \otimes |\Omega\rangle\langle\Omega|)\ |\Phi\rangle|\psi\rangle\| \leq \Delta \quad \textit{for all} \quad |\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B.$$

In our first result, we give a lower bound on the communication complexity of implementing this operator in terms of the entanglement spread of the ground state.

Theorem 41 (Lower bound on the complexity of EPR-assisted AGSP)**.** *Let $|\Omega\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ be the ground state of a local Hamiltonian shared between Alice and Bob. For any error $\Delta \in (0,1)$, the communication complexity of implementing the two-outcome measurement $\{K, \mathbb{1}-K\}$ where K is the EPR-assisted AGSP corresponding to $|\Omega\rangle$ is lower bounded by*

$$\mathrm{C}_\Delta(\Omega) \geq \mathrm{ES}_{(4\sqrt{2}\Delta)^{2/3}}(\Omega) - 1 = S_{\max}^{(4\sqrt{2}\Delta)^{2/3}}(\Omega_A) - S_{\min}^{(4\sqrt{2}\Delta)^{2/3}}(\Omega_A) - 1. \tag{3.3}$$

We note that the above theorem applies to any state $|\Omega\rangle$, as long as an approximate projection operator K (similar to Definition 40) exists. But we keep the "ground state" terminology in our discussion, to fit the context.

3.2.2 Communication protocol for approximate ground space projection

In Theorem 41, we stated a lower bound on $\mathrm{C}_\Delta(\Omega)$, the communication complexity of approximately measuring the ground state. The main claim of this section is an *upper bound* on the communication cost $\mathrm{C}_\Delta(\Omega)$ which is obtained from an explicit communication protocol that approximately implements a ground state measurement.

Let $|A|$ be total number of qudits on Alice's side. Alice and Bob can trivially implement K by exchanging $|A|$ qudits. Our result shows that when the input state is the ground state of a gapped Hamiltonian, the communication complexity can be improved to $O(|\partial A|)$, where $|\partial A|$ is the number of terms in the Hamiltonian that act on both Alice and Bob's registers, see Figure 3-1.

Theorem 42 (Communication protocol for projecting onto the ground space)**.** *Suppose the state $|\Omega\rangle$ is the ground state of a local Hamiltonian with spectral gap γ (See Section 3.5 for a formal definition of "local Hamiltonian."). Let $|\partial A|$ be the number of terms in the Hamiltonian that acts on both Alice and Bob's qudits. Then, there exists a protocol that implements the measurement $\{K, \mathbb{1}-K\}$, where K is an EPR-assisted AGSP satisfying*

$$\|(K - \mathbb{1} \otimes |\Omega\rangle\langle\Omega|)\ |\Phi\rangle|\psi\rangle\| \leq \Delta \quad \textit{for all} \quad |\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B.$$

and has the communication cost

$$c = O\left(\frac{|\partial A|}{\gamma} \cdot \log \frac{1}{\Delta} \cdot \text{polylog}(\frac{|\partial A|}{\gamma}, \frac{1}{\Delta})\right) \tag{3.4}$$

As a result of Theorem 41 in the previous section, we know that the communication complexity of performing an AGSP gives us an upper bound on the entanglement spread of the ground state. When combined with the bound (3.4), this establishes an area law for the entanglement spread, meaning that across a given cut in the ground state of a gapped Hamiltonian, the Schmidt coefficients can be spread out at most proportional to the size of the cut.

Corollary 43 (Area law for entanglement spread)**.** *Under the assumptions of Theorem 41 and Theorem 42, the following bound on the δ-smooth entanglement spread of the ground state $|\Omega\rangle$ of a gapped local Hamiltonian holds,*

$$\text{ES}_\delta(\Omega) \leq O\left(\frac{|\partial A|}{\gamma} \cdot \log \frac{1}{\delta} \cdot \text{polylog}(\frac{|\partial A|}{\gamma}, \frac{1}{\delta})\right) \tag{3.5}$$

Note that the range of applicability of Corollary 43 is quite general. The bound (3.5) holds for any local Hamiltonian over an arbitrary interaction (hyper)graph. In particular, we do not assume the Hamiltonian is also *geometrically local* or the qudits are arranged on a *lattice*. In fact, when the Hamiltonian is restricted to any finite dimensional lattice, we can obtain tighter bounds on the entanglement spread by lifting the powerful machinery of AGSPs based on the Chebyshev polynomials [AKLV13, ALV12] from 1D geometries to higher dimensions. This quadratically improves the bound (3.4) to $c = \tilde{O}(\sqrt{|\partial_w A|/\gamma})$ at the cost of including an extended boundary $\partial_w A$ of constant width w instead of the original boundary ∂A. Note that $|\partial_w A| = O(|\partial A|)$ on lattices when $w = O(1)$. This is an intriguing result since it shows a "sub-volume" scaling for an information theoretic quantity in the gapped ground states. We further discuss this in Section 3.4. In this setting, it is convenient to view 2^c as the Schmidt rank of the AGSP operator across the cut. We also do not rely on shared EPR pairs in this setup. To distinguish things from our previous construction, we refer to this operator as the Chebyshev-AGSP. More precisely, we have:

Theorem 44 (Chebyshev-AGSP for lattices)**.** *Suppose, H is a geometrically-local Hamiltonian with gap γ over a finite-dimensional lattice. Let $(A:B)$ be a bipartition of the lattice. There is an operator K with the Schmidt rank 2^c across the partition $(A:B)$ such that $\|K - |\Omega\rangle\langle\Omega|\| \leq \Delta$ and*

$$c = O\left(\sqrt{\frac{|\partial_w A|}{\gamma}} \cdot \log \frac{1}{\Delta} \cdot \text{polylog}\left(\frac{|\partial_w A|}{\gamma}, \frac{1}{\Delta}\right)\right), \tag{3.6}$$

where $\tilde{O}$ hides constant factors related to the geometry of the Hamiltonian and $w = O(1)$. Here, $|\partial_w A|$ is the number of terms in the Hamiltonian that act on the qudits in some extended boundary of constant width around ∂A.

Corollary 45 (Tighter bounds on entanglement spread on lattices)**.** *Under the same conditions in Theorem 41 and Theorem 44, the δ-smooth entanglement spread of the ground state of geometrically*

local Hamiltonians is bounded by

$$\mathrm{ES}_\delta(\Omega) \leq O\left(\sqrt{\frac{|\partial_w A|}{\gamma}} \cdot \log\frac{1}{\delta} \cdot \mathrm{polylog}\left(\frac{|\partial_w A|}{\gamma}, \frac{1}{\delta}\right)\right) \tag{3.7}$$

where $w = O(1)$.

3.3 Main ideas

Here, we describe the main ideas and technical tools used in the proof of our results.

3.3.1 AGSP from quantum phase estimation

One ingredient of our proofs is a novel construction of an AGSP for the ground state of gapped Hamiltonians based on the quantum phase estimation (QPE) algorithm. We find a protocol between Alice and Bob that allows them to jointly apply this AGSP with communication complexity $\tilde{O}(|\partial A|/\gamma)$. As mentioned in Section 3.2.2, one advantage of using QPE compared to the conventional Chebyshev polynomials (reviewed in Section 3.3.4) is that it applies not only to geometrically-local Hamiltonians on lattices, but also continues to work for any local Hamiltonian on arbitrary interaction graphs.

One can view QPE as a procedure that given an eigenstate of a Hamiltonian H, uses $O(\log(1/\gamma))$ many ancillary qubits, makes $O(1/\gamma)$ queries to the Hamiltonian simulation oracle e^{-iH}, and determines the energy of the input state with accuracy $\gamma/2$. By letting γ be the gap of the Hamiltonian, this algorithm basically performs a two outcome measurement $\{\Omega, \mathbb{1} - \Omega\}$ on any input state, where Ω is the ground state of H.

To implement this algorithm in a distributed fashion involving two parties, Alice and Bob need to prepare and reflect about the state $\frac{1}{\sqrt{T+1}} \sum_{t=0}^{T} |t\rangle|t\rangle$ for $T = O(1/\gamma)$ and work together to apply the operator e^{-itH} conditioned on the register $|t\rangle$. In the next section, we show how to achieve this.

3.3.2 Communication protocol based on interaction picture Hamiltonian simulation

For a given partition of the qudits between Alice and Bob, we can write the Hamiltonian as $H = H_A + H_{\partial A} + H_B$ where $[H_A, H_B] = 0$. One of our main technical contributions is designing a communication protocol for performing the Hamiltonian simulation operator e^{-itH} with a communication cost that scales as $O(t\|H_{\partial A}\|)$ instead of the conventional $O(t\|H\|)$.

It is not hard to see how one can achieve this if the boundary term $H_{\partial A}$ also commutes with H_A and H_B. In that case, we have $e^{-itH} = e^{-itH_A} e^{-itH_{\partial A}} e^{-itH_B}$ and the parties can implement e^{-itH} if one of them sends the boundary qudits that are in the support of $H_{\partial A}$ to the other. This yields a communication cost that scales with $|\mathrm{supp}(H_{\partial A})| = O(|\partial A|)$. In general, however, $H_{\partial A}$ does not commute with H_A and H_B and finding a non-trivial protocol for the Hamiltonian simulation becomes challenging.

One attempt to remedy this might be to use the Trotterization technique. That is, to divide the simulation into η segments and implement $e^{-itH/\eta}$ for η consecutive times. If η is large enough, $[H_{\partial A}/\eta, H_{A \text{ or } B}/\eta] \approx 0$, and we again recover the commuting case. That is, the parties

collaboratively implement $e^{-itH_{\partial A}/\eta}$. Unfortunately, for this to work, we need η (and therefore, the communication cost) to be $O(t\|H\|)$, which is far from the bound $O(t\|H_{\partial A}\|)$ we are aiming for.

We instead use a recent framework for Hamiltonian simulation developed in [LW18] known as the "interaction picture" Hamiltonian simulation. Intuitively, one can view this as a sophisticated change of variables that is widely used in physics and allows us to separate the contribution of the boundary term from H_A and H_B. Suppose we want to prepare the state $|\psi(t)\rangle = e^{-itH}|\psi(0)\rangle$. For any $|\psi(t)\rangle$, we define its counterpart in the interaction picture by

$$|\psi_I(t)\rangle = e^{it(H_A+H_B)}|\psi(t)\rangle. \tag{3.8}$$

Since the operator $e^{it(H_A+H_B)}$ can be applied locally by the parties, the states $|\psi_I(t)\rangle$ and $|\psi(t)\rangle$ can be switched with each other with no extra communication. The point of this transformation is that the state $|\psi_I(t)\rangle$ can be prepared starting from $|\psi(0)\rangle$ by applying a unitary $U(t)$ which is the Hamiltonian simulation operator associated with a *time dependent* Hamiltonian

$$H_I(t) = e^{it(H_A+H_B)}H_{\partial A}e^{-it(H_A+H_B)}. \tag{3.9}$$

Putting the time-dependence of $H_I(t)$ aside (we discuss that in more details in Section 3.7.1), the main gain is that $\|H_I(t)\| = \|H_{\partial A}\|$. This solves the issue we mentioned before because here, the length of Trotter step η in implementing $U(t)$ can be taken as small as $O(t\|H_{\partial A}\|)$ instead of the original $O(t\|H\|)$. The remaining task is to find a communication protocol for performing $U(t/\eta)$, which now, is not simply the operator $e^{-itH_{\partial A}/\eta}$ that we had before. This is done in [BCC+15, LW18] using the Linear Combination of Unitaries (LCU) method. Our next idea is a modification of this algorithm that suits our framework better.

3.3.3 EPR-assisted communication and the LCU method

Our results regarding the ground state entanglement and the communication complexity are information theoretic in nature. In particular, the running time or other algorithmic aspects of the tools we use, such as the Hamiltonian simulation, do not affect our conclusions. Here, we explain how we can use this observation to simplify the analysis of a part of our protocol.

In the LCU method, we express the Hamiltonian simulation operator $U(t)$ as sum of unitaries $U(t) \approx \sum_k \alpha_k u_k^{(A)} \otimes u_k^{(B)}$ for some choice of coefficients $\alpha_k \in \mathbb{R}$ and unitaries $u_k^{(A)}$ and $u_k^{(B)}$ that act on Alice and Bob's qudits respectively. Such decomposition as a sum of unitaries could come from the truncation of a Taylor expansion. To keep the running time efficient such Taylor expansions are truncated at low orders.

When Alice and Bob jointly implement the LCU algorithm, they need to prepare and share the ancillary state $|\alpha\rangle = (\sum_k \alpha_k)^{-1/2} \sum_k \sqrt{\alpha_k}|k\rangle_A|k\rangle_B$. Then, they proceed by applying the unitaries $u_k^{(A)} \otimes u_k^{(B)}$ conditioned on their register $|k\rangle$. Now suppose instead of truncating the expansions, we continue adding higher terms. Of course, the issue is that the number of coefficients α_k and thus, the communication cost of sharing $|\alpha\rangle$ and reflecting about $|\alpha\rangle$ also increases. On the other hand, we know that if instead of $|\alpha\rangle$, the parties share a maximally entangled state, the bound (3.5) on the entanglement spread remains intact. In other words, it is not the number of exchanged ancillary qubits in the protocol, but their entanglement spread that affects our final bound (3.5).

We fix this problem by modifying the LCU algorithm such that instead of the state $|\alpha\rangle$, Alice and Bob only share the maximally entangled state (or equivalently some number of EPR pairs).

This state only needs to be shared once, which can be done outside the protocol, and then many reflections about it can be done with a cost independent of the size of the state. Now we can keep an unbounded number of terms in the expansions and avoid similar approximations in our protocol. This blows up the running time of these procedures, but maintains the communication complexity.

3.3.4 AGSP for lattices

Our improved bound for the lattice Hamiltonians in Theorem 44 are obtained using the AGSPs based on the Chebyshev polynomials. These were first developed in the context of the area law for entanglement entropy in 1D systems [ALV12, AKLV13, ALVV17]. The AGSP framework [AALV09, ALV12] in itself provides a framework to connect the min-entropy and entanglement entropy (see [AALV09, Lemma 5.3] or [ALV12, Lemma III.3]). But this connection does not give us the desired bound on entanglement spread, as it relates entanglement entropy and min-entropy by a certain *multiplicative factor*, that may be large. For instance, [ALV12, Lemma III.3] implies that by choosing the Chebyshev-based AGSP which has a shrinking of $O(1)$ and the Schmidt rank of $2^{O(\sqrt{|\partial A|})}$, we get

$$S(\Omega_A) = O(\sqrt{|\partial A|} S_{\min}(\Omega_A)), \tag{3.10}$$

where $S(\Omega_A)$ is the von-Neumann entropy of (Ω_A).

Here, we show that a simple adaptation of the Chebyshev-based AGSP, along with appropriate smoothing, leads to a stronger theorem for lattices, which shows that entanglement spread scales as $O(\sqrt{|\partial A|})$ (see discussion section for the interpretation). We utilize the "truncation step" [AKLV13] which is used to lower the norm of the Hamiltonian away from a cut while maintaining its gap and ground state. We apply the truncation to both the frustration-free and frustrated cases. In the former, we use the Detectability Lemma operator [AALV09], while in the latter, we rely on the recent techniques of [KS20a] to perform the truncation.

3.4 Discussion and connection to previous work

Quadratically better scaling on lattices: In Corollary 45, we have shown that the entanglement spread on lattices scales as $\sqrt{|\partial A|}$ The intuition behind this comes from the exponential decay of correlations which is shown to hold for gapped Hamiltonians on any finite dimensional lattice [Has04, HK06, NS06]. The decay of correlations implies that the distant qudits along the boundary ∂A are almost uncorrelated. This suggests that the ground state across the boundary is roughly in a product form $|\phi\rangle_{AB}^{\otimes |\partial A|}$ composed of $O(|\partial A|)$ partially entangled states $|\phi\rangle_{AB}$. By using conventional concentration bounds [LP99], it can be shown that the *smooth* entanglement spread obeys $\mathrm{ES}_\delta(\phi^{\otimes k}) = O(\sqrt{k})$, which is quadratically smaller than the $\delta = 0$ case where $\mathrm{ES}(\phi^{\otimes k}) = O(k)$. Thus, an entanglement spread of $O(\sqrt{|\partial A|})$ that we prove for gapped lattice Hamiltonians matches our intuitive expectation.

One might wonder if our quadratic bound in Theorem 44 for lattices can be improved. Here, we show that this is not possible in general. Consider a D-dimensional cubic lattice $[n]^D$ for an even n such that the qubits are located on the vertices of the lattice. Let $A = [\frac{n}{2} - 1] \times [n] \times \ldots \times [n]$ define a bipartition of this lattice. Suppose, we have a Hamiltonian H on the lattice given by

$$H = \sum_{i_1,\ldots,i_D : i_1 = \text{even}} (\mathbb{1} - \Psi_{i_1,\ldots i_D}),$$

where the entangled states

$$|\Psi\rangle_{i_1,\dots i_D} = \left(\sqrt{\frac{2}{3}}|0\rangle_{i_1,\dots i_D}|0\rangle_{i_1+1,\dots i_D} + \sqrt{\frac{1}{3}}|1\rangle_{i_1,\dots i_D}|1\rangle_{i_1+1,\dots i_D}\right)$$

is defined between qudits $(i_1, i_2, \dots i_D)$ and $(i_1+1, i_2, \dots i_D)$. Then, the ground state is the simple two-qudit product state

$$\bigotimes_{i_1,\dots i_D : i_1 = \text{even}} |\Psi\rangle_{i_1,\dots i_D}.$$

It is easily seen that the entanglement spread across this partition is at least $\Omega(\sqrt{n^{D-1}}) = \Omega(\sqrt{|\partial A|})$ achieving the bound in Theorem 44.

Is it possible to prove an analog of Theorem 44 with $S_{\min}^{(4\sqrt{2}\Delta)^{2/3}}(\Omega_A)$ replaced by $S_{\min}(\Omega_A)$? This cannot be done without changing the upper bound from $O(|\partial A|^{1/2})$ to $O(|\partial A|)$, since the two-qubit product ground state constructed above has the property that $S_{\min}(\Omega_A) \leq |\partial A|/4$ and $S_{\max}^{(4\sqrt{2}\Delta)^{2/3}}(\Omega_A) \geq 0.92|\partial A|$. Improving $S_{\max}^{(4\sqrt{2}\Delta)^{2/3}}(\Omega_A)$ to $S_{\max}(\Omega_A)$ is also not possible since there are Hamiltonians such as the transverse field Ising model that are gapped but have $S_{\max}(\Omega_A)$ scaling with the system size [CC04].

Locality of modular Hamiltonians: As mentioned in Section 3.1.1, given the reduced state Ω_A, the modular (or entanglement) Hamiltonian H_{mod} is defined such that $\Omega_A = e^{-H_{\text{mod}}}$. In other words, Ω_A corresponds to the Gibbs (thermal) state of H_{mod}. The eigenvalues of H_{mod} are known as the entanglement spectrum [LH08]. An application of our result in Theorem 44 and Corollary 45 is to give formal evidence for a conjecture regrading the entanglement spectrum studied in previous works [LH08, SPCPG13, CPSV11, KBa19a, DVZ18]). According to this conjecture, in 2D gapped systems, the entanglement spectrum often has similar features to that of the spectrum of a 1D local Hamiltonian. In particular, the entanglement entropy area law predicts an $O(|\partial A|)$ scaling for the entropy of Ω_A which matches the entropy of the Gibbs state of a 1D Hamiltonian. A more general question is what aspects of the modular Hamiltonian H_{mod} are similar to that of a 1D local Hamiltonian beyond simply the $O(|\partial A|)$-scaling of the entropy. We contribute to this by showing one further feature, namely the $O(\sqrt{|\partial A|})$-scaling of the entanglement spread of Ω_A.

To see why the $O(\sqrt{|\partial A|})$ scaling predicted by our bound (3.7) for the gapped ground states is in agreement with H_{mod} being a 1D local Hamiltonian, we use the well-known fact that at thermal equilibrium, the energy distribution of a many-body quantum system is concentrated around the average energy. Indeed if H_{mod} is a sum of $O(|\partial A|)$ local terms, each of norm $O(1)$, then the energy variance $\sigma^2 = \text{tr}[H_{\text{mod}}^2 \Omega_A] - \text{tr}[H_{\text{mod}}\Omega_A]^2 \leq O(|\partial A|)$, which can be shown using the exponential decay of correlations in the Gibbs state of such Hamiltonians [Ara69, PGPH20]. It follows from the Chebyshev's inequality on the concentration of probability distributions that the δ-smooth entanglement spread of the spectrum of the Gibbs state satisfies $\text{ES}_\delta(\Omega_A) \leq O(\sqrt{|\partial A|}/\delta)$, yielding the same $O(\sqrt{|\partial A|})$ dependency as in bound (3.7). A more sophisticated argument in Corollary 1 of [KS20b] can improve this to $\text{ES}_\delta(\Omega_A) \leq O\left(\sqrt{|\partial A|}\log\frac{1}{\delta}\right)$.

Contracting PEPS with exact entanglement spread sub-area law: There is an interesting connection between entanglement spread and efficient algorithms for estimating the expectation value of local observables in 2D gapped ground states. We thank Itai Arad for suggesting this

argument. Suppose a ground state $|\Omega\rangle$ is given by its PEPS representation [VMC08] of bond dimension D. For any region A on the lattice, one can compute a subspace Π such that Ω_A is supported in Π. This can be achieved by computing $D^{|\partial A|}$ vectors and then finding their span [AAJ16]. This computation takes time polynomial in $D^{|\partial A|}$, which is efficient if A is of constant size and D is polynomial in the number of particles. If the PEPS is injective [PGVWC08], the support of Ω_A coincides with Π.

It is well known that $|\Omega\rangle$ satisfies an exponential decay of correlation with length $\xi = \mathcal{O}\left(\frac{1}{\gamma}\right)$ [Has04, NS06]. We further assume the *exact* entanglement spread in any region A is $c|\partial A|^{1-\kappa}$ for a constant c and a parameter $\kappa > 0$. This means that given the reduced state Ω_A, we have $e^{-c|\partial A|^{1-\kappa}} \leq \frac{\lambda_{\max}(\Omega_A)}{\lambda_{\min}(\Omega_A)} \leq e^{c|\partial A|^{1-\kappa}}$. Note that the assumption of a 'sub-area' scaling for the exact entanglement spread could be violated in a very simple example of a gapped local Hamiltonian, as given in the tightness argument earlier in this section. Extending our current result to the case of *smooth* entanglement spread, for which we establish a sub-area scaling, is an interesting open problem.

Consider a qudit and let O be an operator supported on it with $\|O\| \leq 1$. Let A be a ball of radius r around this qudit. Let d be the dimension of the projector Π for this region computed from the PEPS. We prove the following in Section 3.9.

Theorem 46. *Fix any $\epsilon \in (0, \frac{1}{2})$. Let $\langle O\rangle := \langle\Omega|O|\Omega\rangle$. It holds that*

$$\left|\frac{\operatorname{tr} O\Pi}{d} - \langle O\rangle\right| \leq \epsilon + \frac{1}{\epsilon}\cdot e^{30cr^{1-\kappa} - r/\xi}.$$

In particular, choosing $r = (60c\xi)^{\frac{1}{\kappa}} + 2\xi\log\frac{1}{\epsilon}$, we can efficiently estimate $\langle O\rangle$ with error 2ϵ by computing $\frac{\operatorname{tr} O\Pi}{d}$.

Implications for proving area law for the entanglement entropy: Theorem 44 shows that if one can prove an area law for $S_{\min}^{(4\sqrt{2}\Delta)^{2/3}}(\Omega_A)$, then this implies an area law for $S_{\max}^{(4\sqrt{2}\Delta)^{2/3}}(\Omega_A)$ and hence, for the entanglement entropy. In contrast, prior work [ALV12] (see Equation 3.10) would show that an area law for the min-entropy $S_{\min}(\Omega_A) = O(|\partial A|)$ leads to a sub-volume law $O(|\partial A|^{3/2})$ on the entanglement entropy. We cannot directly compare our result with this, as $S_{\min}(\Omega_A)$ is smaller than $S_{\min}^{(4\sqrt{2}\Delta)^{2/3}}(\Omega_A)$.

As mentioned earlier, we cannot replace $S_{\min}^{(4\sqrt{2}\Delta)^{2/3}}(\Omega_A)$ with $S_{\min}(\Omega_A)$ without changing our upper bound on the entanglement spread to $O(|\partial A|)$. Achieving such a bound is an interesting open problem since it would rigorously prove that min-entropy area law implies entanglement entropy area law. The utility of this is that min-entropy area law may be easier to prove in comparison to the entanglement entropy area law. For instance, for specific models such as stoquastic local Hamiltonians, proving min-entropy area law can be reduced to a classical problem [BDOT08]

Connection to the counter example to the area law in [AHL+14] Our setup is closest to [AHL+14], where the authors construct a family of gapped Hamiltonians whose ground states violate the entropy area law. This is done by connecting a protocol (without any EPR assistance) for testing maximally entangled states to the ground state of a local Hamiltonian using Kitaev's circuit-to-Hamiltonian construction. The obtained ground state admits a bipartition into parts A and B such that a single Hamiltonian term crosses the cut, but it enforces a maximally entangled

state $\Phi_{|A|}$ between A and B. This causes the entropy of part A to be $O(|A|)$ violating the entropy area law. Nevertheless, we see that this construction still satisfies our entanglement spread area law simply because the maximally entangled state has zero entanglement spread, and the ground state has entanglement spread at most $O(1)$. When combined with our previous discussion on area law, this loosely suggests the following: The ground state of a gapped Hamiltonian always exhibits a small entanglement spread. But it either has a large min-entropy (such as maximally entangled states in the counter-example Hamiltonian) hence not obeying an entropy area law, or it possesses small min-entropy (such as the 1D ground states) thus obeying an entropy area law.

Hamiltonian simulation by Trotterization: While we use the interaction picture Hamiltonian simulation algorithm, it would be interesting to achieve the same result by directly using the Trotterization method along the lines of [Ber07]. That is, we want to simulate the Hamiltonian $e^{-i\tau H_{\partial A}}$ for some small $\tau \approx 1/\|H\|$ with the communication cost $O(\tau\|H_{\partial A}\|)$. By repeating this step for $1/\tau = O(\|H\|)$ times, we obtain the desired overall scaling of $O(\|H_{\partial A}\|)$. The issue with naively using this approach is that each simulation step requires exchanging one qubit of communication resulting in a large communication complexity. Note, however, that the *entropy* of this exchanged qubit is $O(\tau)$. Hence, we expect the quantum information cost of this step [Tou15] to also be $O(\tau)$. We anticipate that performing quantum information theoretic *compression* on such a protocol would lead to a new protocol achieving the desired bound. Finally, note that recent results in [CST+21] achieve a similar bound for the clustered Hamiltonians by a tighter analysis of the Trotter error.

Compression of Schmidt rank: A by-product of our techniques is a compression tool for the Schmidt rank of any AGSP using EPR assistance. Since this might find other applications beyond our work, we formally state it in the following proposition.

Proposition 47. *Fix an AGSP*

$$K = \sum_i \alpha_i U_i \otimes V_i,$$

with $\alpha_i > 0$, $\|U_i\|, \|V_i\| \leq 1$, and U_i, V_i acting on subsystem A, B respectively. Suppose

$$\|K - |\Omega\rangle\langle\Omega|\| \leq \Delta.$$

Then there exists an EPR-assisted AGSP K' (as in Definition 40) with Schmidt rank $\left(\frac{\sum_i \alpha_i}{\Delta}\right)^{\mathcal{O}(1)}$ such that

$$\|(K' - \mathbb{1} \otimes |\Omega\rangle\langle\Omega|)\ |\Phi\rangle|\psi\rangle\| \leq 2\Delta \quad \textit{for all} \quad |\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B.$$

Hence, given an AGSP K, we can use Proposition 47 to construct an EPR-assisted AGSP with similar shrinking Δ but a Schmidt rank only polynomial in ℓ_1-norm of the coefficients α_i in the Schmidt decomposition of K.

3.5 Preliminaries

Local Hamiltonians: Let $\mathcal{S}$ be a collection of n spins, each with dimension s. The interactions between these spins are described by a local Hamiltonian $H = \sum_{k=1}^{N} h_k$ where the operators $0 \preceq$

$h_k \preceq \mathbb{1}$ act nontrivially only on at most κ spins. Let H_X denote the Hamiltonian restricted to region $X \subseteq \mathcal{S}$. For a bipartition $(A:B)$ of the set $\mathcal{S}$, we write $H = H_A + H_B + H_{\partial A}$, where $H_{\partial A}$ is the collection of interaction terms acting on both A and B. We denote the Hilbert space of these partitions and the whole system by $\mathcal{H}_A$, $\mathcal{H}_B$ and $\mathcal{H}_{AB}$ respectively.

We denote the spectrum of H by $E_0 \leq E_1 \leq \cdots \leq E_{\max}$. For convenience, we assume that $E_0 = 0$. Let $|\Omega\rangle$ be the *unique* ground state of H and $|E_1\rangle, |E_2\rangle, \ldots, |E_{\max}\rangle$ the other eigenstates. The spectral gap of the Hamiltonian H is a constant γ such that $E_1 = E_0 + \gamma$. We use the notation $\Omega := |\Omega\rangle\langle\Omega|$ and more generally $\Phi := |\Phi\rangle\langle\Phi|$ for any state $|\Phi\rangle$.

Communication protocols: In what follows, we consider quantum communication protocols between Alice and Bob. We assume, a bipartition $(A:B)$ of the set $\mathcal{S}$ is shared between the parties such that Alice has access to spins in region A while Bob has access to those in region B. Both parties also have their own additional registers.

The parties communicate by sending qubits, and can cooperate to implement an operator supported on $A \cup B$. The communication complexity of implementing such an operator is defined as the total number of exchanged qubits.

Two-party entanglement Given a state $|\psi\rangle_{AB}$ shared between Alice and Bob with the Schmidt coefficients $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$, the Rényi entropy of order α of the reduced state $\psi_A = \mathrm{tr}_B\, |\psi\rangle\langle\psi|$ is defined as

$$S_\alpha(\psi_A) = \frac{1}{1-\alpha} \log\left(\mathrm{tr}\, \psi_A^\alpha\right) = \frac{1}{1-\alpha} \log\left(\sum_{i=1}^{d} \lambda_i^\alpha\right), \quad 0 < \alpha < \infty. \tag{3.11}$$

Specifically for $\alpha = 0, 1, \infty$, we define the max- and min- entropies by $S_{\max}(\psi_A) = \log\left(\mathrm{rank}\left(\psi_A\right)\right)$ and $S_{\min}(\psi_A) = -\log \lambda_1$. The von Neumann entropy $S(\psi_A)$ is the limiting case of $\lim_{\alpha\to 1} S_\alpha(\psi_A) = -\,\mathrm{tr}[\psi_A \log \psi_A]$. In this paper, we mostly use a robust version of these entropies defined as follows.

Definition 48 (Smooth Rényi entropies and entanglement spread)**.** *Consider a state ρ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$. For $\delta \in (0,1)$, let*

$$r_\delta(\rho) = \{L \subseteq [d] : \sum_{i\in L} \lambda_i \geq 1 - \delta\}. \tag{3.12}$$

We define the δ-smooth max- and min- entropies of the state ρ by

$$S^\delta_{\max}(\rho) = \min_{L \in r_\delta(\rho)} \log |L|, \tag{3.13}$$

$$S^\delta_{\min}(\rho) = -\min_{L \in r_\delta(\rho)} \log\left(\max_{i\in L} \lambda_i\right). \tag{3.14}$$

The δ-smooth entanglement spread of the state ρ is defined as

$$\mathrm{ES}_\delta(\rho) = S^\delta_{\max}(\rho) - S^\delta_{\min}(\rho) \tag{3.15}$$

Lemma 49 (Young-Eckart theorem)**.** *Consider a bipartite state $|\psi\rangle \in \mathcal{H}_{AB}$ with the Schmidt coefficients $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$. Let $|\phi\rangle \in \mathcal{H}_{AB}$ be the state with the Schmidt rank $\leq r$ which has the largest overlap with $|\psi\rangle$. It holds that $|\langle\phi|\psi\rangle|^2 \leq \sum_{i=1}^r \lambda_i$.*

We will use the operator Schmidt decomposition in which an operator T acting on two systems is decomposed as $T = \sum_{i\in[r]} \lambda_i A_i \otimes B_i$ such that $\lambda_i > 0$ and $\mathrm{tr}[A_i^\dagger A_j] = \mathrm{tr}[B_i^\dagger B_j] = \delta_{ij}$. The number of terms r is called the Schmidt rank of T and satisfies an important submultiplicativity bound: $\mathrm{SR}(T_1T_2) \leq \mathrm{SR}(T_1)\mathrm{SR}(T_2)$. This includes the case when one operator is a state so that we have $\mathrm{SR}(T|\psi\rangle) \leq \mathrm{SR}(T)\mathrm{SR}(|\psi\rangle)$.

3.6 Approximate ground space projector and entanglement spread

We being by proving the connection between the entanglement spread and the communication complexity of testing bipartite states.

Definition 50 (EPR-assisted AGSP, restatement of Definition 40)**.** *Fix a bipartition $(A:B)$ of the spins, let $|\Omega\rangle \in \mathcal{H}_{AB}$ be the ground state of a local Hamiltonian H and $|\Phi\rangle = \frac{1}{\sqrt{p}} \sum_{j=1}^{p} |j\rangle_{A_0}|j\rangle_{B_0}$ be a maximally entangled state with dimension $p \geq 1$ shared between Alice and Bob who control A_0 and B_0 respectively. We say that an operator K is a (D,Δ)-EPR-assisted AGSP if*

- *The Schmidt rank of K is at most D and*
- *It holds that*

$$\|(K - \mathbb{1} \otimes |\Omega\rangle\langle\Omega|)\ |\Phi\rangle|\psi\rangle\| \leq \Delta \quad \text{for all} \quad |\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B. \tag{3.16}$$

Remark 51. *The Schmidt rank of the AGSP equals 2^c where c is the communication complexity of implementing it. We switch between D and c depending on which one is more convenient. Also, we use EPR-assistance only in our AGSP construction based on the quantum phase estimation and not the Chebyshev-AGSPs. Nevertheless, the following theorem applies generally to both cases.*

Theorem 52 (Bounding entanglement spread using AGSP)**.** *Suppose there exists a (D,Δ)-EPR-assisted AGSP with respect to a partition $(A:B)$ such that $\Delta < \frac{1}{4\sqrt{2}}$. Then the entanglement spread across $(A:B)$ is bounded by*

$$S_{\max}^{(4\sqrt{2}\Delta)^{2/3}}(\Omega_A) - S_{\min}^{(4\sqrt{2}\Delta)^{2/3}}(\Omega_A) \leq \log D + 1. \tag{3.17}$$

Proof. Let $\varepsilon = (2\Delta)^{2/3}$. Consider the Schmidt decomposition

$$|\Omega\rangle = \sum_i \sqrt{\lambda_i}|i\rangle_A|i\rangle_B,$$

where $\{\lambda_i\}$ are in descending order. Let b be the smallest integer such that $\varepsilon' := \sum_{i<b} \lambda_i \geq \varepsilon$ and define

$$|\Omega_{\text{heavy}}\rangle = \sum_{1\leq i<b} \sqrt{\frac{\lambda_i}{\varepsilon'}}|i\rangle_A|i\rangle_B,$$

$$|\Omega_{\text{light}}\rangle = \sum_{i\geq b} \sqrt{\frac{\lambda_i}{1-\varepsilon'}}|i\rangle_A|i\rangle_B.$$

Hence, $|\Omega\rangle = \sqrt{\varepsilon'}|\Omega_{\text{heavy}}\rangle + \sqrt{1-\varepsilon'}|\Omega_{\text{light}}\rangle$. We are also given the Schmidt decomposition of $|\Phi\rangle$ as

$$|\Phi\rangle = \frac{1}{\sqrt{p}} \sum_{j=1}^{p} |j\rangle_{A_0} |j\rangle_{B_0},$$

for some integer p. From the closeness of K to $|\Omega\rangle\langle\Omega|$ as in (3.16) and the identity $\langle\Omega|\Omega_{\text{heavy}}\rangle = \sqrt{\varepsilon'}$, we have

$$\|K|\Phi\rangle|\Omega_{\text{heavy}}\rangle - |\Phi\rangle \otimes |\Omega\rangle\langle\Omega|\Omega_{\text{heavy}}\rangle\| = \|K|\Phi\rangle|\Omega_{\text{heavy}}\rangle - \sqrt{\varepsilon'}|\Phi\rangle|\Omega\rangle\| \leq \Delta. \tag{3.18}$$

The Schmidt rank of $\frac{1}{\sqrt{\varepsilon'}}K$ is the same as K which equals D. The Schmidt rank of $|\Omega_{\text{heavy}}\rangle$ is $b-1$. Hence, the Schmidt rank of $\frac{1}{\sqrt{\varepsilon'}}K|\Phi\rangle|\Omega_{\text{heavy}}\rangle$ is at most $p(b-1)D$. From (3.18), we have that

$$\left| \langle\Phi|\langle\Omega| \frac{\left(\frac{1}{\sqrt{\varepsilon'}}K\right)|\Phi\rangle|\Omega_{\text{heavy}}\rangle}{\left\|\left(\frac{1}{\sqrt{\varepsilon'}}K\right)|\Phi\rangle|\Omega_{\text{heavy}}\rangle\right\|} \right| \geq 1 - 2\frac{\Delta}{\sqrt{\varepsilon'}}. \tag{3.19}$$

Following [ALV12], we use the Young-Eckart theorem (Lemma 49) along with the above bound. This implies that the sum of the largest $p(b-1)D$ eigenvalues of $\Omega_A \otimes \Phi_{A_0}$ is at least $(1 - \frac{2\Delta}{\sqrt{\varepsilon'}})^2$. However, since the eigenvalues of Φ_{A_0} are all equal to $\frac{1}{p}$, this sum is equal to $\sum_{i=1}^{(b-1)D} \lambda_i$. This is the key point in our proof where we use the fact that $|\Phi\rangle$ is maximally entangled; replacing it with a different state, such as an embezzling state, would cause this step to fail. Hence, we have

$$\sum_{i=1}^{(b-1)D} \lambda_i \geq (1 - \frac{2\Delta}{\sqrt{\varepsilon'}})^2 \geq 1 - \frac{4\Delta}{\sqrt{\varepsilon'}}.$$

From the definition of the smooth max-entropy (3.13), we see that this statement is equivalent to

$$S_{\max}^{\frac{4\Delta}{\sqrt{\varepsilon'}}}(\Omega_A) \leq \log D + \log(b-1).$$

Since $S_{\max}^{\frac{4\Delta}{\sqrt{\varepsilon}}}(\Omega_A) \leq S_{\max}^{\frac{4\Delta}{\sqrt{\varepsilon'}}}(\Omega_A)$, we conclude that

$$S_{\max}^{\frac{4\Delta}{\sqrt{\varepsilon}}}(\Omega_A) \leq \log D + \log(b-1).$$

Now, consider the following two cases:

1) $\varepsilon \geq \lambda_1$**:** From the definition of b,

$$\varepsilon' = \sum_{i<b} \lambda_i \leq \varepsilon + \lambda_b \leq 2\varepsilon.$$

Since λ_i are arranged in descending order, we also have $\varepsilon' \geq (b-1)\lambda_{b-1}$. This implies

$$\log(b-1) \leq \log \varepsilon' + \log \frac{1}{\lambda_{b-1}} \leq \log(2\varepsilon) + \log \frac{1}{\lambda_b}.$$

By Definition 48, $\log \frac{1}{\lambda_b} = S_{\min}^{\varepsilon'}(\Omega_A) \leq S_{\min}^{2\varepsilon}(\Omega_A)$. From this, we conclude that

$$S_{\max}^{\frac{4\Delta}{\sqrt{\varepsilon}}}(\Omega_A) - S_{\min}^{2\varepsilon}(\Omega_A) \leq \log D + \log(2\varepsilon) \leq \log D + 1.$$

2) $\varepsilon < \lambda_1$**:** In this case, $b = 2$. Thus, $S_{\max}^{\frac{4\Delta}{\sqrt{\varepsilon}}}(\Omega_A) \leq \log D$. Since $S_{\min}^{2\varepsilon}(\Omega_A) \geq 0$, we have

$$S_{\max}^{\frac{4\Delta}{\sqrt{\varepsilon}}}(\Omega_A) - S_{\min}^{2\varepsilon}(\Omega_A) \leq \log D \leq \log D + 1.$$

By plugging in the value of $\varepsilon = (2\Delta)^{2/3}$, we arrive at (3.17) which concludes the proof. □

Theorem 52 implies that we can bound the entanglement spread in the ground state by finding an appropriate AGSP. In the next sections, we achieve this using two distinct approaches. First in Section 3.7, we use the phase estimation algorithm to construct an AGSP for a gapped Hamiltonian on an arbitrary graph with $D = O(|\partial A|/\gamma)$. Next in Section 3.8, we find an AGSP using the Chebyshev polynomial with a quadratically improved scaling of $D = O(\sqrt{|\partial A|/\gamma})$.

3.7 AGSP for general graphs using quantum phase estimation

Here, we prove in detail how a distributed version of the quantum phase estimation can be used to test gapped ground states. This is done in three parts. In Section 3.7.1, we review the Hamiltonian simulation algorithm based on the interaction picture. Then in Section 3.7.2, we turn this algorithm into a communication protocol. Finally in Section 3.7.3, we combine the Hamiltonian simulation protocol with the quantum phase estimation to obtain an AGSP.

3.7.1 Hamiltonian simulation in the interaction picture

In this section, we describe a communication protocol between Alice and Bob that allows them to approximately implement the evolution operator $e^{-itH} = e^{-it(H_A+H_B+H_{\partial A})}$ using $\tilde{O}(\|H_{\partial A}\|t)$ qubits of communication. The conventional Hamiltonian simulation techniques work in the Schrödinger picture. Naively using these techniques results in communication complexity that scales with $\|H\|$ instead of $\|H_{\partial_A}\|$. To get around this issue, we instead use the recent Hamiltonian simulation algorithm in the interaction picture [LW18] along with the Linear Combination of Unitaries (LCU) method [BCC+15].

In the Hamiltonian simulation, the goal is to prepare the state $|\psi(t)\rangle = e^{-it(H_A+H_B+H_{\partial A})}|\psi(0)\rangle$ for any initial state $|\psi(0)\rangle$. This is conventionally done by directly implementing the unitary $e^{-it(H_A+H_B+H_{\partial A})}$. In the interaction picture, we work in the rotating frame $|\psi_I(t)\rangle := e^{it(H_A+H_B)}|\psi(t)\rangle$. There, the evolution of a time-independent Hamiltonian H is transformed to the evolution by a time-dependent Hamiltonian

$$\frac{\mathrm{d}}{\mathrm{d}t}|\psi_I(t)\rangle = -iH_I(t)|\psi_I(t)\rangle \tag{3.20}$$

$$H_I(t) = e^{it(H_A+H_B)}H_{\partial A}e^{-it(H_A+H_B)}. \tag{3.21}$$

We can divide the evolution of duration t to L shorter segments of length $\tau = t/L$. The state $|\psi(t)\rangle$ can be expressed in this picture by

$$|\psi(t)\rangle = e^{-it(H_A+H_B)}|\psi_I(t)\rangle = \left(e^{-i\tau(H_A+H_B)}\mathcal{T}\left[e^{-i\int_0^\tau H_I(s)ds}\right]\right)^L |\psi(0)\rangle, \tag{3.22}$$

where $\mathcal{T}\left[\exp\left(-i\int_0^\tau H_I(s)ds\right)\right]$ is the time-ordered propagator. One advantage of working in the interaction picture is that $\|H_I(t)\| = \|H_{\partial A}\|$. Hence, the cost of implementing the propagation operator $\mathcal{T}\left[\exp\left(-i\int_0^\tau H_I(s)ds\right)\right]$ scales with $\|H_{\partial A}\|$ instead of $\|H_A\|$.

Lemma 53 (cf. [LW18], Lemma 5)**.** *The time-ordered propagator can be written as*

$$\mathcal{T}\left[e^{-i\int_0^\tau H_I(s)ds}\right] = \lim_{M,K\to\infty}\sum_{k=0}^{K}\frac{(-i\tau)^k}{M^k}\sum_{0\le m_1<\cdots<m_k<M} H_I(m_k\tau/M)\cdots H_I(m_1\tau/M). \tag{3.23}$$

The order of the M, K limit and the speed of convergence will not matter to us since we will see that our communication cost is completely independent of M, K.

The boundary term $H_{\partial A}$ can be decomposed as a sum of unitary operators, i.e.,

$$H_{\partial A} = \sum_{j=1}^{J}\beta_j u_j^{(A)}\otimes u_j^{(B)},$$

where $u_j^{(A)}$ ($u_j^{(B)}$) acts on Alice's (Bob's) spins. We can always absorb the phase of β_j in $u_j^{(A)}$ and assume $\beta_j > 0$. Similarly, the interaction Hamiltonian is

$$H_I(m_k\tau/M) = \sum_{j=1}^{J}\beta_j(e^{im_kH_A\tau/M}u_j^{(A)}e^{-im_kH_A\tau/M})\otimes(e^{im_kH_B\tau/M}u_j^{(B)}e^{-im_kH_B\tau/M}).$$

By plugging this into (3.23), we see that the time-ordered propagator can be expressed as a linear combination of unitary operators. For convenience, we define a collective index set

$$I_{M,K} = \{(k, m_1, \ldots, m_k, j_1, \ldots, j_k) : 0 \le k \le K, 0 \le m_1 < \cdots < m_k < M, j_1, \ldots, j_k \in [J]\}.$$

For some $\ell = (k, m_1, \ldots, m_k, j_1, \ldots, j_k)$, define

$$\begin{aligned}
\alpha_\ell &= (\tau^k/M^k)\beta_{j_k}\ldots\beta_{j_1},\\
v_\ell^{(A)} &= (-i)^k(e^{im_kH_A\tau/M}u_{j_k}^{(A)}e^{-im_kH_A\tau/M})\ldots(e^{im_1H_A\tau/M}u_{j_1}^{(A)}e^{-im_1H_A\tau/M}),\\
v_\ell^{(B)} &= (e^{im_kH_B\tau/M}u_{j_k}^{(B)}e^{-im_kH_B\tau/M})\ldots(e^{im_1H_B\tau/M}u_{j_1}^{(B)}e^{-im_1H_B\tau/M}).
\end{aligned}$$

Note that

$$\lim_{M,K\to\infty}\sum_{\ell\in I_{M,K}}\alpha_\ell = \exp(\tau\sum_{j=1}^{J}\beta_j). \tag{3.24}$$

where J is the number of unitaries in the decomposition $H_{\partial A} = \sum_{j=1}^{J}\beta_j u_j^{(A)}\otimes u_j^{(B)}$. Using this

notation, Lemma 53 can be expressed as

$$\mathcal{T}\left[e^{-i\int_0^\tau H_I(s)ds}\right] = \lim_{M,K\to\infty} \sum_{\ell\in I_{M,K}} \alpha_\ell \; v_\ell^{(A)} \otimes v_\ell^{(B)}. \tag{3.25}$$

3.7.2 Communication protocol for Hamiltonian simulation

Our results regarding the ground-state entanglement and the communication complexity are information theoretic in nature. In particular, the running time or other algorithmic aspects of the tools we use, such as the Hamiltonian simulation, do not affect our conclusions. Here, we explain how we can use this observation to simplify the analysis of a part of our protocol.

In the LCU method, we express the Hamiltonian simulation operator $U(t)$ as sum of unitaries $U(t) \approx \sum_k \alpha_k u_k^{(A)} \otimes u_k^{(B)}$ for some choice of coefficients $\alpha_k \in \mathbb{R}$ and unitaries $u_k^{(A)}$ and $u_k^{(B)}$ that act on Alice and Bob's qudits respectively. Such decomposition as a sum of unitaries could come from the truncation of a Taylor expansion. To keep the running time efficient such Taylor expansions are truncated at low orders.

When Alice and Bob jointly implement the LCU algorithm, they need to prepare and share the ancillary state $|\alpha\rangle = (\sum_k \alpha_k)^{-1/2} \sum_k \sqrt{\alpha_k}|k\rangle_{A_0}|k\rangle_{B_0}$. Then, they proceed by applying the unitaries $u_k^{(A)} \otimes u_k^{(B)}$ conditioned on their register $|k\rangle$. Now suppose instead of truncating the expansions, we continue adding higher terms. Of course, the issue is that the number of coefficients α_k and thus, the communication cost of sharing $|\alpha\rangle$ and reflecting about $|\alpha\rangle$ also increases. On the other hand, we know that if instead of $|\alpha\rangle$, the parties share a maximally entangled state, the bound (3.5) on the entanglement spread remains intact. In other words, it is not the number of exchanged ancillary qubits in the protocol, but their entanglement spread that affects our final bound (3.5).

We fix this problem by modifying the LCU algorithm such that instead of the state $|\alpha\rangle$, Alice and Bob only share the maximally entangled state (or equivalently some number of EPR pairs). This state only needs to be shared once, which can be done outside the protocol, and then many reflections about it can be done with a cost independent of the size of the state. Now we can keep an unbounded number of terms in the expansions and avoid similar approximations in our protocol. This blows up the running time of these procedures, but maintains the communication complexity.

In the following we show how this can be achieved more formally. Our first step is to prove how using $O(\log(1/\varepsilon))$ qubits, Alice and Bob can perform a reflection about the maximally entangled state $|\Phi_p\rangle = \frac{1}{\sqrt{p}} \sum_{j=1}^p |j\rangle_{A_0}|j\rangle_{B_0}$ with an arbitrary dimension p up to an error ε. The proof follows from the EPR testing protocol (i.e. performing the two-outcome measurement $\{\Phi_p, \mathbb{1} - \Phi_p\}$) of [AHL+14].

Theorem 54 (Reflection about $|\Phi_p\rangle$, cf. [AHL+14])**.** *For any p and any $\varepsilon > 0$, there exists a protocol for performing $\mathbb{1} - 2\Phi_p$, the reflection about the maximally entangled state, using $O(\log(1/\varepsilon))$ qubits of communication.*

Next we focus on the Hamiltonian simulation protocol and analyze its communication cost.

Theorem 55 (Communication protocol for Hamiltonian simulation)**.** *There exists a communication protocol between Alice and Bob, summarized in Protocol 1, that uses a shared maximally entangled state $|\Phi_p\rangle$ for an arbitrarily large p and $O(t|\partial A|\log(t|\partial A|/\varepsilon))$ extra qubits of communication and*

implements an operator W_t such that

$$\left\| W_t |\Phi_p\rangle|\psi\rangle - |\Phi_p\rangle e^{-it(H_A+H_{\partial A}+H_B)}|\psi\rangle \right\| \leq \varepsilon. \tag{3.26}$$

Proof of Theorem 55. The evolution is divided into L segments of length τ as in (3.22). In each segment, the operator $e^{-i\tau(H_A+H_B)} = e^{-i\tau H_A} e^{-i\tau H_B}$ can be implemented without any communication. Thus we focus on the cost of performing the time-ordered propagator $\mathcal{T}\left[\exp\left(-i\int_0^\tau H_I(s)ds\right)\right]$.

Following [LW18, BCC+15], we use the LCU method to simulate the time-ordered operator given as a sum of unitaries in (3.25). In the original LCU algorithm, to implement a sum of unitaries such as $\sum_{\ell\in I_{M,K}} \alpha_\ell \, v_\ell^{(A)} \otimes v_\ell^{(B)}$. Alice and Bob need to share (and later reflect about) the state $\sum_{\ell\in I_{M,K}} \sqrt{\alpha_\ell}|\ell\rangle_{A_0}|\ell\rangle_{B_0}$. In general, sharing such a state results in extra entanglement spread between the parties. To avoid this, we modify the sum in (3.25) so that all α_ℓ are equal and Alice and Bob can instead use their shared maximally entangled state $|\Phi_p\rangle$ which has zero entanglement spread.

We achieve this by rounding off the coefficients α_ℓ to the nearest multiple of $k_\delta = 2^{-\lceil \log(\delta^{-1})\rceil}$ denoted by $\tilde{\alpha}_\ell$ such that $|\alpha_\ell - \tilde{\alpha}_\ell| \leq \delta \ll 1$. The choice of δ depends on M and K. In particular as $M, K \to \infty$, we have $\delta \to 0$. We can re-express the sum in (3.25) by repeating each term $v_\ell^{(A)} \otimes v_\ell^{(B)}$ for $\tilde{\alpha}_\ell/k_\delta$ times. This means for a fixed M, K, and δ, Alice and Bob wish to implement the sum

$$k_\delta \cdot \sum_{\ell=1}^{p} v_\ell^{(A)} \otimes v_\ell^{(B)} \tag{3.27}$$

with some extended set of indices ℓ with size $p \leq \sum_{\ell\in I_{M,K}} \lceil \alpha_\ell/k_\delta \rceil < e^{\tau\sum_{j=1}^{J}\beta_j}/k_\delta$ (using (3.24)). The simulation protocol consists of the following steps summarized in Protocol 1:

1. Alice and Bob perform the following operator on the state $|\psi\rangle$ and the maximally entangled state $|\Phi_p\rangle = \frac{1}{\sqrt{p}}\sum_{\ell=1}^{p}|\ell\rangle_{A_0}|\ell\rangle_{B_0}$ shared between them:

$$\begin{aligned} \mathrm{SEL} &= \mathrm{SEL}_{A_0A} \otimes \mathrm{SEL}_{B_0B} \\ \mathrm{SEL}_{A_0A} &= \sum_{\ell=1}^{p} |\ell\rangle\langle\ell|_{A_0} \otimes v_\ell^{(A)} \\ \mathrm{SEL}_{B_0B} &= \sum_{\ell=1}^{p} |\ell\rangle\langle\ell|_{B_0} \otimes v_\ell^{(B)}. \end{aligned}$$

   To implement this, Alice (Bob) applies the unitary $v_\ell^{(A)}$ ($v_\ell^{(B)}$) on their spins conditioned on register $|\ell\rangle_{A_0}$ ($|\ell\rangle_{B_0}$). Hence, the operator SEL can be implement by the parties only using local unitaries. Their state after this step is

$$\mathrm{SEL}\,|\Phi_p\rangle|\psi\rangle = \frac{1}{\sqrt{p}}\sum_{\ell=1}^{p} |\ell\rangle_{A_0}|\ell\rangle_{B_0} \otimes \left(v_\ell^{(A)} \otimes v_\ell^{(B)}\right)|\psi\rangle \tag{3.28}$$

2. Next, the parties implement the oblivious amplitude amplification subroutine [BCC+15] which for given any state $|\psi\rangle$, *turns* the state $\mathrm{SEL}\,|\Phi_p\rangle|\psi\rangle$ into the desired state

$|\Phi_p\rangle \left(\sum_\ell k_\delta \; v_\ell^{(A)} \otimes v_\ell^{(B)}\right) |\psi\rangle$ using a "Grover type" rotation. In our setting, this means the parties apply the rotation operator $-\mathrm{SEL}(2\Phi_p - \mathbb{1})\,\mathrm{SEL}^\dagger(2\Phi_p - \mathbb{1})$. It is shown in [BCC+15] that if $\sum_{\ell \in I_{M,K}} \tilde{\alpha}_\ell = 2$, one application of this operator suffices. It follows from Eq. (3.24) that in the limit of $M, K \to \infty$ and $\delta \to 0$, $\sum_{\ell \in I_{M,K}} \tilde{\alpha}_\ell = \exp(\tau \sum_{j=1}^J \beta_j)$. Hence, if $t \cdot \sum_{j=1}^J \beta_j$ is a multiple of $\ln(2)$, we can choose the number of segments as $L = t \sum_j \beta_j / \ln(2) = O\left(t|\partial A|\right)$ to ensure $\sum_{\ell \in I_{M,K}} \tilde{\alpha}_\ell = 2$. Otherwise, we set $L = \lceil t \sum_j \beta_j / \ln(2) \rceil$ and adjust the angle between the initial and target states in the last segment by appending an additional ancillary register $e^{i\theta X}|0\rangle$ with a suitably small θ.

Similar to step 1, the operators $-\mathrm{SEL}$ and $\mathrm{SEL}^\dagger$ are performed locally. The reflection operator $2\Phi_p - \mathbb{1}$ is performed using the protocol in Theorem 54. For an overall error of ε, we need error ε/L per segment, which requires $O\left(\log(L/\varepsilon)\right) = O\left(\log(t|\partial A|/\varepsilon)\right)$ qubits of communication per segment.

In the described protocol, we can take $M, K \to \infty$. By doing so $\delta \to 0$. This will only increase the size p of the shared maximally entangled state $|\Phi_p\rangle$ and not the qubits communicated when implementing $2\Phi_p - \mathbb{1}$.

The state after performing oblivious amplitude amplification is

$$\Big(-\mathrm{SEL}(2\Phi_p - \mathbb{1})\,\mathrm{SEL}^\dagger(2\Phi_p - \mathbb{1})\Big)\,\mathrm{SEL}\,|\Phi_p\rangle|\psi\rangle \approx_{\varepsilon/L} |\Phi_p\rangle \left(\sum_{\ell'} k_\delta \; v_{\ell'}^{(A)} \otimes v_{\ell'}^{(B)}\right) |\psi\rangle. \quad (3.29)$$

3. Alice performs $e^{-i\tau H_A}$ and Bob performs $e^{-i\tau H_B}$ on their spins.

4. They repeat the steps 1-3 for $L = O\left(t|\partial A|\right)$ times.

Hence, the number of qubits exchanged during the whole protocol is $O(t|\partial A| \log(t|\partial A|/\varepsilon))$. □

Protocol 1 Hamiltonian simulation protocol between Alice and Bob

---

Input: Unbounded shared maximally entangled state $|\Phi\rangle$, a shared state $|\psi\rangle$, H and t.
Goal: Implement W_t such that $\|W_t|\Phi\rangle|\psi\rangle - |\Phi\rangle e^{-itH}|\psi\rangle\| \leq \varepsilon$.
Procedure:
For $L = O\left(t|\partial A|\right)$ times, perform the following protocol:

1. Alice and Bob implement $\mathrm{SEL}\,|\Phi\rangle|\psi\rangle$ by applying local controlled-unitaries,

2. Using the protocol in Theorem 54, Alice and Bob approximately perform the rotation operator $\big(-\mathrm{SEL}(2\Phi - \mathbb{1})\,\mathrm{SEL}^\dagger(2\Phi - \mathbb{1})\big)$ using $O\left(\log(t|\partial A|/\varepsilon)\right)$ qubits of communication. The parties use $O(1)$ qubits of communication in the L'th iteration to adjust the angle between states in the oblivious amplitute amplification subroutine (see step 2 of the proof of Theorem 55).

3. Alice applies $e^{-iH_A t/L}$ and Bob applies $e^{-iH_B t/L}$ locally.

3.7.3 A communication protocol for measuring the ground state

In this section, we use the Hamiltonian simulation protocol of Section 3.7.1 along with the quantum phase estimation algorithm to implement an approximate version of the measurement $\{\Omega, \mathbb{1} - \Omega\}$.

The phase estimation algorithm is an operator PHASE that is implemented on some target state $|\psi\rangle$ in the following steps. (1) Starting with $f = \log(1/(\gamma t_0)) + O(1)$ ancillary qubits in the $|0\rangle$ state, a uniform superposition $\frac{1}{\sqrt{2^f}} \sum_{j=0}^{2^f-1} |jt_0\rangle$ is created. Here, the parameter t_0 is chosen such that $|\partial A| t_0 \ll 1$ and following the discussion in the beginning of Section 3.7.2, we simply consider the limit $t_0 \to 0$ at the end. Hence, this register is a uniform superposition over time steps $\{0, t_0, 2t_0, \ldots, O(1/\gamma)\}$. (2) Controlled on the register $|jt_0\rangle$, the Hamiltonian simulation operator $e^{ijt_0 H}$ is performed on the target state $|\psi\rangle$. (3) Finally, an inverse Fourier transform is applied to the ancillary registers. The action of this operator on the eigenstates of H is (assuming that the e^{-itH} oracle is perfect):

$$\begin{aligned} \text{PHASE } |0\rangle|\Omega\rangle &= |0\rangle|\Omega\rangle \\ \text{PHASE } |0\rangle|E_i\rangle &= \sqrt{\sigma_i}|0\rangle|E_i\rangle + \sqrt{1-\sigma_i}|0_i^\perp\rangle|E_i\rangle, \quad \forall i \geq 1, \end{aligned}$$

where the $|\sigma_i|$ are all less than some universal constant $\sigma < 1$ and the states $|0_i^\perp\rangle$ are all orthogonal to $|0\rangle$. By repeating this operator for $O(\log(1/\Delta))$ times in parallel, we can reduce the error to $\leq \Delta$. Let us define the two-outcome POVM $\{K, \mathbb{1} - K\}$ such that

$$K = \text{PHASE}^\dagger (|0\rangle\langle 0| \otimes \mathbb{1}) \, \text{PHASE}\,.$$

We see that $\{K, \mathbb{1} - K\}$ provides a good approximation to the measurement $\{\Omega, \mathbb{1} - \Omega\}$ that we intend to perform [CSS18, Equation 10-12]:

$$\| (K - \mathbb{1} \otimes |\Omega\rangle\langle\Omega|) \, |0\rangle|\psi\rangle \| \leq \Delta. \tag{3.30}$$

In our EPR-assisted communication protocol for implementing the PHASE operator, the Hamiltonian simulation is implemented approximately, hence introducing an additional error in (3.30). Moreover, we modify steps (1) and (3) in the above description of the phase estimation algorithm to replace the ancillary registers with the available shared EPR pairs. In step (1), the parties use their shared EPR states instead of preparing a uniform superposition starting from $|0\rangle$. In Step (3), instead of performing the inverse Fourier transform and the measurement $\{|0\rangle\langle 0|, \mathbb{1} - |0\rangle\langle 0|\}$, Alice and Bob jointly perform the equivalent EPR testing measurement $\{\Phi, \mathbb{1} - \Phi\}$ as in [AHL+14] (see also Theorem 54). In the following, we give the details of this protocol (Protocol 2) and its analysis which shows how to implement the two-outcome measurement $\{K, \mathbb{1} - K\}$.

Theorem 56 (Communication protocol for measuring the ground state)**.** *Following Protocol 2, Alice and Bob can implement a measurement $\{K, \mathbb{1} - K\}$ such that*

$$\| (K - \mathbb{1} \otimes |\Omega\rangle\langle\Omega|) \, |0\rangle|\Phi\rangle|\psi\rangle \| \leq \Delta$$

while sharing unlimited EPR pairs $|\Phi\rangle$ and with the communication cost

$$O\left(\frac{|\partial A|}{\gamma} \log \left(\frac{|\partial A|}{\gamma \Delta} \log \frac{1}{\Delta} \right) \log \frac{1}{\Delta} \right). \tag{3.31}$$

Proof. The phase estimation algorithm PHASE is repeated $O(\log(1/\Delta))$ times. In each application, the Hamiltonian simulation Protocol 1 is run once for time up to $O(|\partial A|/\gamma)$. According to Theorem 55, the communication cost of implementing Protocol 1 in each round is bounded by

Protocol 2 Protocol for measuring the ground state

Input: Unbounded shared maximally entangled state $|\Phi\rangle$, ancillary state $|0\rangle$, a shared state $|\psi\rangle$, and the Hamiltonian H.
Goal: Perform POVM $\{K, \mathbb{1} - K\}$ such that $\| (K - \mathbb{1} \otimes \mathbb{1} \otimes |\Omega\rangle\langle\Omega|) |0\rangle|\Phi\rangle|\psi\rangle \| \leq \Delta$.
Procedure:

1. For $k = O(\log(1/\Delta))$ times, repeat the following steps i.-iii. to perform the operator PHASE k times in parallel:
   i. Using their unlimited EPR pairs available, Alice and Bob share the state $|\Phi_f\rangle = \frac{1}{2^{f/2}} \sum_{j=0}^{2^f-1} |jt_0\rangle_{a_k} |jt_0\rangle_{b_k}$ for $f = \log(1/(\gamma t_0)) + O(1)$ and some $t_0 \to 0$.
   ii. Conditioned on registers $|j\rangle_{a_k}|j\rangle_{b_k}$, the parties implement the Hamiltonian simulation protocol W_j (Protocol 1) on the state $|\psi\rangle$.
   iii. Using additional ancillary states, they jointly implement the EPR testing protocol of [AHL+14] that approximately implements the two-outcome measurement $\{\Phi_f, \mathbb{1} - \Phi_f\}$ on registers $a_k \otimes b_k$.
2. If all the k EPR tests in step iii. accept, the parties accept (the outcome corresponding to measuring K), otherwise they reject (the outcome corresponding to measuring $\mathbb{1} - K$).

$O\left(\frac{|\partial A|}{\gamma} \log\left(\frac{|\partial A|}{\gamma\varepsilon}\right)\right)$. Additionally, the EPR testing in step iii. of Protocol 2 requires a communication cost of $O(\log(1/\varepsilon))$ to achieve an error of $O(\varepsilon)$. In order to have an overall error of $O(\Delta)$ in the phase estimation, it suffices choose

$$\varepsilon = O\left(\frac{\Delta}{\log(1/\Delta)}\right).$$

Adding these costs, we get (3.31). Note that taking $t_0 \to 0$ in Protocol 2 only blows up the local running time of Alice and Bob who call the Hamiltonian simulation oracle conditioned on $|jt_0\rangle$. □

3.8 AGSP for lattice Hamiltonians using Chebyshev polynomials

The construction of an AGSP based on the Chebyshev polynomials relies heavily on the truncation of the Hamiltonian. The idea of truncation, introduced in [AKLV13], allows one to control the norm of Hamiltonian away from a bipartite cut. In this section we (1) review the previous techniques for truncation in frustration-free and general Hamiltonians and (2) adapt them from 1D systems to an arbitrary lattice. First, we explain how to perform truncation in the frustration-free case.

3.8.1 Truncation: frustration-free case

Without loss of generality, we assume that h_k are projectors and the ground energy $E_0 = 0$. Let DL be the detectability lemma operator [AALV09] corresponding to H defined as follows.

Definition 57 (Detectability lemma operator)**.** *Partition the terms of the Hamiltonian* $H = \sum_{i=k}^{N} h_k$ *into* w *groups* $\{T_1, \ldots T_w\}$*, where the terms in each group mutually commute. For a finite*

dimensional lattice, w is a constant. The detectability lemma operator is defined by

$$\mathrm{DL} = \prod_{\alpha=1}^{w} \left(\prod_{k \in T_\alpha} (\mathbb{1} - h_k) \right).$$

The operator DL defines an AGSP for the Hamiltonian H. In particular, since H is frustration-free, the terms $\mathbb{1} - h_k$ preserve the ground state and we have $\mathrm{DL}\,|\Omega\rangle = |\Omega\rangle$. Since $\|\,\mathrm{DL}\,\| \leq 1$, when this operator is applied to the states orthogonal to $|\Omega\rangle$, their norm shrinks by a factor ≤ 1. More precisely, it is shown in [AALV09, AAV16] that we have

$$\mathrm{DL}\,|\Omega\rangle = |\Omega\rangle, \quad \|\,\mathrm{DL} - |\Omega\rangle\langle\Omega|\| \leq \frac{1}{1+\gamma/g^2}. \tag{3.32}$$

where γ is the spectral gap and g is the number of interactions in the Hamiltonian not commuting with a given interaction h_k.

Remark 58. *For a D-dimensional lattice with a κ-local Hamiltonian, we have $w \leq (2D)^{2\kappa}$ and $g \leq \kappa(2D)^{\kappa-1}$ (see for instance, [AAV16, Section II])*

In order to obtain a truncation for the Hamiltonian H, which as we explain later allows us to control the norm of Hamiltonian away from a bipartite cut, we consider a slightly different AGSP than DL. Consider any bipartition $(A:B)$ of the lattice. Let Π_A be the projector onto the ground space of the Hamiltonian H_A and Π_B be the projector onto the ground space of H_B. Let C be the set of all interactions contained within $\partial_w A$. Using the "absorption" argument from [AALV09, ALV12], the following equality can be shown:

$$(\Pi_A \Pi_B)\,\mathrm{DL} = \Pi_A \Pi_B \prod_{\alpha=1}^{w} \left(\prod_{k \in T_\alpha \cap C} (\mathbb{1} - h_k) \right), \tag{3.33}$$

which can be verified by noticing that $(\mathbb{1} - h_k)\Pi_R = \Pi_R$, where Π_R is the projector onto the ground space of a region R on which h_k is supported, we can absorb terms from DL into $\Pi_A\Pi_B$ except for those that are hindered due to the boundary.

By applying (3.33) in (3.32), and using $\Pi_A\Pi_B|\Omega\rangle = |\Omega\rangle$, we conclude that $(\Pi_A\Pi_B)\,\mathrm{DL}$ is also an AGSP, i.e.

$$\left\| \Pi_A \Pi_B \prod_{\alpha=1}^{w} \left(\prod_{k \in T_\alpha \cap C} (\mathbb{1} - h_k) \right) - |\Omega\rangle\langle\Omega| \right\| \leq \frac{1}{1+\gamma/g^2}. \tag{3.34}$$

Next, we use this operator to truncate the Hamiltonian H outside some region A.

Definition 59 (Truncated Frustration-free Hamiltonian)**.** *The truncation of a frustration-free Hamiltonian H with respect to the partition $(A:B)$ is defined by:*

$$\tilde{H} = \sum_{\alpha, k \in T_\alpha \cap C} h_k + (\mathbb{1} - \Pi_A) + (\mathbb{1} - \Pi_B), \tag{3.35}$$

Theorem 60 (Truncation in the frustration-free case)**.** *The truncated Hamiltonian $\tilde{H}$ has the follownng properties:*

1. *$\tilde{H}$ is frustration free*

2. *$\|\tilde{H}\| \leq 2 + w^2|\partial A|$*

3. *the spectral gap of $\tilde{H}$ is $\geq \gamma/4g^2$*

Proof. One can see that $|\Omega\rangle$ is a ground state of $\tilde{H}$. In order to lower bound the spectral gap of $\tilde{H}$, we use the fact that the detectability lemma operator and the Hamiltonian have very similar spectral gaps. This was described as a converse to the detectability lemma in [AAV16]. More precisely, using Theorem 1.1b of [Gao15], we obtain that for any state $|\psi\rangle$,

$$4\langle\psi|\tilde{H}|\psi\rangle \geq 1 - \left\| \Pi_A \Pi_B \prod_{\alpha=1}^{w} \left(\prod_{k \in T_\alpha \cap C} (\mathbb{1} - h_k) \right) |\psi\rangle \right\|^2 .$$

If $|\psi\rangle$ is orthogonal to $|\Omega\rangle$, (3.34) ensures that

$$4\langle\psi|\tilde{H}|\psi\rangle \geq 1 - \left(\frac{1}{1+\gamma/g^2}\right)^2 \geq \frac{\gamma}{g^2}.$$

Thus the spectral gap of $\tilde{H}$ is at least $\frac{\gamma}{4g^2}$. Finally we can bound the norm with

$$\|\tilde{H}\| \leq \|\mathbb{1} - \Pi_A\| + \|\mathbb{1} - \Pi_B\| + \sum_{\alpha, k \in T_\alpha \cap C} \|h_k\| \leq 2 + w^2|\partial A|$$

□

3.8.2 Truncation: frustrated case

Here, we consider truncation in the more general case of frustrated Hamiltonians. This is first achieved in [AKLV13]. We will directly use the following theorem from [KS20a], which built upon [AKL16]. For a partition $(A : B)$, the truncation in [AKL16] is defined by removing the high energy spectrum of H_B. The improvement in [KS20a] allows one to truncate both H_A and H_B while leaving the boundary term $H_{\partial_w A}$ untouched, where $w = O(1)$.

Definition 61 (Truncation of H up to energy ξ)**.** *Fix a bipartition $(A : B)$ such that $H = H_A + H_B + H_{\partial_w A}$ and $w = O(1)$. Let $\Pi_A^{<\xi}$ and $\Pi_A^{\geq\xi}$ denote the projectors onto the eigenstates of H_A with energy $< \xi$ and $\geq \xi$ respectively. Similarly, we assign $\Pi_B^{<\xi}$ and $\Pi_B^{\geq\xi}$ for region B. The truncation of H_A (or H_B) up to energy ξ is defined as*

$$\tilde{H}_A = H_A \Pi_A^{<\xi} + \xi \Pi_A^{\geq\xi}.$$

Moreover, the truncation of H up to energy ξ with respect to the partition $(A : B)$ is defined by

$$\tilde{H} = \tilde{H}_A + \tilde{H}_B + H_{\partial_w A}.$$

Theorem 62 (Truncation in the frustrated case, cf. [KS20a], Theorem 5)**.** *Let $\tilde{H}$ be the truncation of the Hamiltonian H with respect to the partition $(A : B)$ up to energy $\xi = O\left(w^2|\partial_w A| + \log\frac{1}{\gamma}\right)$, where $w = O(1)$. Then, it holds that*

i. The spectral gap of $\tilde{H}$ is at least $\frac{\gamma}{2}$,

ii. The ground state $|\Omega'\rangle$ of $\tilde{H}$ satisfies $|\langle\Omega|\Omega'\rangle| \geq 1 - e^{-0.1|\partial_w A|}$,

iii. $\|\tilde{H} - E_0'\mathbb{1}\| \leq 100w^2|\partial_w A|$, where E_0' is the ground-state energy of $\tilde{H}$.

3.8.3 Chebyshev-AGSP

The previous subsections show that the Hamiltonian H can be truncated to $\tilde{H}$ such that the spectral gap stays $\geq \gamma/2$ and the norm of the Hamiltonian is at most $E_0' + O(w^2|\partial_w A|)$. Furthermore, the ground state $|\Omega'\rangle$ is close to $|\Omega\rangle$ with fidelity at least $1 - e^{-0.1|\partial_w A|}$. This means we can instead construct an AGSP for the ground state $|\Omega'\rangle$ of the truncated Hamiltonian $\tilde{H}$.

Definition 63 (Chebyshev-AGSP)**.** *Define the Chebyshev-AGSP as the following polynomial of $\tilde{H}$ of degree q:*

$$Q_q(\tilde{H}) = \frac{1}{T_q\left(1 + \frac{2E_1' - 2E_0'}{E_{\max}' - E_1'}\right)} T_q\left(1 + \frac{2E_1' - 2\tilde{H}}{E_{\max}' - E_1'}\right),$$

where $E_{\max}'$ is the largest eigenvalue of $\tilde{H}$ and T_q is the degree-q Chebyshev polynomial of the first kind defined by $T_q(\cos\theta) = \cos(q\theta)$.

The reason for this definition is that (i) $Q_q(E_0') = 1$, (ii) the eigenvalues in the range $[E_1', E_{\max}']$ are sent by $x \mapsto 1 + \frac{2E_1' - 2x}{E_{\max}' - E_1'}$ to the range $[-1, 1]$ and $T_q([-1, 1]) = [-1, 1]$, (iii) the denominator is exponentially large in $q^2\left(\frac{2E_1' - 2E_0'}{E_{\max}' - E_1'}\right)$. Thus it yields an AGSP with properties given by the following theorem.

Theorem 64. *There is a constant ξ depending on the geometry of the Hamiltonian such that if we let $q = \sqrt{w^2 g^2 \frac{|\partial_w A|}{\xi^2 \gamma}} \log \frac{4}{\Delta}$, then $Q_q(\tilde{H})$ is a (D, Δ)-AGSP with respect to the partition $(A : B)$ (see Definition 50). That is, $\||\Omega\rangle\langle\Omega| - Q_q(\tilde{H})\| \leq \Delta$ and the Schmidt rank is bounded by*

$$D = \exp\left(\sqrt{\frac{|\partial_w A|}{\gamma}} \cdot \log\frac{4}{\Delta} \cdot \frac{wg}{c} \cdot \log\left(\frac{|\partial_w A|^2}{\gamma} w^2 g^2 s^b \log^2\left(\frac{4}{\Delta}\right)\frac{1}{\xi^2}\right)\right), \tag{3.36}$$

Before stating the proof of Theorem 64, we need the following lemma:

Lemma 65 (Adapted from [AKLV13])**.** *The Schmidt rank of $Q_q(\tilde{H})$ is at most $D \leq e^{q \log\left(q^2 s^b |\partial_w A|\right)}$.*

Proof. First, $\mathrm{SR}(Q_q(\tilde{H})) \leq q\mathrm{SR}(\tilde{H})^q$. Thus, we upper bound $\mathrm{SR}(\tilde{H})$. In both the frustration-free ((3.35)) and the frustrated case (Theorem 62), we can write $\tilde{H} = H_{\partial_w A} + X + Y$, where X is an operator supported on region A and Y is an operator acting on region B. Consider the following expansion:

$$(\tilde{H})^q = \sum_{\ell=1}^{q+1} \left(X^{f_1} Y^{g_1}\right) H_{\partial_w A} \left(X^{f_2} Y^{g_2}\right) H_{\partial_w A} \left(X^{f_3} Y^{g_3}\right) \ldots H_{\partial_w A} \left(X^{f_\ell} Y^{g_\ell}\right).$$

In each term, $H_{\partial_w A}$ occurs $\ell - 1$ times, and the tuple of non-negative integers $(f_1, g_1, \dots f_\ell, g_\ell)$ satisfies

$$\sum_{i=1}^{\ell} (f_i + g_i) = q - \ell + 1.$$

The number of possible such tuples is equal to $\binom{q+\ell+1}{2\ell} \leq \binom{2q+2}{2\ell} \leq (q+1)^\ell$. Since none of $X^{f_i} Y^{g_i}$ change the Schmidt rank across the bipartition, and $H_{\partial_w A}$ changes the Schmidt rank by at most $s^b |\partial_w A|$, we obtain that the Schmidt rank of $(\tilde{H})^q$ is at most

$$(s^b |\partial_w A|)^q \cdot (q+1)^{q+2} \leq e^{q \log(q^2 s^b |\partial_w A|)}.$$

This completes the proof. □

Proof of Theorem 64. By a result of [AKLV13], we have

$$\|Q_q(\tilde{H}) - |\Omega'\rangle\langle\Omega'|\| \leq 2e^{-2q\sqrt{\frac{E'_1 - E'_0}{E'_{\max} - E'_1}}} \leq 2e^{-\xi q \sqrt{\frac{\gamma}{w^2 g^2 |\partial_w A|}}},$$

where ξ is an absolute constant determined by the lattice structure and the locality of the Hamiltonian. For a given Δ, choose $q = \sqrt{w^2 g^2 \frac{|\partial_w A|}{\xi^2 \gamma}} \log \frac{4}{\Delta}$. Then

$$\||\Omega\rangle\langle\Omega| - Q_q(\tilde{H})\| \leq \Delta.$$

By plugging in the choice of $q = \sqrt{w^2 g^2 \frac{|\partial_w A|}{\xi^2 \gamma}} \log \frac{4}{\Delta}$ in Lemma 65, we see that the Schmidt rank is at most

$$D = \exp\left(\sqrt{\frac{|\partial_w A|}{\gamma}} \cdot \log \frac{4}{\Delta} \cdot \frac{wg}{\xi} \cdot \log\left(\frac{|\partial_w A|^2}{\gamma} w^2 g^2 s^b \log^2(\frac{4}{\Delta}) \frac{1}{\xi^2}\right)\right).$$

□

3.9 Proof of Theorem 46

Theorem 66 (Theorem 46 restated)**.** *Fix any $\epsilon \in (0, \frac{1}{2})$. Let $\langle O\rangle := \langle\Omega|O|\Omega\rangle$. It holds that*

$$\left|\frac{\operatorname{tr} O\Pi}{d} - \langle O\rangle\right| \leq \epsilon + \frac{1}{\epsilon} \cdot e^{30cr^{1-\kappa} - r/\xi}.$$

In particular, choosing $r = (60c\xi)^{\frac{1}{\kappa}} + 2\xi \log \frac{1}{\epsilon}$, we can efficiently estimate $\langle O\rangle$ with error 2ϵ by computing $\frac{\operatorname{tr} O\Pi}{d}$.

Proof. Divide the spectrum of Ω_A into $N = \frac{2c(|\partial A|)^{1-\kappa}}{\epsilon}$ blocks, such that the ratio between smallest and largest eigenvalues in each block is $\leq 1 + \epsilon$. Let Π_i be the projector onto the ith block, d_i be the dimension of Π_i and $\lambda^i_{\min}$ be the smallest eigenvalue in the ith block. Since $|\Omega\rangle$ is a pure state, there is a projector Π'_i acting on A^c, such that $(\Pi_i \otimes \mathrm{id}_{A^c})|\Omega\rangle = (\mathrm{id}_A \otimes \Pi'_i)|\Omega\rangle$.

Using decay of correlation, we can write for each block i,

$$|\langle\Omega|(\Pi'_i \otimes O)|\Omega\rangle - \langle\Omega|\Pi'_i|\Omega\rangle\langle\Omega|O|\Omega\rangle| \leq e^{-r/\xi}.$$

Note that $\langle\Omega|(\Pi_i' \otimes O)|\Omega\rangle = \mathrm{tr}(O\Pi_i\Omega_A)$ and $\langle\Omega|\Pi_i'|\Omega\rangle = \mathrm{tr}(\Pi_i\Omega_A)$. Thus, we can rewrite the above equation as

$$\left|\frac{\mathrm{tr}(O\Pi_i\Omega_A)}{\mathrm{tr}(\Pi_i\Omega_A)} - \langle O\rangle\right| \leq \frac{e^{-r/\xi}}{\mathrm{tr}(\Pi_i\Omega_A)}. \tag{3.37}$$

Since all the eigenvalues in Π^i are within a multiplicative factor $1+\epsilon$ of $\lambda^i_{\min}$, we have

$$\lambda^i_{\min}\Pi_i \preceq \Pi_i\Omega_A \preceq (1+\epsilon)\lambda^i_{\min}\Pi_i,$$

and hence

$$(1-\epsilon)\frac{\mathrm{tr}(O\Pi_i)}{\mathrm{tr}(\Pi_i)} \leq \frac{\mathrm{tr}(O\Pi_i\Omega_A)}{\mathrm{tr}(\Pi_i\Omega_A)} \leq (1+\epsilon)\frac{\mathrm{tr}(O\Pi_i)}{\mathrm{tr}(\Pi_i)}.$$

This allows us to rewrite Equation 3.37 as (also using $d_i = \mathrm{tr}(\Pi_i)$)

$$\left|\frac{\mathrm{tr}(O\Pi_i)}{d_i} - \langle O\rangle\right| \leq \epsilon\frac{\mathrm{tr}(O\Pi_i)}{d_i} + \frac{e^{-r/\xi}}{\mathrm{tr}(\Pi_i\Omega_A)}. \tag{3.38}$$

Now, we sum both sides over i, with weights $\frac{d_i}{d}$. Since $\Pi = \sum_i \Pi_i$, we obtain

$$\left|\frac{\mathrm{tr}(O\Pi)}{d} - \langle O\rangle\right| \leq \epsilon\frac{\mathrm{tr}(O\Pi)}{d} + e^{-r/\xi}\sum_i \frac{d_i}{d\,\mathrm{tr}(\Pi_i\Omega_A)} \leq \epsilon + e^{-r/\xi}\sum_i \frac{1}{d\lambda^i_{\min}}. \tag{3.39}$$

Using the entanglement spread condition, for each i we have $d\lambda^i_{\min} \geq e^{-c|\partial A|^{1-\kappa}} d\lambda_{\max}(\Omega_A) \geq e^{-c|\partial A|^{1-\kappa}}$. Since $|\partial A| \leq 10r$, we obtain

$$\begin{aligned}\left|\frac{\mathrm{tr}(O\Pi)}{d} - \langle O\rangle\right| &\leq \epsilon + e^{-r/\xi}Ne^{c(10r)^{1-\kappa}} \leq \epsilon + e^{-r/\xi}\cdot\frac{2c(|\partial A|)^{1-\kappa}}{\epsilon}\cdot e^{c(10r)^{1-\kappa}} \\ &\leq \epsilon + e^{-r/\xi}\cdot\frac{1}{\epsilon}\cdot e^{30cr^{1-\kappa}}.\end{aligned} \tag{3.40}$$

This completes the proof. □

3.10 Proof of Proposition 47

Proposition 67 (Proposition 47 restated)**.** *Fix an AGSP*

$$K = \sum_i \alpha_i U_i \otimes V_i,$$

with $\alpha_i > 0$, $\|U_i\|, \|V_i\| \leq 1$, and U_i, V_i acting on subsystem A, B respectively. Suppose

$$\|K - |\Omega\rangle\langle\Omega|\| \leq \Delta.$$

Then there exists an EPR-assisted AGSP K' (as in Definition 40) with Schmidt rank $\left(\frac{\sum_i \alpha_i}{\Delta}\right)^{\mathcal{O}(1)}$ such that

$$\|(K' - \mathbb{1}\otimes|\Omega\rangle\langle\Omega|)\,|\Phi\rangle|\psi\rangle\| \leq 2\Delta \quad \textit{for all} \quad |\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B.$$

Proof. Suppose $\alpha_i = \frac{w_i}{N}$ are rational numbers, with w_i and N nonnegative integers. This can be assumed with arbitrarily small error. We re-write

$$K = \frac{1}{N} \sum_i \sum_{j=1}^{w_i} U_i \otimes V_i.$$

Let $M = \sum_i w_i$. Introduce a maximally entangled state

$$|\Phi\rangle_{ab} = \frac{1}{\sqrt{M}} \sum_i \sum_{j=1}^{w_i} |i,j\rangle_a |i,j\rangle_b,$$

where Alice's and Bob's auxiliary registers are denoted by a and b respectively. This leads to the following representation of K:

$$|\Phi\rangle\langle\Phi|_{ab} \otimes K = \frac{M}{N} \left(|\Phi\rangle\langle\Phi|_{ab} \left(\sum_{i,j} |i,j\rangle\langle i,j|_a \otimes U_i \right) \otimes \left(\sum_{i,j} |i,j\rangle\langle i,j|_b \otimes V_i \right) |\Phi\rangle\langle\Phi|_{ab} \right).$$

From Theorem 54, there exists an operator L with Schmidt rank $\frac{1}{\epsilon^{\mathcal{O}(1)}}$, such that

$$\|L - |\Phi\rangle\langle\Phi|_{ab}\| \leq \epsilon.$$

Letting $\epsilon = \frac{N\Delta}{M}$, we obtain the following approximation to $K \otimes |\Phi\rangle\langle\Phi|_{ab}$:

$$K' = \frac{M}{N} \left(L \left(\sum_{i,j} |i,j\rangle\langle i,j|_a \otimes U_i \right) \otimes \left(\sum_{i,j} |i,j\rangle\langle i,j|_b \otimes V_i \right) |\Phi\rangle\langle\Phi|_{ab} \right)$$

such that for all $|\psi\rangle$,

$$\|K' - |\Phi\rangle\langle\Phi|_{ab} \otimes K\| \leq \Delta \implies \| \left(K' - \mathbb{1} \otimes |\Omega\rangle\langle\Omega|_{ab} \right) |\Phi\rangle_{ab} |\psi\rangle \| \leq 2\Delta.$$

When K' acts on a state $|\psi\rangle_A |\psi\rangle_B |\Phi\rangle_{ab}$, the Schmidt rank is increased by at most the Schmidt rank of L, which is

$$\left(\frac{M}{N\Delta} \right)^{\mathcal{O}(1)} = \left(\frac{1}{\Delta} \sum_i \frac{w_i}{N} \right)^{\mathcal{O}(1)} = \left(\frac{\sum_i \alpha_i}{\Delta} \right)^{\mathcal{O}(1)}.$$

By definition, this is the Schmidt rank of the EPR-assisted AGSP $K' (\mathbb{1} \otimes |\Phi\rangle_{ab})$. This completes the proof. □

Chapter 4

Learning quantum interactions

Chapter summary: In this chapter, we study the problem of learning the Hamiltonian of a quantum many-body system given samples from its Gibbs (thermal) state. The classical analog of this problem, known as learning graphical models or Boltzmann machines, is a well-studied question in machine learning and statistics. We present the first sample-efficient algorithm for the quantum Hamiltonian learning problem. In particular, we prove that polynomially many samples in the number of particles (qudits) are necessary and sufficient for learning the parameters of a geometrically-local Hamiltonian in ℓ_2-norm.

Our main contribution is in establishing the strong convexity of the log-partition function of quantum many-body systems, which along with the maximum entropy estimation yields our sample-efficient algorithm. Classically, the strong convexity for partition functions follows from the Markov property of Gibbs distributions. This is, however, known to be violated in its exact form in the quantum case. We introduce several new ideas to obtain an unconditional result that avoids relying on the Markov property of quantum systems, at the cost of a slightly weaker bound. In particular, we prove a lower bound on the variance of quasi-local operators with respect to the Gibbs state, which might be of independent interest. Our work paves the way toward a more rigorous application of machine learning techniques to quantum many-body problems. This chapter is based on:

[AAKS21] Anurag Anshu, Srinivasan Arunachalam, Tomotaka Kuwahara, and Mehdi Soleimanifar. *Sample-efficient learning of interacting quantum systems. Nature Physics, 17(8):931–935*, 2021. Also in *Proceedings of IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS), pages 685–691,* 2020.

4.1 Introduction

The fundamental interactions in materials, given by the electromagnetic forces between electrons and nuclei, are often too complicated to be grasped in their entirety. This has shifted the attention to understanding the effective interactions between particles, described by local Hamiltonians, that if accurately chalked out, can be used to describe a variety of properties of the system and novel phases of matter. Recent advances in experimental techniques allow for the synthesis and study of increasingly complex interacting quantum systems [BSK+17, ZPH+17, AAB+19, SBM+11]. This prompts an important question: How can we infer the interactions between particles in such complicated systems or certify that their Hamiltonians indeed match the theoretically predicted models? A similar question arises in the context of near-term quantum computers. A major application

of quantum computers is to simulate the dynamics or particular states of a quantum system governed by a given Hamiltonian. In particular, the task of preparing the thermal state of a target Hamiltonian, known as quantum Gibbs sampling, is the backbone of many quantum algorithms such as semi-definite programming solvers [BS17b, vAGGdW20, BKF22], quantum simulated annealing [Mon15b, HW20], quantum machine learning [WKS16], and quantum simulations at finite temperature [MST+20]. Since near-term quantum computers will be noisy, an important problem in the scalable development of these devices is to verify their performance and calibrate them. For various quantum algorithms this means devising classical algorithms that, using measurement data, determine if the correct Hamiltonian has been implemented on the quantum device.

In this chapter, we focus on one version of this problem where we aim to learn the underlying Hamiltonian of a quantum system given multiple identical copies of its Gibbs state at a known fixed temperature. We refer to this task as the *quantum Hamiltonian learning problem*. More formally, we consider a geometrically-local Hamiltonian H acting on n qudits which are arranged on the vertices of a finite dimensional lattice. In general, we can parameterize H by

$$H(\mu) = \sum_{\ell=1}^{m} \mu_\ell E_\ell$$

where $\mu_\ell \in \mathbb{R}$ and the operators E_ℓ are Hermitian and $\{E_\ell\}$ forms an orthogonal basis for the space of operators. We say the Hamiltonian H is κ-local when the number of qubits in the support of each operator E_ℓ is at most κ. In geometrically-local Hamiltonians, the maximum distance between the qubits in the support of operators E_ℓ (measured with respect to the underlying lattice) is also bounded by a constant (see Section 4.6.2 and Definition 79 for more formal definitions). For instance in the case of qubits, operators E_ℓ are tensor products of at most κ Pauli operators that act non-trivially only on spatially-close qubits. We let the vector $\mu = (\mu_1, \ldots, \mu_m)^\top$ be the vector of interaction coefficients. In our setup, without loss of generality we assume the Hamiltonian is traceless, i.e., for the identity operator $E_\ell = \mathbb{1}$, the coefficient $\mu_\ell = 0$. At a inverse-temperature β, the qudits are in the Gibbs state defined as

$$\rho_\beta(\mu) = \frac{e^{-\beta H(\mu)}}{\mathrm{tr}[e^{-\beta H(\mu)}]}.$$

In the quantum Hamiltonian learning problem, we are given multiple copies of $\rho_\beta(\mu)$ and can perform arbitrary *local measurements* on them. This allows us to estimate all the κ-local *marginals* of $\rho_\beta(\mu)$ denoted by

$$e_\ell = \mathrm{tr}[\rho_\beta(\mu) E_\ell] \quad \textit{for } \ell \in [m].$$

The goal is to learn the coefficients μ_ℓ of the Hamiltonian H using the result of these measurements.

Learning the Hamiltonian of a quantum system has a natural classical analog, known as learning Boltzmann machines or, more generally, graphical models which is a central problem in machine learning and modern statistical inference. When expressed using physics terminology, Boltzmann machines correspond to Ising models, a prototypical classical spin Hamiltonian, and the learning problem is equivalent to inferring the corresponding Hamiltonian using samples from its Gibbs (thermal) distribution. Due to the wide-spread application of these models [WJ08], a large body of work has been devoted to studying them, resulting in highly efficient methods for learning Boltz-

mann machines that operate even in the regime where simulating the equivalent Ising model is computationally NP-hard [CL68, HS+86, Bre15, KM17, VMLC16].

Given the practical importance of quantum Hamiltonian learning and the remarkable achievements of machine learning methods for the classical version of this problem, various theoretical studies [BAL19, SK14, QR19], experimental implementations [WPS+17, SSR+14], and a great number of heuristic algorithms [BAL19, EHF19, BGP+20, SML+11, WGFC14b, WGFC14a] for this task have appeared. Despite these efforts, the primary challenge in the quantum Hamiltonian learning problem has remained long open, that is, to devise learning methods that are *provably efficient* in terms of the resources required for inferring the interactions with high accuracy. Most significantly, a prerequisite for any useful Hamiltonian learning method is an *efficient sample (or copy) complexity*. That is, the number of performed measurements, or equivalently, the number of identically prepared samples of the system, should ideally scale efficiently with the number of particles. The previous results, however, use a number of samples that in general can grow exponentially, lack a rigorous performance guarantee, assume additional control of the system, or only apply to special systems with few particles.

A fundamental obstacle to achieving efficient methods for quantum Hamiltonian learning is a striking feature of interacting quantum systems compared to their classical counterparts. At thermal equilibrium, the classical spin systems always satisfy what is known as the Markov property. That is, the correlations between two distant spins are mediated by the intermediate spins located in between them such that by conditioning on the state of these intermediate spins, the two distant spins become independent of each other (see Section 4.3.3 for more details). A crucial implication of the Markov property for classical many-body systems is that their Hamiltonian can be robustly learned using local measurements. Surprisingly, quantum spin systems are known to violate the Markov property in its exact form [LP08]. This makes it unclear whether, in principle, recovering quantum interactions from the results of local measurements is possible and complicates efforts to extend the machine learning techniques from classical to quantum Hamiltonians. Resolving this question and obtaining an algorithm for inferring the local Hamiltonian given efficient number of copies of the Gibbs state of the system is the main objective of this chapter.

4.2 Main result

Motivated by these applications, we now formally define the Hamiltonian learning problem. Our main result is a sample-efficient algorithm for the Hamiltonian learning problem.

Theorem 68 (Sample-efficient Hamiltonian learning)**.** *Consider a geometrically-local Hamiltonian* $H(\mu) = \sum_{\ell=1}^{m} \mu_\ell E_\ell$ *that acts on* n *qudits and consists of* m *local terms such that* $\max_{\ell\in[m]} |\mu_\ell| \leq 1$*. There is a Hamiltonian learning algorithm that uses*

$$N = \mathcal{O}\left(\frac{e^{\mathcal{O}(\beta^c)}}{\beta^{\tilde{c}}\varepsilon^2} \cdot m^3 \cdot \log\left(\frac{m}{\mathrm{err}}\right)\right) \tag{4.1}$$

copies of the Gibbs state $\rho_\beta(\mu) = e^{-\beta H(\mu)}/\mathrm{tr}[e^{-\beta H(\mu)}]$ *at a fixed inverse-temperature* β*, and obtains an estimate* $\hat{\mu} = (\hat{\mu}_1, \ldots, \hat{\mu}_m)$ *of the coefficients* μ_ℓ *such that with probability at least* $1 - \mathrm{err}$*,*

$$\|\mu - \hat{\mu}\|_2 \leq \varepsilon.$$

Here $c, \tilde{c} \geq 1$ *are constants depending on the geometry of the Hamiltonian and* $\|\mu - \hat{\mu}\|_2 = \left(\sum_{\ell=1}^{m} |\mu_\ell - \hat{\mu}_\ell|^2\right)^{\frac{1}{2}}$ *is the* ℓ_2*-norm of the difference of* μ *and* $\hat{\mu}$*.*

For goemtrically-local Hamiltonians the number of interaction terms m scales as $O(n)$. Hence, our result in Theorem 68 implies a sample complexity polynomial in the number of qudits. The number of samples in (4.1) increases as $\beta \to \infty$ or $\beta \to 0$. As the temperature increases ($\beta \to 0$), the Gibbs state approaches the maximally mixed state independent of the choice of parameters μ. At low temperatures ($\beta \to \infty$), the Gibbs state is in the vicinity of the ground space, which for instance, could be a product state $|0\rangle^{\otimes n}$ for the various choices of μ. In either cases, more sample are required to distinguish the parameters μ.

To complement our upper bound, we also obtain a $\Omega(\sqrt{m})$ lower bound for the Hamiltonian learning problem with ℓ_2 norm using a simple reduction to the state discrimination problem. The proof appears in Section 4.7. Hence, our upper bound in Theorem 68 is tight up to polynomial factors.

Theorem 69. *The number of copies* N *of the Gibbs state needed to solve the quantum Hamiltonian learning problem and output a* $\hat{\mu}$ *satisfying* $\|\hat{\mu} - \mu\|_2 \leq \varepsilon$ *with probability* $1 - \mathrm{err}$ *is lower bounded by*

$$N \geq \Omega\Big(\frac{\sqrt{m} + \log(1 - \mathrm{err})}{\beta\varepsilon}\Big).$$

4.3 Proof overview

In order to prove our main result, we introduce several new ideas. In this section, we provide a sketch of the main ingredients in our proof.

4.3.1 Maximum entropy estimation and sufficient statistics

In statistical learning theory, a conventional method for obtaining the parameters of a probability distribution from data relies on the concepts of *sufficient statistics* and the *maximum entropy estimation*. Suppose $p(x; \mu)$ is a family of probability distributions parameterized by μ that we want to learn. This family could for instance be various normal distributions with different mean or variance. Let $X_1, \ldots, X_m \sim p(x; \mu)$ be m samples from a distribution in this family. A sufficient *statistic* is a function T of these samples $T(X_1, \ldots, X_m)$ such that conditioned on that, the original date set $X_1, \ldots, X_m$ does not depend on the parameter μ. For example, the sample mean and variance are well known sufficient statistic functions.

After obtaining the sufficient statistic of a given data set given classical samples, there is a natural algorithm for estimating the parameter μ: among all the distributions that match the observed statistic $T(X)$ find the one that maximizes the Shannon entropy. Intuitively, this provides us with the least biased estimate given the current samples [Jay57a, Jay82]. This algorithm, which is closely related to the maximum likelihood estimation, is commonly used for analyzing the sample complexity of classical statistical problems.

Our first observation when addressing the Hamiltonian learning problem is that this method can be naturally extended to the quantum problem as previously considered in [Jay57b, SK14]. Indeed, the maximum entropy principle has already appeared in other quantum algorithms

such as [BKL+19]. More formally, in the following proposition, whose proof is given in Section 4.8.1, we first show that the marginals $\mathrm{tr}[E_\ell \rho]$ for $\ell \in [m]$ form a sufficient statistic for the Hamiltonian learning problem.

Proposition 70 (Matching local marginals implies global equivalence)**.** *Consider the following two Gibbs states*

$$\rho_\beta(\mu) = \frac{e^{-\beta \sum_\ell \mu_\ell E_\ell}}{\mathrm{tr}[e^{-\beta \sum_\ell \mu_\ell E_\ell}]}, \quad \rho_\beta(\lambda) = \frac{e^{-\beta \sum_\ell \lambda_\ell E_\ell}}{\mathrm{tr}[e^{-\beta \sum_\ell \lambda_\ell E_\ell}]} \tag{4.2}$$

such that $\mathrm{tr}[\rho_\beta(\lambda) E_\ell] = \mathrm{tr}[\rho_\beta(\mu) E_\ell]$ *for all* $\ell \in [m]$*, i.e. all the* κ*-local marginals of* $\rho_\beta(\lambda)$ *match that of* $\rho_\beta(\mu)$*. Then, we have* $\rho_\beta(\lambda) = \rho_\beta(\mu)$*, which in turns implies* $\lambda_\ell = \mu_\ell$ *for* $\ell \in [m]$*.*

Similar to the classical case discussed above, one implication of Proposition 70 is a method for learning the Hamiltonian H: first measure all the κ-local marginals of the Gibbs state e_ℓ, then among all the states of the form (4.2), find the one that matches those marginals. Finding such a state can be naturally formulated in terms of an optimization problem known as the *maximum entropy problem*:

$$\begin{aligned} \max_{\sigma} \quad & S(\sigma) \\ \text{s.t.} \quad & \mathrm{tr}[\sigma E_\ell] = e_\ell, \quad \forall \ell \in [m] \\ & \sigma > 0, \quad \mathrm{tr}[\sigma] = 1. \end{aligned} \tag{4.3}$$

where $S(\sigma) = -\mathrm{tr}[\sigma \log \sigma]$ is the *von Neumann entropy* of the state σ. The optimal solution of this program is a quantum state with a familiar structure [Jay57b]. Namely, it is a Gibbs state $\rho(\lambda)$ for some set of coefficients $\lambda = (\lambda_1, \ldots, \lambda_m)$. The coefficients λ are the *Lagrange multipliers* corresponding to the dual of this program. Indeed, we can write the dual program of Eq. (4.3) as follows:

$$\mu = \operatorname*{arg\,min}_{\lambda = (\lambda_1, \ldots, \lambda_m)} \log Z_\beta(\lambda) + \beta \cdot \sum_{\ell=1}^{m} \lambda_\ell e_\ell, \tag{4.4}$$

where $Z_\beta(\lambda) = \mathrm{tr}\big(e^{-\beta \cdot \sum_\ell \lambda_\ell E_\ell}\big)$ is the *partition function* at inverse-temperature β. In principle, according to the result of Proposition 70, we could solve the Hamiltonian learning problem by finding the optimal solution of the dual program in (4.4). Of course, the issue with this approach is that since we have access to limited number of samples of the original Gibbs state $\rho_\beta(\mu)$, instead of the exact marginals e_ℓ, we can only *approximately* estimate the e_ℓs. We denote these estimates by $\hat{e}_\ell$. This means instead of solving the dual program (4.4), we solve its *empirical* version

$$\hat{\mu} = \operatorname*{arg\,min}_{\lambda = (\lambda_1, \ldots, \lambda_m)} \log Z_\beta(\lambda) + \beta \cdot \sum_{\ell=1}^{m} \lambda_\ell \hat{e}_\ell. \tag{4.5}$$

The main technical problem that we address in the upcoming sections is analyzing the robustness of the programs (4.3) and (4.4) to the statistical error in the marginals as appears in (4.5). This is an instance of a *stochastic optimization* which is a well-studied problem in optimization. In the next section, we review the ingredients from convex optimization that we need in our analysis.

4.3.2 Strong convexity

One approach to incorporate the effect of the statistical errors in the marginals e_ℓ into the estimates for μ_ℓ is to use Proposition 70. It is not hard to extend this proposition to show that if a Gibbs states $\rho_\beta(\lambda)$ *approximately* matches the marginals of $\rho_\beta(\mu)$ up to some error ε, then $\|\rho_\beta(\mu) - \rho_\beta(\lambda)\|_1^2 \leq \mathcal{O}(m\varepsilon)$ (see Remark 91 for more details). This bound, however, is not strong enough for our purposes. This is because if we try to turn this bound to a one on the coefficients μ_ℓ of the Hamiltonian, we need to bound $\|\log \rho_\beta(\mu) - \log \rho_\beta(\lambda)\|$. Unfortunately, the function $\log(x)$ does not have a bounded gradient (i.e., it is not Lipschitz) over its domain and in general $\|\log \rho_\beta(\mu) - \log \rho_\beta(\lambda)\|$ can be exponentially worse than $\|\rho_\beta(\mu) - \rho_\beta(\lambda)\|_1$. In order to overcome the non-Lipschitz nature of the logarithmic function and bound $\|\log \rho_\beta(\mu) - \log \rho_\beta(\lambda)\|$, we prove a property of the dual objective function (4.4) known as the *strong convexity*, which we define now.

Definition 71. *Consider a convex function $f : \mathbb{R}^m \mapsto \mathbb{R}$ with gradient $\nabla f(x)$ and Hessian $\nabla^2 f(x)$ at a point x.*[1] *This function f is said to be α-strongly convex in its domain if it is differentiable and for all x, y,*

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) + \frac{1}{2}\alpha \|y - x\|_2^2,$$

or equivalently if the minimum eigenvalue of the Hessian $\sigma_{\min}(\nabla^2 f(x))$ satisfies

$$\sigma_{\min}(\nabla^2 f(x)) \geq \alpha. \tag{4.6}$$

In other words, for any vector $v \in \mathbb{R}^m$, it holds that $\sum_{i,j} v_i v_j \frac{\partial^2}{\partial x_i \partial x_j} f(x) \geq \alpha \|v\|_2^2$.

Roughly speaking, strong convexity puts a limit on how *slow* a convex function $f(x)$ changes.[2] This is particularly useful because given two points x, y and an upper bound on $|f(y) - f(x)|$ and $\nabla f(x)^\top (y - x)$, it allows us to infer an upper bound on $\|y - x\|_2$. For our application, we think of f as being $\log Z_\beta(\cdot)$. Then the difference $|f(y) - f(x)|$ is the difference between the optimal solution of the original program in Eq. (4.4) and that of its empirical version in Eq. (4.5) which includes the statistical error. We apply this framework to our optimization (4.5) in two steps:

1) Proving the strong convexity of the objective function: This is equivalent to showing that the log-partition function (aka the free energy) is strongly convex, i.e., $\sigma_{\min}(\nabla^2 \log Z_\beta(\lambda)) \geq \alpha$ for some positive coefficient α. In particular, this means that the optimization (4.5) is a convex program. This result is the main technical contribution of this Chapter and is stated in the following theorem whose proof is overviewed in Section 4.3.4 and given in detail in Section 4.9.

   Theorem 72 (Informal: strong convexity of log-partition function)**.** *Let $H = \sum_{\ell=1}^m \mu_\ell E_\ell$ be a κ-local Hamiltonian over a finite dimensional lattice with $\|\mu\| \leq 1$. For a given inverse-temperature β, there are constants $c, c' > 3$ depending on the geometric properties of the lattice such that*

$$\sigma_{\min}\left(\nabla^2 \log Z_\beta(\mu)\right) \geq e^{-\mathcal{O}(\beta^c)} \frac{\beta^{c'}}{m}, \tag{4.7}$$

[1] Recall that the entries of the Hessian matrix $\nabla^2 f(x)$ are given by $\frac{\partial^2}{\partial x_i \partial x_j} f(x)$

[2] This should not be confused with a related property called the smoothness which limits how fast the function grows.

i.e., for every vector $v \in \mathbb{R}^m$ *we have* $v^T \cdot \nabla^2 \log Z_\beta(\mu) \cdot v \geq e^{-\mathcal{O}(\beta^c)} \frac{\beta^{c'}}{m} \cdot \|v\|_2^2$.

2) Bounding the error in estimating μ in terms of the error in estimating the marginals e_ℓ: In this step we show that as long as the statistical error of the marginals is small, using the strong convexity property from step (1), we can still prove an upper bound on the difference between the solutions of the convex programs (4.4), (4.5). We discuss this in more details later in Section 4.8.2. The result can be stated as follows:

Theorem 73 (Error bound from strong convexity)**.** *Let* $\delta, \alpha > 0$. *Suppose the marginals* e_ℓ *are determined up to error* err*, i.e.,* $|e_\ell - \hat{e}_\ell| \leq \delta$ *for all* $\ell \in [m]$. *Additionally assume* $\sigma_{\min}(\nabla^2 \log Z_\beta(\lambda)) \geq \alpha$ *and* $\|\lambda\| \leq 1$. *Then the optimal solution to the program* (4.5) *satisfies*

$$\|\mu - \hat{\mu}\|_2 \leq \frac{2\beta\sqrt{m}\delta}{\alpha}.$$

Combining Theorem 72 and Theorem 73, we obtain our claimed sample complexity result. We now proceed to sketch the proof of Theorem 72.

4.3.3 Strong convexity of classical log-partition functions

In order to better understand the motivation behind our quantum proof, it is insightful to start with the *classical* Hamiltonian learning problem. This helps us better describe various subtleties (briefly mentioned in the main Article) and what goes wrong when trying to adapt the classical techniques to the quantum case. We continue using the quantum notation here, but the reader can replace the Hamiltonian H, for instance, with the classical Ising model $H = \sum_{i \sim j} J_{ij} x_i x_j$ (where $x_i \in \{-1, 1\}$ and $J_{ij} \in \mathbb{R}$).

The entries of the Hessian $\nabla^2 \log Z_\beta(\mu)$ for classical Hamiltonians are given by

$$\frac{\partial^2}{\partial \mu_i \partial \mu_j} \Big[\log Z_\beta(\mu) \Big] = \mathrm{Cov}[E_i, E_j] \tag{4.8}$$

where Cov is the covariance function which is defined as $\mathrm{Cov}[E_i, E_j] = \langle E_i E_j \rangle - \langle E_i \rangle \langle E_j \rangle$ with the expectation taken with respect to the Gibbs distribution (i.e., $\langle E \rangle = \mathrm{tr}[E \cdot \rho_\beta(\mu)]$). To prove the strong convexity of the log-partition function at a constant β (defined in Definition 92), using (4.8) it suffices to show that for every vector v, we have

$$\sum_{i,j} v_i v_j \frac{\partial^2}{\partial \mu_i \partial \mu_j} \log Z_\beta(\mu) = \mathrm{Var}\left[\sum_{\ell=1}^m v_\ell E_\ell \right] \geq \Theta(1) \cdot \sum_{\ell=1}^m v_\ell^2. \tag{4.9}$$

Although the operator $\sum_\ell v_\ell E_\ell$ is a local Hamiltonian, note the mismatch between this operator and the original Hamiltonian in the Gibbs state $\sum_{\ell=1}^m \mu_\ell E_\ell$. Note that the inequality (4.9) is stronger than our main technical contribution in the quantum setting (i.e., for the case of quantum partition functions, we proved the analogue of inequality (4.9) when $\Theta(1)$ replaced by $\Theta(1/m)$). Before proving Eq. (4.9), we remark that an *upper bound* of $\mathrm{Var}[\sum_{\ell=1}^m v_\ell E_\ell] \leq \mathcal{O}(1)\|v\|_2^2$ is known in literature, under various conditions like the decay of correlations both in classical and quantum settings [Ara69, Gro79, PY95, Uel04, KGK+14, FU15]. This upper bound intuitively makes sense because the variance of the thermal state of a Hamiltonian and other local observables are expected

to be *extensive*, i.e., they scale with the number of particles (spins) or norm of the Hamiltonian, which is replaced by $\|v\|_2^2$ in our setup. However, in the classical Hamiltonian learning problem, we are interested in obtaining a *lower bound* on the variance. To this end, a crucial property of the (classical) Gibbs distributions that allows us to prove the inequality (4.9) is the conditional independence or the *Markov property* of classical systems.

Definition 74 (Markov property)**.** *Suppose the interaction graph is decomposed into three disjoint regions A, B, and C such that region B "shields" A from C, i.e., the vertices in region A are not connected to those in C. Then, conditioned on the sites in region B, the distribution of sites in A is independent of those in C. This is often conveniently expressed in terms of the conditional mutual information by $I(A:C|B) = 0$.*

It is known by the virtue of the Hammersley-Clifford theorem [HC71] that the family of distributions with the Markov property coincides with the Gibbs distributions. Using this property, we can lower bound $\mathrm{Var}\left[\sum_{\ell=1}^m v_\ell E_\ell\right]$ in terms of variance of local terms E_ℓ by *conditioning* on a subset of sites. To this end, we consider a partition of the interaction graph into two sets A and B. The set B is chosen, suggestively, such that the vertices in A are not connected (via any edges) to each other. We denote the spin configuration of sites in B collectively by s_B. Then using the concavity of the variance and the Markov property, we have

$$\begin{aligned}
\mathrm{Var}\left[\sum_{\ell=1}^m v_\ell E_\ell\right] &\overset{(1)}{\geq} \mathbb{E}_{s_B}\left[\mathrm{Var}\left[\sum_{\ell=1}^m v_\ell E_\ell \;\middle|\; s_B\right]\right] \\
&\overset{(2)}{=} \sum_{x\in A} \mathbb{E}_{s_B}\left[\mathrm{Var}\left[\sum_{\ell:E_\ell \text{ acts on } x} v_\ell E_\ell \;\middle|\; s_B\right]\right] \\
&\overset{(3)}{\geq} \Theta(1)\cdot\sum_{\ell=1}^m v_\ell^2,
\end{aligned} \tag{4.10}$$

where inequality (1) follows from the law of total variance, equality (2) can be justified as follows: by construction, the local terms E_ℓ either completely lie inside region B or intersect with only one of the sites in region A. In the former, the local conditional variance $\mathrm{Var}\left[E_\ell\,|s_B\right]$ vanishes, whereas in the latter, the interaction terms E_ℓ that act on different sites $x \in A$ become uncorrelated and the global variance decomposes into a sum of local variance. Finally, inequality (3) is derived by noticing that at any constant inverse-temperature β, the local variance is lower bounded by a constant that scales as $e^{-\Theta(\beta)}$. By carefully choosing the partitions A and B such that $|A| = \mathcal{O}(n)$, we can make sure that the variance in inequality (2) is a constant fraction of the $\sum_{\ell=1}^m v_\ell^2$ as in (4.10) (see [Mon15a, VMLC16] for details).

4.3.4 Strong convexity of quantum log-partition functions

If we try to directly quantize the proof strategy of the classical case in the previous section, we immediately face several issues. In what follows we describe the challenges in obtaining a quantum proof along with our techniques to overcome them which allow us to establish Theorem 98. The content of our proof of the quantum case is divided into four steps whose overview is given below.

Overview of step 1: Relating the Hessian to a variance

The first problem is that we cannot simply express the entries of the Hessian matrix $\nabla^2 \log Z_\beta(\mu)$ in terms of $\mathrm{Cov}[E_i, E_j]$ as in (4.8). This expression in (4.8) only holds for Hamiltonians with *commuting* terms, i.e., $[E_i, E_j] = 0$ for all $i, j \in [m]$. The Hessian for the non-commuting Hamiltonians takes a complicated form (see Lemma 99 for the full expression) that makes its analysis difficult. Our first contribution, covered in Section 4.9.1, is to recover a similar result to (4.9) in the quantum case by showing that, for every v, we can still *lower bound* $v^\top \cdot \nabla^2 \log Z_\beta(\mu) \cdot v$ by the variance of a suitably defined *quasi-local* operator. We later define what we mean by "quasi-local" more formally (see Definition 80), but for now one can assume such an operator is, up to some small error, sum of local terms.

Lemma 75 (A lower bound on $v^\top \cdot \nabla^2 \log Z_\beta \cdot v$)**.** *For any vector $v \in \mathbb{R}^m$, define a quasi-local operator $\widetilde{W} = \sum_{\ell=1}^m v_\ell \widetilde{E}_\ell$, where the operators $\tilde{E}_\ell$ are defined by*

$$\widetilde{E}_\ell = \int_{-\infty}^{\infty} f_\beta(t)\ e^{-iHt}\ E_\ell\ e^{iHt} dt. \tag{4.11}$$

Here $f_\beta(t) = \frac{2}{\beta\pi} \log \frac{e^{\pi|t|/\beta}+1}{e^{\pi|t|/\beta}-1}$ is defined such that $f_\beta(t)$ scales as $\frac{1}{\beta} e^{-\pi|t|/\beta}$ for large t and $f_\beta(t) \propto \log(1/t)$ for $t \to +0$. Then

$$\sum_{i,j} v_i v_j \frac{\partial^2}{\partial \mu_i \partial \mu_j} \log Z_\beta(\mu) \geq \beta^2 \operatorname{Var}[\widetilde{W}]. \tag{4.12}$$

This implies the bound $\sigma_{\min}(\log Z_\beta(\mu)) \geq \beta^2 \operatorname{Var}[\widetilde{W}]$ on the the minimum eigenvalue of $\log Z_\beta(\mu)$.

Overview of step 2: From global to local variance

As a result of Lemma 75, we see that from here onwards, it suffices to lower bound the variance of the quasi-local operator $\widetilde{W} = \sum_{\ell=1}^m v_\ell \tilde{E}_\ell$. One may expect the same strategy based on the Markov property in (4.10) yields the desired lower bound. Unfortunately, it is known that a natural extension of this property to the quantum case, expressed in terms of the *quantum conditional mutual information* (qCMI), does not hold. In particular, example Hamiltonians are constructed in [LP08] such that for a tri-partition A, B, C as in Definition 74, their Gibbs states have non-zero qCMI, i.e., $I(A:C|B) > 0$. Nevertheless, it is *conjectured* that an *approximate* version of this property can be recovered i.e., $I(A:C|B) \leq e^{-\Omega(\mathrm{dist}(A,C))}$. In other words, the approximate property claims that qCMI is exponentially small in the *width* of the shielding region B. Thus far, this conjecture has been proved only at sufficiently high temperatures [KKBa20] and on 1D chains [KBa19b]. Even assuming this conjecture is true, we currently do not know how to recover the argument in (4.10). Given this issue we ask,

> *Can we obtain an unconditional lower bound on the variance of a quasi-local observable at any inverse-temperature β without assuming quantum conditional independence?*

Our next contribution is to give an affirmative answer to this question. To achieve this, we modify the classical strategy as explained below.

One ingredient in the classical proof is to lower bound the global variance $\operatorname{Var}[\sum_\ell v_\ell E_\ell]$ by sum of local conditional variances $\operatorname{Var}[E_\ell | s_B]$ as in (4.10). We prove a similar but slightly weaker result

in the quantum regime. To simplify our discussion, let us ignore the fact that $\widetilde{W} = \sum_\ell v_\ell \tilde{E}_\ell$ is a quasi-local operator and view it as (strictly) local. Consider a special case in which v is such that the operator $\widetilde{W}$ is supported on a small number of sites. For instance, it could be that $v_1 > 0$ while $v_2, \ldots, v_m = 0$. Then the variance $\mathrm{Var}[\widetilde{W}]$ can be easily related to the local variance $\mathrm{Var}[E_1]$ and since $E_1^2 = \mathbb{1}$, $|\mathrm{tr}[E_1\rho_\beta]| < 1$, we get

$$\mathrm{Var}[\widetilde{W}] = v_1^2 \cdot \left(\mathrm{tr}[E_1^2\rho_\beta] - \mathrm{tr}[E_1\rho_\beta]^2\right) \geq \Theta(1) \cdot v_1^2$$

In Section 4.9.3, we show that even in the general case, where $v_1, \ldots, v_m$ are all non-zero, we can still relate $\mathrm{Var}[\widetilde{W}]$ to the variance of a local operator supported on a constant region. Compared to the classical case in (4.10), where the lower bound on $\mathrm{Var}[W]$ includes a sum of $\mathcal{O}(m)$ local terms, our reduction to a *single* local variance costs "an extra factor of m" in our strong convexity bound.

Our reduction to local variance is based on the following observation. By applying Haar-random local unitaries on a site i, we can remove all the terms of the operator $\widetilde{W}$ except those that act on an arbitrary qudit at the site (see also Section 4.6.5). We denote the remainder terms by $\widetilde{W}_{(i)}$ defined via

$$\widetilde{W}_{(i)} = \widetilde{W} - \mathbb{E}_{U_i \sim \mathrm{Haar}}[U_i^\dagger \widetilde{W} U_i].$$

By using the triangle inequality this relation implies

$$\mathrm{Var}[\widetilde{W}] \geq \frac{1}{2}\mathrm{tr}[\widetilde{W}_{(i)}^2\rho_\beta] - \mathbb{E}_{U_i}\left[\mathrm{tr}[\widetilde{W}^2 \cdot U_i\rho_\beta U_i^\dagger]\right]. \tag{4.13}$$

Hence, if we could carefully analyze the effect of the term $\mathbb{E}_{U_i}[\mathrm{tr}[\widetilde{W}^2 \cdot U_i\rho_\beta U_i^\dagger]]$, this will allow us to relate the global variance $\mathrm{Var}[\widetilde{W}]$ to the local variance $\mathrm{tr}[\widetilde{W}_{(i)}^2\rho_\beta]$. We discuss this next.

Overview of step 3: Bounding the effect of local unitaries

While applying the above reduction helps us to go to an easier local problem, we need to deal with the changes in the spectrum of the Gibbs state due to applying the random local unitaries U_i. Could it be that the unitaries U_i severely change the spectral relation between $\widetilde{W}$ and ρ_β? We show that this is not the case, relying on the facts: (1) local unitaries cannot mix up subspaces of $\widetilde{W}$ and H that are energetically far away and (2) the weight given by the Gibbs state ρ_β to nearby subspaces of H are very similar at small β. Thus, (1) allows us to focus on the subspaces that are close in energy and (2) shows that similar weights of these subspaces do not change the variance by much. In summary, in Section 4.9.4 we prove:

Proposition 76 (Invariance under local unitaries, informal)**.** *Let U_X be a local unitary operator acting on region X that has a constant size. There exists a constant $c \leq 1$ such that*

$$\mathrm{tr}\left[\widetilde{W}^2 \cdot U_X\rho_\beta U_X^\dagger\right] \lesssim \left(\mathrm{Var}[\widetilde{W}]\right)^c. \tag{4.14}$$

When combined with (4.13), the inequality (4.14) implies the following loosely stated local lower bound on the global variance:

$$\mathrm{Var}[\widetilde{W}] \gtrsim \left(\mathrm{tr}\left[\widetilde{W}_{(i)}^2\rho_\beta\right]\right)^{\frac{1}{c}}.$$

Overview of step 4: Reduction to infinite temperature variance

With this reduction, it remains to find a constant lower bound on $\mathrm{tr}[\widetilde{W}^2_{(i)}\rho_\beta]$. This can be done, again, by applying a local unitary U. Roughly speaking, we use this unitary to perform a "change of basis" that relates the local variance at finite temperature to its infinite-temperature version. The spectrum of ρ_β majorizes the maximally mixed state η. Hence, by applying a local unitary, we can rearrange the eigenvalues of $\widetilde{W}^2_{(i)}$ in the same order as that of ρ_β such that when applied to both ρ_β and η, we have $\mathrm{tr}[\widetilde{W}^2_{(i)}U\rho_\beta U^\dagger] \geq \mathrm{tr}[\widetilde{W}^2_{(i)}\eta]$. Formally, in Section 4.9.5, we show that:

Proposition 77 (Lower bound on the local variance, informal)**.** *There exists a unitary U supported on $\mathcal{O}(1)$ sites such that*

$$\mathrm{tr}\left[\widetilde{W}^2_{(i)}U\rho_\beta U^\dagger\right] \geq \mathrm{tr}\left[\widetilde{W}^2_{(i)}\eta\right],$$

where η is the maximally mixed state or the infinite temperature Gibbs state.

In summary, starting from (4.13) and following Proposition 76 and Proposition 77, the lower bound on the global variance takes the following local form:

$$\mathrm{tr}\left[\widetilde{W}^2\rho_\beta\right] \geq \left(\mathrm{tr}\left[\widetilde{W}^2_{(i)}\eta\right]\right)^{\Theta(1)}.$$

Lower bounding the quantity $\mathrm{tr}\left[\widetilde{W}^2_{(i)}\eta\right]$ by a constant is now an easier task, which we explain in more detail later in Lemma 111 and Theorem 102.

4.4 Further discussions

4.4.1 Connection to previous works

There have been a number of proposals for the Hamiltonian learning problem in the past. In [BAL19, EHF19, QR19] learning the Hamiltonian from local measurements is considered. Their approach is based on setting up a linear system of equations whose constraints (i.e., the matrix of coefficients) are determined from the measurement outcomes. The solution of these equations is the parameter μ_k of the Hamiltonian. The sample complexity in this approach depends inverse polynomially on the "spectral gap" of the matrix of coefficients which thus far has not been rigorously bounded. Another line of work considers learning the Hamiltonian using a trusted quantum simulator [WGFC14b, WGFC14a, VMN+19] which is analyzed using a combination of numerical evidence and heuristic arguments. Amin et al. [AAR+18] quantized classical Boltzmann machines and proposed a method to train and learn quantum Boltzmann machines using gradient descent.

As mentioned earlier, there has been a fruitful series of works on the classical analog of the Hamiltonian learning problem (see e.g. [Bre15, KM17, VMLC16]). In this chapter, we assume it is a priori known that the interaction graph of the Hamiltonian is geometrically local. We then estimate the parameters in ℓ_2-norm using $\mathrm{poly}(n)$ samples which is polynomially tight even for classical Hamiltonians. If we instead consider estimation in ℓ_∞-norm, the classical algorithms can achieve a stronger result. That is, given $\mathcal{O}(\log n)$ samples, they succeed in efficiently learning the structure of the underlying graph and its parameters in ℓ_∞-norm. If we apply our current analysis to this setup, we cannot improve our $\mathrm{poly}(n)$ sample complexity to $\mathcal{O}(\log n)$. This is in part because the classical results devise a more efficient convex program that learns the parameters node-wise

(this relies on the commutativity of the Hamiltonian terms), and partly because their required strong convexity assumptions is based on the Markov property, none of which are known to be quantizable.

4.4.2 Open questions

In Section 4.3.1 we explained our approach to analyzing the Hamiltonian learning problem based on reducing data to its sufficient statistics and using maximum entropy estimation. An issue with this approach is the blowup in the computationally complexity. It is shown in [Mon15a] that this approach basically requires approximating the partition function which is NP-hard. Ideally, one would like to have an algorithm for the Hamiltonian learning problem that requires small number of samples, but also has an efficient running time. Satisfying both these constraints for all inverse-temperatures β even in the classical learning problems is quite challenge. It was only recently that more efficient algorithms are devised for learning graphical models [KM17, VMLC16]. Here, we focus on the less demanding but still non-trivial question of bounding the sample complexity and leave obtaining an efficient running time for future work. Below we mention some of the open problems in this direction.

Our lower bound on the variance in Theorem 72 is obtained for any constant inverse-temperature β. It is an interesting open question to improve this lower bound on the variance bound (4.7), ideally to a constant independent of system size, assuming physically-motivated conditions such as the decay of correlations or the decay of conditional mutual information. Another approach might be to derive such a bound at high temperatures where powerful tools such as cluster expansions are available [KKBa20] (see [HKT21] for progress on Hamiltonian learning in the high-temperature regime). We also expect our bounds can be improved for commuting Hamiltonians. Indeed, using structural results such as [BV03, AE11], one should be able to follow the same strategy as in Section 4.3.3 to find a constant lower bound on the variance of commuting Hamiltonians.

There are recent results on efficiently computing the partition function of quantum many-body systems under various assumptions [BG17, HMS20, KKBa20]. We expect by combining these results with our maximum entropy estimation algorithm in Section 4.3.1, one can obtain efficient classical algorithms for the Hamiltonian learning problem. Another approach might be to use calibrated quantum computers (or Gibbs samplers) as in [BK16, BKL$^+$19] to solve the maximum entropy optimization using multiplicative weight update method and learn the parameters of another quantum device.

Finally, an important future direction is to devise more refined objective functions for the Hamiltonian learning problem that matches the performance of the learning algorithms for the classical problem as discussed in Section 4.4.1. Given the non-commutative nature of quantum Hamiltonians, this seems to require substantially new ideas and advances in characterizing the information theoretic properties of the quantum Gibbs states.

4.5 A guide to the remainder of this chapter

In the rest of this chapter, we provide the detailed proofs and analysis of the results presented so far. To aid the readers, we first begin with giving an overview of the subject of each future section.

- **Preliminaries (Section 4.6)**: We gather the notations, definitions, facts, and well-known technical tools that are frequently used in the subsequent sections. This includes

- Quantum belief propagation (Section 4.6.3)
- Change in the spectrum after applying local operators (Section 4.6.4)
- Local reduction of global operators (Section 4.6.5)

- **A lower bound on the sample complexity (Section 4.7)**: The proof of the $\sqrt{m}$ lower bound on the sample complexity of learning quantum Hamiltonians is given.

- **Efficient sample complexity assuming strong convexity (Section 4.8)**: We show how assuming the strong convexity of the log-partition function of quantum systems (whose proof is given in Section 4.9) implies an efficient sample complexity for the quantum Hamiltonian learning. This is done in two parts:
  - Sufficient statistics (Section 4.8.1)
  - Stochastic convex optimization (Section 4.8.2)

- **Proof of strong convexity of log-partition function (Section 4.9)**: Our main technical contribution is presented in depth here. The proof the strong convexity is divided into four steps. We have already seen a brief overview of these steps in Section 4.3.4, where an intuitive and broad perspective of them is provided. A diagram in Figure 4-1 depicts the theorems and lemmas used in each of these four steps and the connection between them.
  - Step 1: Relating the Hessian to a variance (Section 4.9.1)
  - Step 2: From global to local variance (Section 4.9.3)
  - Step 3: Bounding the effect of local unitaries (Section 4.9.4)
  - Step 4: Reduction to infinite temperature variance (Section 4.9.5)
  - Final step: Putting things together (Section 4.9.6)

4.6 Preliminaries

4.6.1 Some mathematical facts

Here we summarize some of the basic mathematical facts used in the proof. Let A, B be arbitrary operator. The *operator norm* of A which is its largest singular value is denoted by $\|A\|$. The minimum singular value of A (or equivalently the minimum eigenvalue when A is Hermitian) is denoted by $\sigma_{\min}(A)$. We also often use the *Frobenius norm* $\|A\|_F := \sqrt{\mathrm{tr}[A^\dagger A]}$ and more generally the Hilbert-Schmidt inner product between A, B defined by $\mathrm{tr}[A^\dagger B]$. Additionally using Hölder's inequality we have,

$$\|AB\|_F = \sqrt{\mathrm{tr}(B^\dagger A A^\dagger B)} \leq \sqrt{\|B\|^2 \mathrm{tr}(AA^\dagger)} = \|B\| \cdot \|A\|_F. \tag{4.15}$$

We define the von Neumann entropy of a quantum state ρ by $S(\rho) = -\mathrm{tr}[\rho \log \rho]$ and the relative entropy between two states ρ_1 and ρ_2 by $S(\rho_1 \| \rho_2) = -\mathrm{tr}[\rho_1 \log \rho_2] - S(\rho_1)$.

The gradient of a real function $f : \mathbb{R}^m \mapsto \mathbb{R}$ is denoted by $\nabla f(x)$ and its *Hessian* (second derivative) matrix by $\nabla^2 f(x)$. The entries of the Hessian matrix are given by $\frac{\partial^2}{\partial x_i \partial x_j} f(x)$. The minimum eigenvalue of the Hessian of $f(x)$ is shown by $\sigma_{\min}(\nabla^2 f(x))$.

We write $A \succeq 0$ to represent a *positive semi-definite* (PSD) operator A, one such example of a PSD operator is the Hessian matrix $\nabla^2 f(x)$ of a convex function $f(x)$.

We will use the standard notations $\mathcal{O}(\cdot), \Theta(\cdot)$, and $\Omega(\cdot)$ to specify the scaling of various parameters. In particular, $\Theta(1)$ will denote a constant.

For convenience, we will also gather a collection of infinite sums over exponentials. For $t > 0$, let

$$\Gamma(t) := \int_0^\infty x^{t-1} e^{-x} dx = \frac{1}{t} \int_0^\infty e^{-x} d\left(x^t\right) = \frac{1}{t} \int_0^\infty e^{-y^{\frac{1}{t}}} dy$$

be the *gamma function*. It holds that $\Gamma(t) \leq t^t$. This can be used to simplify several summations that we encounter later. Finally, we collect a few useful summations that we use in our proofs in the following fact.

Fact 78. *Let $a, c > 0$, $1 \geq p > 0$ be reals and b be a positive integer. Then*

1) $\sum_{j=0}^\infty e^{-cj} \leq \frac{e^c}{c}$.

2) $\sum_{j=0}^\infty j^b e^{-cj^p} \leq \frac{2}{p} \cdot \left(\frac{b+1}{cp}\right)^{\frac{b+1}{p}}$.

3) $\sum_{j=0}^\infty e^{-c(a+j)^p} \leq e^{-\frac{c}{2}a^p} \left(1 + \frac{1}{p}\left(\frac{2}{cp}\right)^{\frac{1}{p}}\right)$.

Proof. The first summation follows from

$$\sum_{j=0}^\infty e^{-cj} = \frac{1}{1-e^{-c}} = \frac{e^c}{e^c - 1} \leq \frac{e^c}{c}.$$

For the second sum, notice that the function $t^b e^{-ct^p}$ achieves the maximum at $t^* = \left(\frac{b}{cp}\right)^{\frac{1}{p}}$. Then

$$\begin{aligned}
\sum_{j=0}^\infty j^b e^{-cj^p} &\leq t^* \left(t^*\right)^b e^{-c(t^*)^p} + \int_0^\infty t^b e^{-ct^p} dt \\
&= \left(\frac{b}{cp}\right)^{\frac{b+1}{p}} e^{-\frac{b}{p}} + \frac{1}{(b+1)c^{\frac{b+1}{p}}} \int_0^\infty e^{-y^{\frac{p}{b+1}}} dy \\
&= \left(\frac{b}{e^{\frac{b}{b+1}} cp}\right)^{\frac{b+1}{p}} + \frac{1}{pc^{\frac{b+1}{p}}} \Gamma\left(\frac{b+1}{p}\right) \\
&\leq \left(\frac{b}{e^{\frac{b}{b+1}} cp}\right)^{\frac{b+1}{p}} + \frac{1}{pc^{\frac{b+1}{p}}} \left(\frac{b+1}{p}\right)^{\frac{b+1}{p}} \leq \frac{2}{p} \cdot \left(\frac{b+1}{cp}\right)^{\frac{b+1}{p}}.
\end{aligned}$$

For the third sum, we will use the identity

$$(a+j)^p \geq \frac{1}{2}\left(a^p + j^p\right).$$

Now, consider the following chain of inequalities and change of variables:

$$\begin{aligned}
\sum_{j=0}^{\infty} e^{-c(a+j)^p} &\leq e^{-\frac{c}{2}a^p} \sum_{\ell=0}^{\infty} e^{-\frac{c}{2}\ell^p} \\
&\leq e^{-\frac{c}{2}a^p} \left(1 + \int_0^{\infty} e^{-\frac{c}{2}t^p} dt\right) \\
&= e^{-\frac{c}{2}a^p} \left(1 + \frac{2^{\frac{1}{p}}}{c^{\frac{1}{p}}} \int_0^{\infty} e^{-y^p} dy\right) \\
&= e^{-\frac{c}{2}a^p} \left(1 + \frac{2^{\frac{1}{p}}}{pc^{\frac{1}{p}}} \Gamma\left(\frac{1}{p}\right)\right) \leq e^{-\frac{c}{2}a^p} \left(1 + \frac{1}{pc^{\frac{1}{p}}} \left(\frac{2}{p}\right)^{\frac{1}{p}}\right).
\end{aligned}$$

This completes the proof. □

4.6.2 Local Hamiltonians and quantum Gibbs states

Local Hamiltonians. As mentioned in Section 4.1, in this chapter, we consider Hamiltonians that are *geometrically local*. To describe this notion more precisely, we consider a D-dimensional lattice $\Lambda \subset \mathbb{Z}^D$ that contains n sites with a d-dimensional qudit (spin) on each site. We denote the dimension of the Hilbert space associated to the lattice Λ by $\mathcal{D}_\Lambda$.

Let $B(r, i) := \{j \in \Lambda | \text{dist}(i, j) \leq r\}$ denote a ball (under the Manhattan distance on the lattice) of size r centered at site i. For a given connected set $X \in \Lambda$, let $\text{diam}(X) := \max\{\text{dist}(i, j) : i, j \in X\}$ denote the *diameter* of this set, $X^c := \Lambda \setminus X$ denote the complement of this set, and ∂X denote its boundary. Given two sets $X, Y \in \Lambda$, we define $\text{dist}(X, Y) := \min\{\text{dist}(i, j) : i \in X, j \in Y\}$.

The Hamiltonian of this system is

$$H = \sum_{X \subset \Lambda} H_X.$$

Each term H_X acts only on the sites in X and X is restricted to be a connected set with respect to Λ. We call the Hamiltonian κ-local when the support of all the local terms H_X is $|X| \leq \kappa$. We further assume for all terms H_X, $\text{diam}(X) \leq \mathcal{O}(1)$. We also define the Hamiltonian restricted to a region $A \subseteq \Lambda$ by $H_A = \sum_{X \subseteq A} H_X$.

In order to describe our Hamiltonians, we consider an orthogonal Hermitian basis for the space of operators acting on each qudit. For instance, for qubits, this basis consists of the Pauli operators. By decomposing each local term H_X in terms of the tensor product of such basis operators, we find the following canonical form for the Hamiltonian H:

Definition 79 (Canonical representation for κ-local Hamiltonians)**.** *A κ-local Hamiltonian H on a lattice Λ can be written as a sum of m Hermitian operators E_ℓ each having a connected support with respect to Λ and acting non-trivially on at most κ qudits. That is,*

$$H = \sum_{\ell=1}^{m} \mu_\ell E_\ell. \tag{4.16}$$

where $\mu_\ell \in \mathbb{R}$ *and we assume* $\|E_\ell\| \leq 1$, $\mathrm{tr}[E_\ell^2] = \mathcal{D}_\Lambda$, $E_\ell^\dagger = E_\ell$ *for* $\ell \in [m]$, *and*

$$\mathrm{tr}[E_k E_\ell] = 0 \quad \textit{for } k \neq \ell. \tag{4.17}$$

Since we assume H is geometrically local (i.e. it is defined over a lattice), it holds that m the number of local terms in (4.16) satisfies $m = \mathcal{O}(|\Lambda|) = \mathcal{O}(n)$.

As discussed earlier, we extensively use the notion of quasi-local operators, which we now formally define.

Definition 80 (Quasi-local operators)**.** *An operator* A *is said to be* (τ, a_1, a_2, ζ)-quasi-local *if it can be written as*

$$\begin{aligned} A &= \sum_{\ell=1}^{n} g_\ell \bar{A}_\ell \quad \text{with} \quad g_\ell \leq a_1 \cdot \exp(-a_2 \ell^\tau), \\ \bar{A}_\ell &= \sum_{|Z|=\ell} a_Z, \quad \max_{i \in \Lambda} \left(\sum_{Z : Z \ni i} \|a_Z\| \right) \leq \zeta, \end{aligned} \tag{4.18}$$

where the sets $Z \subset \Lambda$ *are restricted to be balls.*[3]

Although local operators are morally a special case of quasi-local operators (when $\tau = \infty$), we will reserve the above notation for operators with $\tau \leq 1$. A useful tool for analyzing quasi-locality is the Lieb-Robinson bound, which shows a light-cone like behavior of the time evolution operator.

Fact 81 (Lieb-Robinson bound [LR72], [NS09])**.** *Let* A, B *be operators supported on regions* X, Y *of the* D *dimensional lattice* Λ *respectively. Also, let* H *be a geometrically local Hamiltonian. There exist* $\mathcal{O}(1)$ *constants* v_{LR}, f, c *that only depend on the details of the Hamiltonian such that*

$$\|[e^{iHt} A e^{-iHt}, B]\| \leq f \|A\| \|B\| \cdot \min\left(|\partial X|, |\partial Y|\right) \cdot \min\left(e^{c(v_{\mathrm{LR}}|t| - \mathrm{dist}(X,Y))}, 1\right).$$

Gibbs states. At an *inverse-temperature* β, a quantum many-body system with the Hamiltonian $H(\mu)$ is in the Gibbs (thermal) state if it is given by

$$\rho_\beta(\mu) = \frac{e^{-\beta H(\mu)}}{\mathrm{tr}[e^{-\beta H(\mu)}]}. \tag{4.19}$$

The partition function of this system is defined by $Z_\beta(\mu) = \mathrm{tr}[e^{-\beta H(\mu)}]$.

Remark 82. *In our notation, we sometimes drop the dependency of the partition function or the Gibbs state on* μ*. We also often simply use the term local Hamiltonian* H *or quasi-local operator* A *when referring to Definition 128 and Definition 80.*

[3] The assumption that Z is a ball suffices for us. Our results on quasi-local operators also generalize to the case where Z is an arbitrary *regular* shape, for example when the radii of the balls inscribing and inscribed by Z are of constant proportion to each other.

4.6.3 Quantum belief propagation

Earlier we saw that we could express the Gibbs state of a Hamiltonian H by Eq. (4.19). Suppose we alter this Hamiltonian by adding a term V such that

$$H(s) = H + sV, \quad s \in [0,1]. \tag{4.20}$$

How does the Gibbs state associated with this Hamiltonian change? If the new term V commutes with the Hamiltonian H, i.e., $[H,V]=0$, then the derivative of the Gibbs state of $H(s)$ is given by

$$\frac{d}{ds}e^{-\beta H(s)} = -\beta e^{-\beta H(s)}V = -\frac{\beta}{2}\left\{e^{-\beta H(s)}, V\right\}, \tag{4.21}$$

where $\{e^{-\beta H(s)}, V\} = e^{-\beta H(s)}V + Ve^{-\beta H(s)}$ denotes the anti-commutator. In the non-commuting case though, finding this derivative is more complicated. The *quantum belief propagation* is a framework developed in [Has07b, Kim17, KBa19b] for finding such derivatives in a way that reflects the locality of the system.

Definition 83 (Quantum belief propagation operator)**.** *For every $s \in [0,1]$, $\beta \in \mathbb{R}$, define $H(s) = H + sV$ where $V = \sum_{j,k} V_{j,k}|j\rangle\langle k|$ is a Hermitian operator. Also let $f_\beta(t)$ be a function whose Fourier transform is*

$$\tilde{f}_\beta(\omega) = \frac{\tanh(\beta\omega/2)}{\beta\omega/2}, \tag{4.22}$$

i.e., $f_\beta(t) = \frac{1}{2\pi}\int d\omega \tilde{f}_\beta(\omega)e^{i\omega t}$. The quantum belief propagation operator $\Phi_{H(s)}(V)$ is defined by

$$\Phi_{H(s)}(V) = \int_{-\infty}^{\infty} dt f_\beta(t)\ e^{-iH(s)t}\ V\ e^{iH(s)t}.$$

Equivalently, in the energy basis of $H(s) = \sum_j \mathcal{E}_j(s)\ |j\rangle\langle j|$, we can write

$$\Phi_{H(s)}(V) = \sum_{j,k} |j\rangle\langle k|\ V_{j,k}\ \tilde{f}_\beta(\mathcal{E}_j(s) - \mathcal{E}_k(s)). \tag{4.23}$$

Proposition 84 (cf. [Has07b])**.** *In the same setup as Definition 83, it holds that*

$$\frac{d}{ds}e^{-\beta H(s)} = -\frac{\beta}{2}\left\{e^{-\beta H(s)}, \Phi_{H(s)}(V)\right\}. \tag{4.24}$$

4.6.4 Change in the spectrum after applying local operators

For a Hamiltonian H, let $P^H_{\leq x}$ and $P^H_{\geq y}$ be projection operators onto the eigenspaces of H whose energies are in $\leq x$ and $\geq y$, respectively (we use similar notation $P^A_{\leq x}, P^A_{\geq y}$ for the quasi-local operator A). Consider a quantum state $|\psi\rangle$ in the low-energy part of the spectrum such that $P^H_{\leq x}|\psi\rangle = |\psi\rangle$. Suppose this states $|\psi\rangle$ is perturbed by applying a local operator O_X on a subset $X \subset \Lambda$ of its qudits. Intuitively, we expect that the operator O_X only affects the energy of $|\psi\rangle$ up to $\mathcal{O}(|X|)$, i.e., $\|P^H_{\geq y}O_X|\psi\rangle\| \approx 0$ for $y \gg x + |X|$. A simple example is when $|\psi\rangle$ is the eigenstate of a classical spin system. By applying a local operation that flips the spins in a small region X, the

energy changes at most by $\mathcal{O}(|X|)$. The following lemma rigorously formulates the same classical intuition for quantum Hamiltonians

Lemma 85 (Theorem 2.1 of [AKL16]). *Let H be an arbitrary κ-local operator such that*

$$H = \sum_{|Z|\leq\kappa} h_Z, \quad \max_{i\in\Lambda} \sum_{Z:Z\ni i} \|h_Z\| \leq g. \tag{4.25}$$

Then, for an arbitrary operator O_X which is supported on $X \subseteq \Lambda$, the operator norm of $P^H_{\geq y} O_X P^H_{\leq x}$ is upper-bounded by

$$\|P^H_{\geq y} O_X P^H_{\leq x}\| \leq \|O_X\| \cdot \exp\big(-\frac{1}{2g\kappa}(y - x - 2g|X|)\big). \tag{4.26}$$

In our analysis, we need an different version of this lemma for *quasi-local* operators instead of κ-local operators. The new lemma will play a central role in lower-bounding the variance of quasi-local operators. The proof follows by the analysis of a certain moment function (as opposed to the moment generating function in [AKL16]). Due to formal similarities between the proofs, we defer the proof of the next lemma to Section 4.10.3.

Lemma 86 (Variation of [AKL16] for quasi-local operators). *Let A be a $(\tau, a_1, a_2, 1)$-quasi-local operator, as given in Eq. (4.18), with $\tau \leq 1$. For an arbitrary operator O_X supported on a subset $X \subseteq \Lambda$ with $|X| = k_0$ and $\|O_X\| = 1$, we have*

$$\|P^A_{\geq x+y} O_X P^A_{\leq x}\| \leq c_5 \cdot k_0 \exp\Big(-(\lambda_1 y/k_0)^{1/\tau_1}\Big), \tag{4.27}$$

where $\tau_1 := \frac{2}{\tau} - 1$, c_5 and λ_1 are constants depending on a_1 and a_2 as $c_5 \propto a_2^{2/\tau}$ and $\lambda_1 \propto a_2^{-2/\tau}$ respectively.

4.6.5 Local reduction of global operators

An important notion in our proofs will be a reduction of a global operator to a local one, which has influence on a site i. Fix a subset $Z \subseteq \Lambda$ and an operator O supported on Z. Define

$$O_{(i)} := O - \mathrm{tr}_i[O] \otimes \frac{\mathbb{1}_i}{d}, \tag{4.28}$$

where operator $\mathbb{1}_i$ is the identity operator on the ith site, d is the local dimension, tr_i is the partial trace operation with respect to the site i. Note that $O_{(i)}$ removes all the terms in O that do not act on the ith site. This can be explicitly seen by introducing a basis $\{E^\alpha_Y\}_{\alpha\in\mathbb{N}, Y\subseteq Z}$ of Hermitian operators, where Y labels the support of E^α_Y and α labels several possible operators on the same support. We can assume that $\mathrm{tr}[(E^\alpha_Y)^2] = \mathcal{D}_\Lambda, \mathrm{tr}_i[E^\alpha_Y] = 0$ for every $i \in Y$, and the orthogonality condition $\mathrm{tr}[E^\alpha_Y E^{\alpha'}_{Y'}] = 0$ holds if $\alpha \neq \alpha'$ or $Y \neq Y'$. These conditions are satisfied by the appropriately normalized Pauli operators. Expand

$$O = \sum_{\alpha,Y} g_{\alpha,Y} E^\alpha_Y.$$

Then

$$O_{(i)} = \sum_{\alpha,Y} g_{\alpha,Y} E_Y^\alpha - \sum_{\alpha,Y} g_{\alpha,Y} \mathrm{tr}_i[E_Y^\alpha] \otimes \frac{\mathbb{1}_i}{d} = \sum_{\alpha,Y:Y\ni i} g_{\alpha,Y} E_Y^\alpha.$$

Thus, $O_{(i)}$ is an operator derived from O, by removing all E_Y^α which act as identity on i. The following claim shows that the Frobenius norm of a typical $O_{(i)}$ is not much small in comparison to the Frobenius norm of O.

Claim 87. *For every operator O and $O_{(i)}$ defined in Eq. (4.28), it holds that*

$$\max_{i\in Z} \|O_{(i)}\|_F^2 \geq \frac{1}{|Z|} \sum_{i\in Z} \|O_{(i)}\|_F^2 \geq \frac{1}{|Z|} \|O\|_F^2. \tag{4.29}$$

Proof. Using the identities $\mathrm{tr}[E_Y^\alpha E_{Y'}^{\alpha'}] = 0$ and $\mathrm{tr}[(E_Y^\alpha)^2] = \mathcal{D}_\Lambda$, we have

$$\|O\|_F^2 = \mathcal{D}_\Lambda \sum_{Y,\alpha} g_{\alpha,Y}^2 \leq \mathcal{D}_\Lambda \sum_{i\in Z} \sum_{\alpha,Y:Y\ni i} g_{\alpha,Y}^2 = \sum_{i\in Z} \|O_{(i)}\|_F^2,$$

where the inequality comes from the fact that O is supported on the subset Z (i.e., $Y \subseteq Z$). This completes the proof. □

4.7 A lower bound on the sample complexity

In this section, we prove a lower bound on the sample complexity of the quantum Hamiltonian learning problem as claimed earlier in Theorem 69.

Theorem 88 (Restatement of Theorem 69)**.** *The number of copies N of the Gibbs state needed to solve the quantum Hamiltonian learning problem and outputs a $\hat{\mu}$ satisfying $\|\hat{\mu} - \mu\|_2 \leq \varepsilon$ with probability $1 - \mathrm{err}$ is lower bounded by*

$$N \geq \Omega\Big(\frac{\sqrt{m} + \log(1-\mathrm{err})}{\beta\varepsilon}\Big).$$

Proof. In order to prove the lower bound, we consider learning the parameters $\mu \in \mathbb{R}^m$ of the following class of one-local Hamiltonians on m qubits:

$$H(\mu) = \sum_{i=1}^m \mu_i |1\rangle\langle 1|_i,$$

where $|1\rangle\langle 1|_i$ is projection onto a basis $|1\rangle$ on the ith qudit. Let $T_m : \{\mu \in \mathbb{R}_+^m : \sum_i \mu_i^2 \leq 100\varepsilon^2\}$ be an orthant of the hypersphere of radius θ in $\mathbb{R}_+^m$. We have the following claim.

Claim 89. *There exists a collection of 2^m points in T_m, such that the ℓ_2 distance between each pair is $\geq \varepsilon$.*

Proof. Pick 2^m points uniformly at random in T_m. By union bound, the probability that at least one pair is at a distance of at most ε is at most $(2^m)^2$ times the probability that a fixed pair of points is at a distance of at most ε. But the latter probability is upper bounded by the ratio between

the volume of a hypersphere of radius ε and the volume of T_m, which is $\frac{\varepsilon^m}{(10\varepsilon)^m/2^m} = \frac{1}{5^m}$. Since $(2^m)^2 \frac{1}{5^m} < 1$, the claim concludes. □

Let these set of 2^m points be S. For some temperature $\beta > 0$ and unknown $\mu \in S$, suppose $\mathcal{A}$ is an algorithm that is given N copies of $\rho_\beta(\mu)$ and, with probability $1 - \mathrm{err}$, outputs μ' satisfying $\|\mu' - \mu\|_2 \leq \varepsilon$. We now use $\mathcal{A}$ to assign the estimated $\hat{\mu}$ to exactly one of the parameters μ. Once the learning algorithm obtains an output μ', we can find the closest point in S (in ℓ_2 distance) as our estimate of μ, breaking ties arbitrarily. With probability $1 - \mathrm{err}$, the closest $\mu \in S$ to μ' is the correct μ since by the construction of S, $\|\mu' - \mu\|_2 \leq \varepsilon$. Thus, the algorithm $\mathcal{A}$ can be used to solve the problem of estimating the parameters μ themselves (not only approximating it). We furthermore show that the number of samples required to estimate $\mu \in S$ is large using lower bounds in the quantum state discrimination. We will directly use the lower bound from [HKK08] (as given in [HW12]). Before we plug in their formula, we need to bound on the spectral norm of $\rho_\beta(\mu)$ for an arbitrary $\mu \in S$ (denoted by $\|\rho_\beta(\mu)\|$). That is,

$$\begin{aligned}
\max_{\mu\in S}\{2^m\|\rho_\beta(\mu)\|\} &= \max_{\mu\in S} 2^m \left(\bigotimes_{i=1}^m \left\| \frac{1}{1+e^{-\beta\mu_i}}|0\rangle\langle 0| + \frac{e^{-\beta\mu_i}}{1+e^{-\beta\mu_i}}|1\rangle\langle 1| \right\|\right) \\
&= \max_{\mu\in S} \left(\bigotimes_{i=1}^m \left| \frac{2}{1+e^{-\beta\mu_i}} \right|\right) \\
&= \max_{\mu\in S} \left(\bigotimes_{i=1}^m \left| \frac{2e^{\beta\mu_i}}{e^{\beta\mu_i}+1} \right|\right) \\
&\leq \max_{\mu\in S} \left(\bigotimes_{i=1}^m \left| \frac{2e^{\beta\mu_i}}{2} \right|\right) = \max_{\mu\in S} \left(e^{\beta\sum_{i=1}^m \mu_i}\right) \leq e^{\beta\sqrt{m}\sqrt{\sum_{i=1}^m \mu_i^2}} \leq e^{\beta\sqrt{m}\cdot 10\varepsilon},
\end{aligned}$$

since $\sum_i \mu_i^2 \leq 100\varepsilon^2$ for all $i \in S$. Thus, the lower bound for state identification of $\{H(\mu) : \mu \in S\}$ in [HW12, Equation 2] (cf. [HKK08] for the original statement) implies that

$$N \geq \frac{\log|S| + \log(1-\mathrm{err})}{\log\left(\max_{\mu\in S}\{2^m\|\rho(\mu)_\beta\|\}\right)} = \frac{m\log 2 + \log(1-\mathrm{err})}{10\sqrt{m}\beta\varepsilon} = \Omega\Big(\frac{\sqrt{m} + \log(1-\mathrm{err})}{\varepsilon\beta}\Big).$$

This establishes the lower bound. □

4.8 Efficient sample complexity assuming strong convexity

In this section, we provide more details on how one can find rigorous upper bounds on the sample complexity of the Hamiltonian learning problem using the notions of sufficient statistics and the strong convexity from the fields of stochastic convex optimization and statistics.

4.8.1 Sufficient statistics

As discussed in the main text, local expectations of the Gibbs states can be used to uniquely specify these states. This provides us with "sufficient statistics" for learning the Hamiltonians from ρ_β. This observation is formalized in the following result:

Proposition 90 (Restatement of Proposition 70)**.** *Consider the following two Gibbs states*

$$\rho_\beta(\mu) = \frac{e^{-\beta \sum_\ell \mu_\ell E_\ell}}{\mathrm{tr}[e^{-\beta \sum_\ell \mu_\ell E_\ell}]}, \quad \rho_\beta(\lambda) = \frac{e^{-\beta \sum_\ell \lambda_\ell E_\ell}}{\mathrm{tr}[e^{-\beta \sum_\ell \lambda_\ell E_\ell}]} \tag{4.30}$$

such that $\mathrm{tr}[\rho_\beta(\lambda)E_j] = \mathrm{tr}[\rho_\beta(\mu)E_j]$ *for all* $j \in [m]$*, i.e. all the* κ*-local local expectations of* $\rho_\beta(\lambda)$ *match that of* $\rho_\beta(\mu)$*. Then, we have* $\rho_\beta(\lambda) = \rho_\beta(\mu)$*, which in turns implies* $\lambda_\ell = \mu_\ell$ *for* $\ell \in [m]$*.*

Proof. We consider the relative entropy between $\rho_\beta(\lambda)$ and the Gibbs state $\rho_\beta(\mu)$. We have

$$\begin{aligned} S\left(\rho_\beta(\mu)\|\rho_\beta(\lambda)\right) &= \mathrm{tr}\left[\rho_\beta(\mu)\left(\log \rho_\beta(\mu) - \log \rho_\beta(\lambda)\right)\right] \\ &= -S(\rho_\beta(\mu)) + \beta \cdot \mathrm{tr}\left[\rho_\beta(\mu) \sum_\ell \lambda_\ell E_\ell\right] + \log Z(\lambda) \qquad (4.31) \\ &\overset{(1)}{=} -S(\rho_\beta(\mu)) + \beta \sum_\ell \lambda_\ell \mathrm{tr}[\rho_\beta(\lambda)E_\ell] + \log Z(\lambda) \qquad (4.32) \\ &= -S(\rho_\beta(\mu)) + S(\rho_\beta(\lambda)) \\ &\overset{(2)}{\geq} 0, \qquad (4.33) \end{aligned}$$

where (1) follows because $\mathrm{tr}[\rho_\beta(\mu)E_\ell] = \mathrm{tr}[\rho_\beta(\lambda)E_\ell]$ for all $\ell \in [m]$ and (2) used the positivity of relative entropy. Similarly, we can exchange the role of $\rho(\mu)$ and $\rho(\lambda)$ in (4.33) and get

$$S\left(\rho_\beta(\lambda)\|\rho_\beta(\mu)\right) = -S(\rho_\beta(\lambda)) + S(\rho_\beta(\mu)) \geq 0. \tag{4.34}$$

Combining these bounds imply $S(\rho_\beta(\mu)) = S(\rho_\beta(\lambda))$ and hence from Eq. (4.33), we get $S(\rho_\beta(\mu)\|\rho_\beta(\lambda)) = 0$. It is known that the relative entropy of two distribution is zero only when $\rho_\beta(\mu) = \rho_\beta(\lambda)$. Hence, we also have $\log \rho_\beta(\mu) = \log \rho_\beta(\lambda)$ or equivalently up to an additive term $\sum_{\ell=1}^m \mu_\ell E_\ell = \sum_{\ell=1}^m \lambda_\ell E_\ell$. Since the operators E_ℓ form an orthogonal basis (see Eq. (4.17)), we see that $\lambda_\ell = \mu_\ell$ for all $\ell \in [m]$. □

Remark 91. *When the local expectations of the two Gibbs states only approximately match, i.e.,*

$$|\mathrm{tr}[\rho_\beta(\mu)E_\ell] - \mathrm{tr}[\rho_\beta(\lambda)E_\ell]| \leq \delta$$

for $\ell \in [m]$*, then a similar argument to* (4.33) *shows that* $S(\rho_\beta(\mu)\|\rho_\beta(\lambda)) \leq \mathcal{O}(m\delta)$*. By applying Pinsker's inequality*[4]*, we get* $\|\rho_\beta(\mu) - \rho_\beta(\lambda)\|_1^2 \leq \mathcal{O}(m\delta) = \mathcal{O}(n\delta)$*, where we used the fact that for geometrically-local Hamiltonians* $m = \mathcal{O}(n)$*.*

Given Proposition 90, we ask if there is an algorithm that finds the Gibbs state from its local expectations. A solution to this question is to use the *maximum entropy optimization* [Jay57b, BKL+19]. This is also closely related to the problem of *maximum likelihood estimation* that is used often to analyze the sample complexity of statistical problems. Formally, this problem can be expressed as follows:

[4]Pinsker's inequality states that for two density matrices ρ, σ, we have $\|\rho - \sigma\|_1^2 \leq 2 \ln 2 \cdot S(\rho\|\sigma)$.

$$\begin{aligned} \max_{\rho} \quad & S(\rho) \\ \text{s.t.} \quad & \mathrm{tr}[\rho E_\ell] = e_\ell, \quad \forall \ell \in [m] \\ & \rho > 0, \quad \mathrm{tr}[\rho] = 1. \end{aligned} \tag{4.35}$$

where $S(\rho) = -\mathrm{tr}[\rho \log \rho]$ is the *von Neumann entropy* of ρ.

Below, we will be mainly concerned with the convex dual of this optimization which directly produces the interaction coefficients μ. Understanding the effect of statistical errors on the dual optimization program (and in more generality, stochastic convex optimization) will be the main the subject of the next few sections.

4.8.2 Stochastic convex optimization

Suppose we want to solve the optimization

$$\max_{x \in \mathbb{R}^m} f(x)$$

for a function $f : \mathbb{R}^m \to \mathbb{R}$ which is of the form $f(x) = \mathbb{E}_{y \sim \mathcal{D}}[g(x, y)]$. Here $g(x, y)$ is some convex function and the expectation $\mathbb{E}_{y \sim \mathcal{D}}$ is taken with respect to an unknown distribution $\mathcal{D}$. Algorithms for this maximization problem are based on obtaining i.i.d. *samples* y drawn from the distribution $\mathcal{D}$. In practice, we can only receive finite samples $y_1, y_2, \ldots, y_\ell$ from such a distribution. Hence, instead of the original optimization, we solve an *empirical* version

$$\max_{x \in \mathbb{R}^m} \frac{1}{\ell} \sum_{k=1}^{\ell} g(x, y_k).$$

The natural question therefore is: How many samples ℓ do we need to guarantee the output of the empirical optimization is close to the original solution? One answer to this problem relies on a property of the objective function known as *strong convexity*.

Definition 92 (restatement of Definition 71)**.** *Consider a convex function $f : \mathbb{R}^m \mapsto \mathbb{R}$ with gradient $\nabla f(x)$ and Hessian $\nabla^2 f(x)$ at x. The function f is said to be α-strongly convex in its domain if it is differentiable and for all x, y, and*

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) + \frac{1}{2} \alpha \|y - x\|_2^2,$$

or equivalently if the minimum eigenvalue of the Hessian $\sigma_{\min}(\nabla^2 f(x))$ satisfies

$$\sigma_{\min}(\nabla^2 f(x)) \geq \alpha.$$

In other words, for any vector $v \in \mathbb{R}^m$ it holds that $\sum_{i,j} v_i v_j \frac{\partial^2}{\partial x_i \partial x_j} f(x) \geq \alpha \|v\|_2^2$.

Next, we discuss how the framework of convex optimization and in particular strong convexity, can be applied to the Hamiltonian learning problem. To this end, we define the following optimization problems based on the convex dual of the optimization (4.35)

Definition 93 (Optimization program for learning the Hamiltonian)**.** *We denote the objective function in the Hamiltonian learning problem and its approximate version by $L(\lambda)$ and $\hat{L}(\lambda)$ respectively, i.e.,*

$$L(\lambda) = \log Z_\beta(\lambda) + \beta \cdot \sum_{\ell=1}^{m} \lambda_\ell e_\ell, \quad \hat{L}(\lambda) = \log Z_\beta(\lambda) + \beta \cdot \sum_{\ell=1}^{m} \lambda_\ell \hat{e}_\ell, \tag{4.36}$$

where the partition function is given by $Z_\beta(\lambda) = \mathrm{tr}\big(e^{-\beta \sum_{\ell=1}^{m} \lambda_\ell E_\ell}\big)$. The parameters of the Hamiltonian that we intend to learn are $\mu = \arg\min_{\lambda \in \mathbb{R}^m : \|\lambda\| \le 1} L(\lambda)$. As before, we also define the empirical version of this optimization by

$$\hat{\mu} = \operatorname*{arg\,min}_{\lambda \in \mathbb{R}^m : \|\lambda\| \le 1} \hat{L}(\lambda). \tag{4.37}$$

We prove later in Lemma 100 that $\log Z_\beta(\lambda)$ is a convex function in parameters λ and thus, the optimization in (4.37) is a convex program whose solution can be in principle algorithmically found. In this work, we do not constraint ourselves with the running time of solving (4.37). We instead obtain sample complexity bounds as formulated more formally in the next theorem.

Theorem 94 (Restatement of Theorem 73)**.** *Let $\delta, \alpha > 0$. Suppose the local expectations e_ℓ are determined up to error δ, i.e., $|e_\ell - \hat{e}_\ell| \le \delta$ for all $\ell \in [m]$. Additionally, assume the strong convexity condition $\sigma_{\min}(\nabla^2 \log Z(\lambda)) \ge \alpha$ holds for $\|\lambda\| \le 1$. Then the optimal solution to the program (4.37) satisfies*

$$\|\mu - \hat{\mu}\|_2 \le \frac{2\beta\sqrt{m}\delta}{\alpha}. \tag{4.38}$$

Proof. From the definition of $\hat{\mu}$ as the optimal solution of $\hat{L}$ in (4.37), we see that $\hat{L}(\hat{\mu}) \le \hat{L}(\mu)$. Thus, we get

$$\log Z_\beta(\hat{\mu}) + \beta \cdot \sum_{\ell=1}^{m} \hat{\mu}_\ell \hat{e}_\ell \le \log Z_\beta(\mu) + \beta \cdot \sum_{\ell=1}^{m} \mu_\ell \hat{e}_\ell.$$

or equivalently,

$$\log Z_\beta(\hat{\mu}) \le \log Z_\beta(\mu) + \beta \cdot \sum_{\ell=1}^{m} (\mu_\ell - \hat{\mu}_\ell) \hat{e}_\ell. \tag{4.39}$$

We show later in Lemma 99 that for every $\ell \in [m]$, we have $\frac{\partial}{\partial \mu_\ell} \log Z_\beta(\mu) = -\beta e_\ell$.[5] This along with the assumption $\sigma_{\min}(\nabla^2 \log Z(\mu)) \ge \alpha$ in the theorem statement, implies that for every μ' with $\|\mu'\| \le 1$

$$\log Z_\beta(\mu') \ge \log Z_\beta(\mu) - \beta \cdot \sum_{\ell=1}^{m} (\mu'_\ell - \mu_\ell) e_\ell + \frac{1}{2}\alpha \|\mu' - \mu\|_2^2. \tag{4.40}$$

[5] In particular, see Eq. (4.45), where we showed $\frac{\partial}{\partial \mu_\ell} \log Z_\beta(\mu) = -\beta \cdot \mathrm{tr}[E_\ell \rho_\beta(\mu)] = -\beta e_\ell$.

Hence, by choosing $\mu' = \hat{\mu}$ and combining (4.40) and (4.39), we get

$$\log Z_\beta(\mu) - \beta \cdot \sum_{\ell=1}^{m} (\hat{\mu}_\ell - \mu_\ell) e_\ell + \frac{1}{2}\alpha \|\hat{\mu} - \mu\|_2^2 \leq \log Z_\beta(\mu) + \beta \cdot \sum_{\ell=1}^{m} (\mu_\ell - \hat{\mu}_\ell)\hat{e}_\ell$$

which further implies that

$$\begin{aligned} \frac{1}{2}\alpha \|\hat{\mu} - \mu\|_2^2 &\leq \beta \cdot \sum_{\ell=1}^{m} (\hat{\mu}_\ell - \mu_\ell)(e_\ell - \hat{e}_\ell), \\ &\leq \beta \cdot \|\hat{\mu} - \mu\|_2 \cdot \|\hat{e} - e\|_2. \end{aligned}$$

Hence, we have

$$\|\hat{\mu} - \mu\|_2 \leq \frac{2\beta}{\alpha} \|\hat{e} - e\|_2 \leq \frac{2\beta\sqrt{m}\delta}{\alpha}.$$

□

Remark 95. *We note that the bound (4.38) in Theorem 94 can be also derived when other methods are used instead of the maximum entropy optimization or its dual in Definition 93. Suppose we use an alternative approach, which could be an approximate version of the maximum entropy method, a heuristic quantum algorithm that variationally prepares the Gibbs state, or one that infers the interactions in a different way without relying on the thermal averages. As long as the interaction coefficients inferred via such method (denoted by $\tilde{\mu}$) satisfy $\|\tilde{\mu}\| \leq 1$ and their corresponding Gibbs state possesses local expectations $\tilde{e}_k$ within δ distance of the original values, a similar expression as in (4.38) bounds the error of that approach as well.*

More formally strong convexity leads to Equation 4.40, which says

$$\log Z_\beta(\tilde{\mu}) \geq \log Z_\beta(\mu) - \beta \cdot \sum_{\ell=1}^{m} (\tilde{\mu}_\ell - \mu_\ell) e_\ell + \frac{1}{2}\alpha \|\tilde{\mu} - \mu\|_2^2.$$

Switching μ and $\tilde{\mu}$ as well as e_ℓ and $\tilde{e}_\ell$, we have

$$\log Z_\beta(\mu) \geq \log Z_\beta(\tilde{\mu}) - \beta \cdot \sum_{\ell=1}^{m} (\mu_\ell - \tilde{\mu}_\ell) \tilde{e}_\ell + \frac{1}{2}\alpha \|\tilde{\mu} - \mu\|_2^2.$$

Adding these equations and cancelling some terms, we obtain

$$0 \geq \beta \cdot \sum_{\ell=1}^{m} (\mu_\ell - \tilde{\mu}_\ell)(e_\ell - \tilde{e}_\ell) + \alpha \|\tilde{\mu} - \mu\|_2^2.$$

This leads to the inequality (using Cauchy-Schwartz)

$$\alpha \|\tilde{\mu} - \mu\|_2^2 \leq \beta \cdot \|\tilde{\mu} - \mu\|_2 \cdot \|\tilde{e} - e\|_2,$$

which is equivalent to

$$\|\tilde{\mu} - \mu\|_2 \leq \frac{\beta}{\alpha} \|\tilde{e} - e\|_2.$$

Finally, we obtain the following upper bound on the sample complexity assuming the log-partition function is strongly convex.

Corollary 96 (Sample complexity from strong convexity)**.** *Under the same conditions as in Theorem 94, the number of copies of the Gibbs state ρ_β that suffice to solve the quantum Hamiltonian learning problem and outputs a $\hat{\mu}$ satisfying $\|\hat{\mu} - \mu\|_2 \leq \varepsilon$ with probability $1 - \mathrm{err}$ is*

$$N = \mathcal{O}\left(\frac{\beta^2 2^{\mathcal{O}(\kappa)}}{\alpha^2 \varepsilon^2} \cdot m \cdot \log \frac{m}{\mathrm{err}}\right).$$

Proof. First observe that, using Theorem 94, as long as the error in estimating the local expectations e_ℓ are

$$\delta \leq \frac{\alpha \varepsilon}{2\beta\sqrt{m}}, \tag{4.41}$$

we estimate the coefficients μ by $\hat{\mu}$ such that $\|\hat{\mu} - \mu\|_2 \leq \varepsilon$. The local expectations e_ℓ can be estimated in various ways. One method considered in [CW20, BMBO20] is to group the operators E_ℓ into sets of mutually commuting observables and simultaneously measure them at once. Alternatively, we can use the recent procedure in [HKP20, Theorem 1] based on a variant of shadow tomography. In either case, the number of copies of the state needed to find all the local expectations with accuracy δ and success probability $1 - \mathrm{err}$ is

$$N = \mathcal{O}\left(\frac{2^{\mathcal{O}(\kappa)}}{\delta^2} \log \frac{m}{\mathrm{err}}\right),$$

where recall that κ is the locality of the Hamiltonian. Plugging in Eq. (4.41) gives us the final bound

$$N = \mathcal{O}\left(\frac{\beta^2 2^{\mathcal{O}(\kappa)}}{\alpha^2 \varepsilon^2} m \log \frac{m}{\mathrm{err}}\right).$$

□

Remark 97 (Final upper bound on the sample complexity)**.** *In the next section, we prove that the α parameter in Corollary 96 is $\alpha \leq e^{-\Theta(\beta^c)} \cdot \frac{\beta^{c'}}{m}$. This implies the claimed sample complexity for the Hamiltonian learning problem in the main text, i.e.*

$$N = \mathcal{O}\left(\frac{e^{\mathcal{O}(\beta^c)}}{\beta^{\tilde{c}} \varepsilon^2} \cdot m^3 \cdot \log \frac{m}{\mathrm{err}}\right). \tag{4.42}$$

Notice that although we assume in the main that err*, the probability of error, is a constant (e.g.,* 0.01*), the dependency on* err *in* (4.42) *only appears in the logarithm and can be suitably scaled.*

4.9 Proof of strong convexity of quantum log-partition functions

In this section we prove the strong convexity of quantum log-partition functions.

Theorem 98. *Let $H = \sum_{\ell=1}^{m} \mu_\ell E_\ell$ be a κ-local Hamiltonian over a finite dimensional lattice. For a given inverse-temperature β, there are constants $c, c' > 3$ depending on the geometric properties of the lattice such that the following strong convexity property (see Definition 92) for $\log Z_\beta(\mu)$ holds*

$$\sigma_{\min}(\nabla^2 \log Z_\beta(\mu)) \geq e^{-\Theta(\beta^c)} \cdot \frac{\beta^{c'}}{m}. \tag{4.43}$$

In other words, for every vector $v \in \mathbb{R}^m$ we have $v^\top \cdot \nabla^2 \log Z_\beta(\mu) \cdot v \geq \beta^{c'} \frac{e^{-\Theta(\beta^c)}}{m} \cdot \|v\|_2^2$[6].

The proof of Theorem 98 is explained in multiple steps. In the next few sections we first review the proof of this statement for the classical Hamiltonians. Then, we give a brief overview of the main steps in the proof of the quantum case.

4.9.1 Step 1: Relating the Hessian to a variance

We begin our detailed proofs with finding an expression for the Hessian of $\log Z_\beta(\lambda)$ in terms of the variance of a quasi-local operator.

Lemma 99. *For every vector $v \in \mathbb{R}^m$, define the local operator $W_v = \sum_{i=1}^{m} v_i E_i$ (for notational convenience, later on we stop subscripting W by v). The Hessian $\nabla^2 \log Z_\beta(\lambda)$ satisfies*

$$v^\top \cdot \big(\nabla^2 \log Z_\beta(\lambda)\big) \cdot v = \frac{\beta^2}{2} \mathrm{tr}\Big[\big\{W_v, \Phi_{H(\lambda)}(W_v)\big\}\rho_\beta(\lambda)\Big] - \beta^2 \big(\mathrm{tr}\left[W_v \rho_\beta(\lambda)\right]\big)^2, \tag{4.44}$$

Proof. Since the terms in the Hamiltonian are non-commuting, we use Proposition 84 to find the derivatives of $\log Z_\beta(\lambda)$. We get

$$\begin{aligned}
\frac{\partial}{\partial \lambda_j} \log Z_\beta(\lambda) &= \frac{1}{Z_\beta(\lambda)} \mathrm{tr}\left[-\frac{\beta}{2}\left\{e^{-\beta H(\lambda)}, \Phi_{H(\lambda)}(E_j)\right\}\right] \\
&= \frac{-\beta}{Z_\beta(\lambda)} \mathrm{tr}\left[e^{-\beta H(\lambda)} \int_{-\infty}^{\infty} dt f_\beta(t) e^{-iH(\lambda)t}\ E_j\ e^{iH(\lambda)t}\right] \\
&= -\beta\ \mathrm{tr}\left[E_j \frac{e^{-\beta H(\lambda)}}{Z_\beta(\lambda)}\right],
\end{aligned} \tag{4.45}$$

where the second equality used the definition of the quantum belief propagation operator $\Phi_{H(\lambda)}(E_j) = \int_{-\infty}^{\infty} dt f_\beta(t)\ e^{-iH(\lambda)t}\ E_j\ e^{iH(\lambda)t}$ with f_β as given in Definition 83. The third equality

[6] In future sections, we often use this characterization in terms of vectors v to obtain proper bounds on the minimum eigenvalue $\sigma_{\min}$.

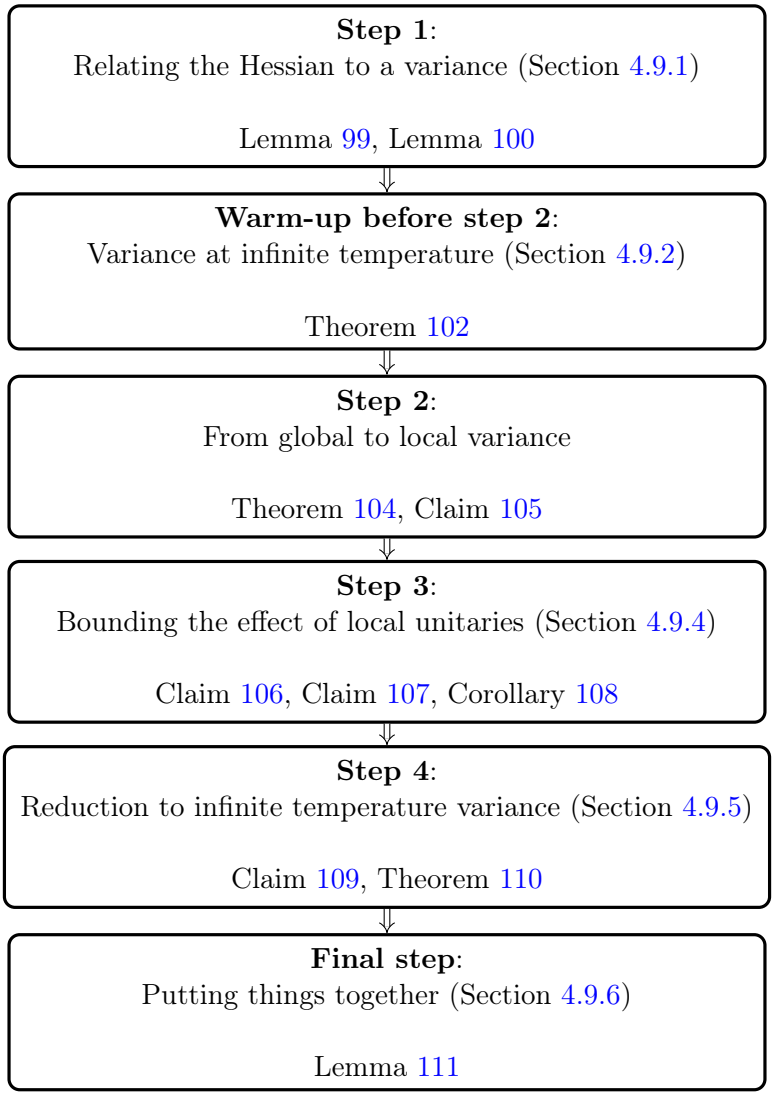


Figure 4-1: Flow of the argument in the proof of Theorem 98.

used the fact that $e^{iH(\lambda)t}$ commutes with $e^{-\beta H(\lambda)}$. Similarly, we have

$$\begin{aligned}
\frac{\partial^2}{\partial\lambda_k\partial\lambda_j}\log Z_\beta(\lambda) &= -\beta\ \mathrm{tr}\left[E_j\cdot\frac{\partial}{\partial\lambda_k}\left(\frac{e^{-\beta H(\lambda)}}{Z_\beta(\lambda)}\right)\right]\\
&= -\beta\ \mathrm{tr}\left[E_j\cdot\frac{1}{Z_\beta(\lambda)}\frac{\partial}{\partial\lambda_k}\left(e^{-\beta H(\lambda)}\right)\right]+\beta\mathrm{tr}\left[E_j\cdot\frac{e^{-\beta H(\lambda)}}{Z_\beta(\lambda)}\right]\cdot\frac{1}{Z_\beta(\lambda)}\frac{\partial}{\partial\lambda_k}Z_\beta(\lambda)\\
&= \frac{\beta^2}{2}\mathrm{tr}\left[E_j\cdot\left\{\rho_\beta(\lambda),\Phi_{H(\lambda)}(E_k)\right\}\right]-\beta^2\ \mathrm{tr}[E_k\rho_\beta(\lambda)]\ \mathrm{tr}[E_j\rho_\beta(\lambda)]\\
&= \frac{\beta^2}{2}\mathrm{tr}\left[\left\{E_j,\Phi_{H(\lambda)}(E_k)\right\}\cdot\rho_\beta(\lambda)\right]-\beta^2\ \mathrm{tr}[E_k\rho_\beta(\lambda)]\ \mathrm{tr}[E_j\rho_\beta(\lambda)].
\end{aligned}$$

One can see from this equation that $\nabla^2\log Z_\beta(H)$ is a symmetric real matrix, [7] and hence its eigenvectors have real entries. Finally, we get

$$\begin{aligned}
v^\top\cdot\left(\nabla^2\log Z_\beta(\lambda)\right)\cdot v &= \sum_{j,k}v_jv_k\frac{\partial^2}{\partial\lambda_k\partial\lambda_j}\log Z_\beta(\lambda) \qquad (4.46)\\
&= \frac{\beta^2}{2}\mathrm{tr}\Big[\left\{W_v,\Phi_{H(\lambda)}(W_v)\right\}\rho_\beta(\lambda)\Big]-\left(\beta\mathrm{tr}\left[W_v\rho_\beta(\lambda)\right]\right)^2.
\end{aligned}$$

□

The statement of Lemma 99 does not make it clear that the Hessian is a variance of a suitable operator, or even is positive. The following lemma shows how to lower bound the Hessian by a variance of a quasi-local operator. The intuition for the proof arises by writing the Hessian in a manner that makes its positivity clear. This in particular, shows that $\log Z_\beta(\mu)$ is a convex function in parameters μ — we later improve this to being strongly convex.

Lemma 100 (A lower bound on $v^\top\cdot\nabla^2\log Z_\beta\cdot v$)**.** *For every $v\in\mathbb{R}^m$ and local operator $W_v=\sum_i v_iE_i$, define another quasi-local operator $\widetilde{W}_v$ such that*

$$\widetilde{W}_v=\int_{-\infty}^{\infty}f_\beta(t)\ e^{-iHt}\ W_v\ e^{iHt}dt, \qquad (4.47)$$

where

$$f_\beta(t)=\frac{2}{\beta\pi}\log\frac{e^{\pi|t|/\beta}+1}{e^{\pi|t|/\beta}-1} \qquad (4.48)$$

is defined such that $f_\beta(t)$ scales as $\frac{4}{\beta\pi}e^{-\pi|t|/\beta}$ for large t. It holds that

$$\begin{aligned}
&\frac{1}{2}\mathrm{tr}\Big[\left\{W_v,\Phi_{H(\lambda)}(W_v)\right\}\rho_\beta(\lambda)\Big]-\left(\mathrm{tr}\left[W_v\rho_\beta(\lambda)\right]\right)^2\\
&\qquad\geq\mathrm{tr}\left[(\widetilde{W}_v)^2\rho_\beta(\lambda)\right]-\left(\mathrm{tr}\left[\widetilde{W}_v\rho_\beta(\lambda)\right]\right)^2 \qquad (4.49)
\end{aligned}$$

[7] The terms $\mathrm{tr}[E_k\rho_\beta(\lambda)]$ and $\mathrm{tr}[E_j\rho_\beta(\lambda)]$ are real, being expectations of Hermitian matrices. Moreover, $\left\{E_j,\Phi_{H(\lambda)}(E_k)\right\}$ is a Hermitian operator, being an anti-commutator of two Hermitian operators. Hence $\mathrm{tr}\left[\left\{E_j,\Phi_{H(\lambda)}(E_k)\right\}\rho_\beta(\lambda)\right]$ is real too.

This via Lemma 99 implies that $v^\top \cdot \left(\nabla^2 \log Z_\beta(\lambda)\right) \cdot v \geq \beta^2 \operatorname{Var}[\widetilde{W}_v]$

Remark 101. *For the rest of the paper, we are going to fix an arbitrary* $v \in \mathbb{R}^m$*, in order to avoid subscripting* $W, \widetilde{W}$ *by* v*.*

Proof of Lemma 100. Let us start by proving a simpler version of Eq. (4.49), where we only show

$$\frac{1}{2}\mathrm{tr}\Big[\big\{W, \Phi_{H(\lambda)}(W)\big\}\rho_\beta(\lambda)\Big] - \big(\mathrm{tr}\left[W\rho_\beta(\lambda)\right]\big)^2 \geq 0. \tag{4.50}$$

Since v is an arbitrary vector, this shows that, as expected, $\nabla^2 \log Z_\beta(\lambda)$ is a positive semidefinite operator.

Consider the spectral decomposition of the Gibbs state $\rho_\beta(\lambda)$: $\rho_\beta(\lambda) = \sum_j r_j(\lambda)|j\rangle\langle j|$. Then observe that

$$\begin{aligned}
&\frac{1}{2}\mathrm{tr}\Big[\big\{W, \Phi_{H(\lambda)}(W)\big\}\rho_\beta(\lambda)\Big] - \big(\mathrm{tr}\left[W\rho_\beta(\lambda)\right]\big)^2 \qquad (4.51)\\
&= \frac{1}{2}\sum_j r_j(\lambda)\langle j|\{W, \Phi^H(W)\}|j\rangle - \left(\sum_j r_j(\lambda)W_{j,j}\right)^2\\
&= \frac{1}{2}\sum_{j,k} r_j(\lambda)\left(W_{j,k}\langle k|\Phi^H(W)|j\rangle + \langle j|\Phi^H(W)|k\rangle W_{k,j}\right) - \left(\sum_j r_j(\lambda)W_{j,j}\right)^2\\
&\overset{(1)}{=} \frac{1}{2}\sum_{j,k} r_j(\lambda)\left(W_{j,k}W_{k,j}\tilde{f}_\beta(\mathcal{E}_k - \mathcal{E}_j) + W_{j,k}W_{k,j}\tilde{f}_\beta(\mathcal{E}_j - \mathcal{E}_k)\right) - \left(\sum_j r_j(\lambda)W_{j,j}\right)^2\\
&\overset{(2)}{=} \sum_{j,k} r_j(\lambda)|W_{j,k}|^2\tilde{f}_\beta(|\mathcal{E}_j - \mathcal{E}_k|) - \left(\sum_j r_j(\lambda)W_{j,j}\right)^2. \qquad (4.52)
\end{aligned}$$

In equality (1), we use Definition 83 and in equality (2) we use the facts that W is Hermitian and resp$\tilde{f}_\beta(\omega) = \tilde{f}_\beta(-\omega)$. Since $\tilde{f}_\beta(0) = 1$ and $\tilde{f}_\beta(\omega) > 0$ for all ω, it is now evident from last equation that

$$\mathrm{tr}\left(\frac{1}{2}\{W, \Phi^H(W)\}\rho_\beta\right) - \mathrm{tr}\left(W\rho_\beta\right)^2 \geq \sum_j r_j(\lambda)|W_{j,j}|^2 - \left(\sum_j r_j(\lambda)W_{j,j}\right)^2 \geq 0.$$

We can improve this bound by using the operator $\widetilde{W}$ in (4.47). The function $f_\beta(t)$ in (4.47) is chosen carefully such that its Fourier transform satisfies $\tilde{f}_\beta(\omega) = \tilde{f}_\beta(|\omega|)$. Then, we have that

$$\widetilde{W} = \int_{-\infty}^{\infty} f_\beta(t)\; e^{-iHt}\; W\; e^{iHt} dt = \sum_{j,k} |j\rangle\langle k|\; W_{j,k}\; \tilde{f}_\beta(|\mathcal{E}_j - \mathcal{E}_k|). \tag{4.53}$$

Similar to (4.52) we get

$$\begin{aligned}\operatorname{tr}(\widetilde{W}^2\rho_\beta(\lambda)) - [\operatorname{tr}(\widetilde{W}\rho_\beta(\lambda))]^2 &= \sum_{j,k} r_j(\lambda)|W_{j,k}|^2 \tilde{f}_\beta(|\mathcal{E}_j - \mathcal{E}_k|)^2 - \left(\sum_j r_j(\lambda) W_{j,j}\right)^2 \\ &\leq \sum_{j,k} r_j(\lambda)|W_{j,k}|^2 \tilde{f}_\beta(|\mathcal{E}_j - \mathcal{E}_k|) - \left(\sum_j r_j(\lambda) W_{j,j}\right)^2 \\ &= \operatorname{tr}\left(\frac{\{W, \Phi^H(W)\}}{2}\rho_\beta(\lambda)\right) - [\operatorname{tr}(W\rho_\beta(\lambda))]^2, \end{aligned} \tag{4.54}$$

where the inequality is derived from $\tilde{f}_\beta(x)^2 \leq \tilde{f}_\beta(x)$ for arbitrary $-\infty < x < \infty$.

□

4.9.2 Warm-up before step 2: Variance at infinite temperature

In the previous section, we showed how to give a lower bound on the Hessian of the logarithm of the partition function. To be precise, for a vector $\lambda = (\lambda_1, \ldots, \lambda_m) \in \mathbb{R}^m$ with $\|\lambda\| \leq 1$, Hamiltonian $H(\lambda) = \sum_i \lambda_i E_i$ and $\rho_\beta(\lambda) = \frac{1}{Z_\beta(\lambda)} e^{-\beta H(\lambda)}$ (where $Z_\beta(\lambda) = \operatorname{tr}(e^{-\beta H(\lambda)})$), we showed in Lemma 100 how to carefully choose a local operator $\widetilde{W}$ such that for every v, we have

$$v^\top \cdot \left(\nabla^2 \log Z_\beta(\lambda)\right) \cdot v \geq \beta^2 \operatorname{Var}[\widetilde{W}]. \tag{4.55}$$

In the next few sections, we further prove that the variance of $\widetilde{W}$ with respect to $\rho_\beta(\lambda)$ can be bounded from below.

Before looking at how this can be achieved for the highly non-trivial case of finite temperature, we will look at a simpler case of *infinite temperature limit*. As we see later, certain elements of the proof strategy in this case extends to the general case too.

Consider the infinite temperature Gibbs state (i.e., the maximally mixed state) $\eta = \frac{\mathbb{1}_\Lambda}{\mathcal{D}_\Lambda}$. Assuming that the locality of W, namely W is κ-local with $\kappa = \mathcal{O}(1)$, the following theorem holds.

Theorem 102. *For $\widetilde{W}$ as defined in Lemma 100, we have*

$$\operatorname{tr}[(\widetilde{W})^2\eta] - \operatorname{tr}[\widetilde{W}\eta]^2 \geq \frac{\Theta(1)}{(\beta \log(m) + 1)^2} \sum_{i=1}^m v_i^2. \tag{4.56}$$

The intuition behind the theorem is as follows. In the above statement, if $\widetilde{W}$ is replaced by W, then the lower bound is immediate (see Eq. (4.57) below). Similarly, if H and W were commuting, then $\widetilde{W}$ would be the same as W and the statement would follow. In order to show (4.56) for $\widetilde{W}$ in general, we expand it in the energy basis of the Hamiltonian and use the locality of W to bound the contribution of cross terms (using Lemma 85). This accounts for the contributions arising due to non-commutativity of W and H.

Proof of Theorem 102. Recall from Lemma 100 that $W = \sum_i v_i E_i$. We first note that $\operatorname{tr}[\widetilde{W}\eta] =$

$\mathrm{tr}[W\eta] = 0$. From the definition, we have $\mathrm{tr}[(\widetilde{W})^2\eta] = \frac{1}{\mathcal{D}_\Lambda}\|(\widetilde{W})^2\|_F^2$. To begin with, we observe

$$\|W\|_F^2 = \mathcal{D}_\Lambda \sum_{i=1}^m v_i^2, \tag{4.57}$$

which holds since the basis E_i satisfies $\|E_i\|_F^2 = \mathcal{D}_\Lambda$ and $\mathrm{tr}[E_iE_j] = 0$ if $i \neq j$. Define P_s^H as the projection onto the energy range $(s, s+1]$ of H.

$$P_s^H := \sum_{j:\mathcal{E}_j \in (s,s+1]} |j\rangle\langle j|. \tag{4.58}$$

Using the identity $\sum_s P_s^H = \mathbb{1}_\Lambda$ and the definition of $\widetilde{W}$, let us expand

$$\begin{aligned}
\|W\|_F^2 &= \sum_{s,s'=-\infty}^{\infty} \|P_{s'}^H W P_s^H\|_F^2, \\
\|\widetilde{W}\|_F^2 &= \sum_{s,s'=-\infty}^{\infty} \left\| \int_{-\infty}^{\infty} dt P_{s'}^H f_\beta(t) e^{-iHt} W e^{iHt} P_s^H \right\|_F^2 \qquad (4.59) \\
&= \sum_{s,s'=-\infty}^{\infty} \left\| \sum_{\substack{j:\mathcal{E}_j \in (s,s+1] \\ k:\mathcal{E}_k \in (s,s+1]}} W_{j,k} \tilde{f}_\beta(|\mathcal{E}_j - \mathcal{E}_k|) P_{s'}^H |j\rangle\langle k| P_s^H \right\|_F^2 . \qquad (4.60)
\end{aligned}$$

By using the inequality

$$\tilde{f}_\beta(\omega) = \frac{\tanh(\beta\omega/2)}{\beta\omega/2} \geq \frac{1}{\frac{\beta}{2}|\omega| + 1},$$

we have

$$\begin{aligned}
&\left\| \sum_{\substack{j:\mathcal{E}_j \in (s,s+1] \\ k:\mathcal{E}_k \in (s,s+1]}} W_{j,k} \tilde{f}_\beta(|\mathcal{E}_j - \mathcal{E}_k|) P_{s'}^H |j\rangle\langle k| P_s^H \right\|_F^2 \\
&= \sum_{j:\mathcal{E}_j \in (s,s+1]} \sum_{k:\mathcal{E}_j \in (s',s'+1]} [\tilde{f}_\beta(|\mathcal{E}_j - \mathcal{E}_k|)]^2 |W_{j,k}|^2 \\
&\geq \frac{1}{\left(\frac{\beta}{2}(|s-s'|+1)+1\right)^2} \sum_{j:\mathcal{E}_j \in (s,s+1]} \sum_{k:\mathcal{E}_j \in (s',s'+1]} |W_{j,k}|^2 = \frac{\|P_{s'}^H W P_s^H\|_F^2}{\left(\frac{\beta}{2}(|s-s'|+1)+1\right)^2}. \qquad (4.61)
\end{aligned}$$

Plugging this lower bound in Eq. (4.60) gives the following lower bound for $\|\widetilde{W}\|_F^2$:

$$\|\widetilde{W}\|_F^2 \geq \sum_{s,s'=-\infty}^{\infty} \frac{\|P_{s'}^H W P_s^H\|_F^2}{\left(\frac{\beta}{2}(|s-s'|+1)+1\right)^2} = \sum_{s_0=-\infty}^{\infty} \sum_{s_1 : \frac{s_0+s_1}{2} \in \mathbb{Z}} \frac{\|P_{(s_0+s_1)/2}^H W P_{(s_0-s_1)/2}^H\|_F^2}{\left(\frac{\beta}{2}(|s_1|+1)+1\right)^2}, \tag{4.62}$$

where we have introduced $s_0 = s + s', s_1 = s - s'$. Let us consider the last expression for a fixed s_0, introducing a cut-off parameter $\bar{s}$ which we fix eventually:

$$\begin{aligned}
&\sum_{s_1 : \frac{s_0+s_1}{2} \in \mathbb{Z}} \frac{\|P^H_{(s_0+s_1)/2} W P^H_{(s_0-s_1)/2}\|_F^2}{[\frac{\beta}{2}(|s_1|+1)+1]^2} \\
\geq &\frac{1}{\left(\frac{\beta}{2}(\bar{s}+1)+1\right)^2} \left(\sum_{s_1 : \frac{s_0+s_1}{2} \in \mathbb{Z}, |s_1| \leq \bar{s}} \|P^H_{(s_0+2s_1')/2} W P^H_{(s_0-2s_1')/2}\|_F^2 \right) \\
= &\frac{1}{\left(\frac{\beta}{2}(\bar{s}+1)+1\right)^2} \left(\sum_{s_1 : \frac{s_0+s_1}{2} \in \mathbb{Z}} \|P^H_{(s_0+s_1)/2} W P^H_{(s_0-s_1)/2}\|_F^2 - \sum_{s_1 : \frac{s_0+s_1}{2} \in \mathbb{Z}, |s_1| > \bar{s}} \|P^H_{(s_0+s_1)/2} W P^H_{(s_0-s_1)/2}\|_F^2 \right).
\end{aligned} \tag{4.63}$$

By combining the inequalities (4.62) and (4.63), we obtain

$$\|\widetilde{W}\|_F^2 \geq \frac{\|W\|_F^2}{\left(\frac{\beta}{2}(\bar{s}+1)+1\right)^2} - \frac{1}{\left(\frac{\beta}{2}(\bar{s}+1)+1\right)^2} \sum_{s_0=-\infty}^{\infty} \sum_{s_1 : \frac{s_0+s_1}{2} \in \mathbb{Z}, |s_1| > \bar{s}} \|P^H_{(s_0+s_1)/2} W P^H_{(s_0-s_1)/2}\|_F^2. \tag{4.64}$$

Now, we will estimate the second term in (4.64). Since the subspaces $P^H_{(s_0+s_1)/2}$ and $P^H_{(s_0-s_1)/2}$ are sufficiently far apart in energy, we can use the exponential concentration on the spectrum [AKL16] (as stated in Lemma 85) to obtain the following: for $W = \sum_i v_i E_i$, we have

$$\begin{aligned}
\|P^H_{(s_0+s_1)/2} W P^H_{(s_0-s_1)/2}\| &\leq \sum_{i=1}^m v_i \|P^H_{(s_0+s_1)/2} E_i P^H_{(s_0-s_1)/2}\| \\
&\leq C e^{-\lambda(|s_1|-1-\kappa)} \sum_{i=1}^m |v_i| \leq C m e^{-\lambda(|s_1|-1-\kappa)} \max_i |v_i|.
\end{aligned} \tag{4.65}$$

where we use the condition that E_i are tensor product of Pauli operators with weight at most κ, and the parameters C and λ are $\mathcal{O}(1)$ constants (see Lemma 85 for their explicit forms). Then, the

second term in (4.64) can be upper-bounded by

$$\begin{aligned}
&\sum_{s_0=-\infty}^{\infty} \sum_{s_1:\frac{s_0+s_1}{2}\in\mathbb{Z},|s_1|>\bar{s}} \|P^H_{(s_0+s_1)/2} W P^H_{(s_0-s_1)/2}\|_F^2 \\
&\overset{(1)}{\leq} \sum_{s_0=-\infty}^{\infty} \sum_{s_1:\frac{s_0+s_1}{2}\in\mathbb{Z},|s_1|>\bar{s}} \|P^H_{(s_0+s_1)/2} W P^H_{(s_0-s_1)/2}\|^2 \cdot \|P^H_{(s_0+s_1)/2}\|_F^2 \\
&\overset{(2)}{\leq} \sum_{|s_1|>\bar{s}} \sum_{s_0:\frac{s_0+s_1}{2}\in\mathbb{Z}} \|P^H_{(s_0+s_1)/2}\|_F^2 \cdot C^2 m^2 e^{-2\lambda(|s_1|-1-\kappa)} \max_i v_i^2 \\
&= \mathcal{D}_\Lambda C^2 m^2 e^{2\lambda(1+\kappa)} \max_i v_i^2 \sum_{|s_1|\geq\bar{s}+1} e^{-2\lambda|s_1|} \\
&\overset{(3)}{\leq} \mathcal{D}_\Lambda \max_i v_i^2 \frac{C^2 m^2 e^{2\lambda(\kappa+1)}}{2\lambda} e^{-2\lambda\bar{s}} \overset{(4)}{\leq} \mathcal{D}_\Lambda \frac{C^2 m^2 e^{2\lambda(\kappa+1)}}{2\lambda} e^{-2\lambda\bar{s}} \left(\sum_i v_i^2\right),
\end{aligned} \tag{4.66}$$

where inequality (1) follows from Eq. (4.15), (2) follows from Eq. (4.65), (3) follows from Fact 78 and (4) follows from $\max_i v_i^2 \leq \sum_i v_i^2$. Therefore, by applying Eq. (4.57) and (4.66) to (4.64), we arrive at the lower bound as

$$\|\widetilde{W}\|_F^2 \geq \frac{\mathcal{D}_\Lambda}{[\beta(\bar{s}+1)/2+1]^2} \left(\sum_{i=1}^m v_i^2\right) \left(1 - \frac{C^2 m^2 e^{2\lambda(\kappa+1)}}{2\lambda} e^{-2\lambda\bar{s}}\right). \tag{4.67}$$

Since $\lambda, C, \kappa = \mathcal{O}(1)$, by choosing $\bar{s} = \mathcal{O}(\log(m))$, we obtain the main inequality (4.56). This completes the proof. □ □

4.9.3 Step 2: From global to local variance

Having seen how to obtain a lower bound for $\widetilde{W}$ at the infinite temperature as in Theorem 102, we move on to the more challenging task of proving a variance lower bound at *finite* temperatures.

A main challenge in bounding the variance of the operator $\widetilde{W}$ is that even though $\widetilde{W}$ is a sum of quasi-local operators, it acts globally on all the sites and hence, its spectral properties might scale badly with the system size. To address this issue, we will follow the strategy introduced in Section 4.6.5 and reviewed in Section 4.3.4 to reduce the problem to the study of a more local operator. For some $i \in \Lambda$, recall the definition of $\widetilde{W}_{(i)}$ from the subsection 4.6.5:

$$\widetilde{W}_{(i)} := \widetilde{W} - \mathrm{tr}_i[\widetilde{W}] \otimes \frac{\mathbb{1}_i}{d} \tag{4.68}$$

The operator $\widetilde{W}_{(i)}$ includes essentially the terms in $\widetilde{W}$ that are supported on the ith site. We now show how the variance of the global operator $\widetilde{W}$ can be related to the variance of $\widetilde{W}_{(i)}$.

Remark 103. *Instead of working with $\widetilde{W}$, it is more convenient to define a quasi-local operator A that is proportional to $\widetilde{W} - \mathrm{tr}[\rho_\beta \widetilde{W}]\mathbb{1}$. This allows us to express $\mathrm{Var}[\widetilde{W}]$ in terms of $\mathrm{tr}[A^2\rho_\beta]$. The exact choice of this operator will be made later, but all the results in this and upcoming sections hold generally for a (τ, a_1, a_2, ζ)-quasi-local operator A (see Eq. (4.18)) where $a_2 = \mathcal{O}(1/\beta), a_1 = \mathcal{O}(1)$*

are constants, $\zeta = 1$, $\tau \leq 1$, *and* $\mathrm{tr}[A\rho_\beta] = 0$. *We note that the quasi-locality of the opeartor* $\widetilde{W}$ *in terms of the definition 80 is given in Section 4.10.4.*

We will interchangeably use the Frobenius norm to write the variance of A as

$$\mathrm{Var}[A] = \mathrm{tr}(A^2\rho_\beta) = \|A\sqrt{\rho_\beta}\|_F.$$

Using Haar random unitaries, we obtain the integral representation of $A_{(i)} = A - \mathrm{tr}_i[A] \otimes \frac{\mathbb{1}_i}{d}$ as

$$A_{(i)} = A - \frac{1}{d}[\mathrm{tr}_i(A)] \otimes \mathbb{1}_i = A - \int d\mu(U_i) U_i^\dagger A U_i, \tag{4.69}$$

where $d\mu(U_i)$ is the Haar measure for unitary operator U_i which acts on the ith site. Using triangle inequality, we have

$$\begin{aligned} \|A_{(i)}\sqrt{\rho_\beta}\|_F &\leq \|A\sqrt{\rho_\beta}\|_F + \int d\mu(U_i) \|U_i^\dagger A U_i \sqrt{\rho_\beta}\|_F \\ &= \|A\sqrt{\rho_\beta}\|_F + \int d\mu(U_i) \|A U_i \sqrt{\rho_\beta}\|_F \end{aligned} \tag{4.70}$$

This implies

$$\|A_{(i)}\sqrt{\rho_\beta}\|_F^2 \leq \Big(\|A\sqrt{\rho_\beta}\|_F + \int d\mu(U_i) \|A U_i \sqrt{\rho_\beta}\|_F\Big)^2 \tag{4.71}$$

$$\leq 2\|A\sqrt{\rho_\beta}\|_F^2 + 2\Big(\int d\mu(U_i) \|A U_i \sqrt{\rho_\beta}\|_F\Big)^2. \tag{4.72}$$

In fact, we need a slightly generalized version of this bound in the forthcoming steps of the proof. This modified version allows us to include the effect of a *further* local unitary U_{X_i} that is applied to the Gibbs state. Here, $X_i := B(R, i)$ is a ball around the site i with a constant radius R that we specify later in Section 4.9.5. We skip the proof since it is the same as the above argument.

Theorem 104. *Let* A *be a quasi-local operator as in Remark 103 and* $A_{(i)} = A - \mathrm{tr}_i[A] \otimes \frac{\mathbb{1}_i}{d}$ *defined as in Section 4.6.5 for some* $i \in \Lambda$. *Consider local unitaries* U_i *and* U_{X_i} *acting on site* i *and* X_i *respectively, where* X_i *is a ball of constant size at site* i . *The following generalization of (4.72) holds:*

$$\|U_{X_i}^\dagger A_{(i)} U_{X_i} \sqrt{\rho_\beta}\|_F^2 \leq 2\|A U_{X_i} \sqrt{\rho_\beta}\|_F^2 + 2\Big(\int d\mu(U_i) \|A U_i U_{X_i} \sqrt{\rho_\beta}\|_F\Big)^2. \tag{4.73}$$

We now move to another ingredient we need for future steps. Although the operator $A_{(i)}$ is mostly localized around the site i, it is still obtained from a quasi-local operator A and hence it is quasi-local itself. Next claim will approximate $A_{(i)}$ by a *local* operator.

Claim 105. *Consider* A *which is a* (τ, a_1, a_2, ζ)*-quasi-local operator (see Eq. (4.18) and Remark 103) where* $a_2 = \mathcal{O}(1/\beta), a_1 = \mathcal{O}(1)$ *are constants,* $\zeta = 1$, *and* $\tau \leq 1$. *For any* $i \in \Lambda$, *there exists an operator* A_{X_i} *supported entirely on* X_i, *such that*

$$\|A_{(i)} - A_{X_i}\| \leq 2a_1 \cdot \left(1 + a_2^{-\frac{1}{\tau}}\right) \cdot \left(\frac{4}{\tau^2}\right)^{\frac{1}{\tau}} \cdot e^{-\frac{a_2}{2}(R)^\tau}.$$

Proof. Using Eq. (4.18) and the fact that local operators not containing i in their support are removed, we can write

$$A_{(i)} = \sum_{\substack{k,Z\subseteq\Lambda:\\ |Z|\leq k, Z\ni i}} g_k \left(a_Z - \frac{1}{d}[\mathrm{tr}_i(a_Z)] \otimes \mathbb{1}_i \right),$$

Define

$$A_{X_i} = \sum_{\substack{k,Z\subseteq X_i:\\ |Z|\leq k, Z\ni i}} g_k \left(a_Z - \frac{1}{d}[\mathrm{tr}_i(a_Z)] \otimes \mathbb{1}_i \right).$$

be the desired approximations of $A_{(i)}$ by removing all operators that are not contained within X_i. Observe that

$$\begin{aligned}
\|A_{(i)} - A_{X_i}\| &\leq 2 \sum_{\substack{k,Z\subseteq\Lambda:\\ Z\not\subset X_i, |Z|\leq k, Z\ni i}} g_k \|a_Z\| \overset{(1)}{\leq} 2 \sum_{\substack{k,Z\subseteq\Lambda:\\ \mathrm{diam}(Z)\geq R, |Z|\leq k, Z\ni i}} g_k \|a_Z\| \\
&\overset{(2)}{\leq} 2 \sum_{k\geq R} g_k \left(\sum_{Z: Z\ni i} \|a_Z\| \right) \overset{(3)}{\leq} 2\zeta \sum_{k\geq R} g_k \\
&\leq 2\zeta a_1 \sum_{k\geq R} e^{-a_2 k^\tau} \overset{(4)}{\leq} 2\zeta a_1 e^{-\frac{a_2}{2}R^\tau} \left(1 + \tau^{-1}[2/(a_2\tau)]^{1/\tau} \right) \\
&\leq 2\zeta a_1 \cdot \left(1 + a_2^{-\frac{1}{\tau}} \right) \cdot \left(\frac{2}{\tau} \right)^{\frac{2}{\tau}} \cdot e^{-\frac{a_2}{2}(R)^\tau}.
\end{aligned} \tag{4.74}$$

For inequality (1), note that since $Z \not\subset X_i$ and $Z \ni i$, the diameter of Z (recall that Z is a ball) must be larger than the radius of X_i, which is R. Inequality (2) holds since $k \geq |Z| \geq \mathrm{diam}(Z)$, inequality (3) uses Definition 80 and inequality (4) uses Fact 78. Since $\zeta = 1$ for the given A, this completes the proof. □

4.9.4 Step 3: Bounding the effect of local unitaries

Recall that we reduced the problem of bounding the variance of A (which as explained in Remark 103) is closely related to our original operator $\widetilde{W}$) to that of the operators $\{A_{(i)}\}_{i\in\Lambda}$. These operators are essentially supported on a small number of sites. However, in this process, we introduced other local unitaries that are applied on the state. In order to handle the action of these unitaries, we will use two claims which show that local unitaries do not make much difference in the relative behavior of spectra of A and H. To elaborate, consider any local operator U_X acting on constant number of sites X on the state ρ_β. It is expected that the quantum state $U_X^\dagger \rho_\beta U_X$ has "similar" spectral properties as ρ_β. So if the eigen-spectrum of the operator A is strongly concentrated for ρ_β, one would expect this behavior to hold even for $U_X^\dagger \rho_\beta U_X$. We make this intuition rigorous in this section.

We begin this by introducing some key quantities that are used repeatedly in the proof. For notational simplicity, let

$$H' := H - \frac{1}{\beta} \log Z_\beta, \tag{4.75}$$

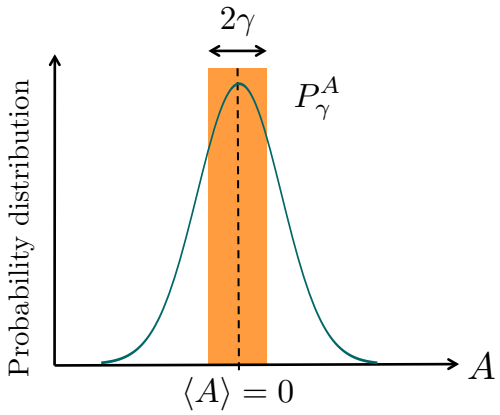


Figure 4-2: Plot of the probability distribution $\mathrm{tr}[\Pi_\omega \rho_\beta]$, where Π_ω is the projector onto the eigenvectors of A with eigenvalue ω. It is assumed that $\mathrm{tr}[A\rho_\beta] = 0$. For a γ to be chosen in the proof, P_γ^A is the projector onto the subspace of eigenvectors of A with eigenvalue between $[-\gamma, \gamma]$. A lower bound on the variance of A follows if we can show that for a constant γ, the probability mass in the colored range is small (see Equation 4.78).

which allows us to write $\rho_\beta = e^{-\beta H'}$. As before, we will interchangeably use the Frobenius norm to write

$$\langle A^2 \rangle = \mathrm{tr}(A^2 \rho_\beta) = \|A\sqrt{\rho_\beta}\|_F.$$

We now define the projection operator P_γ^A as follows (see Figure 4-2):

$$P_\gamma^A := \sum_{\omega \in [-\gamma, \gamma]} \Pi_\omega, \tag{4.76}$$

where Π_ω is the projector onto the eigenspace of A with eigenvalue ω. We then define δ_γ by

$$\delta_\gamma := 1 - \|P_\gamma^A \sqrt{\rho_\beta}\|_F^2. \tag{4.77}$$

Using δ_γ, observe that we can lower bound $\langle A^2 \rangle$ as

$$\langle A^2 \rangle = \sum_\omega \omega^2 \langle \omega | \rho_\beta | \omega \rangle \geq \sum_{|\omega| \geq \gamma} \omega^2 \langle \omega | \rho_\beta | \omega \rangle \geq \gamma^2 \sum_{|\omega| \geq \gamma} \langle \omega | \rho_\beta | \omega \rangle \geq \gamma^2 \delta_\gamma. \tag{4.78}$$

Let $Q_\gamma^A = \mathbb{1} - P_\gamma^A$, then observe that from Eq. (4.77) that

$$\|P_\gamma^A \sqrt{\rho_\beta} - \sqrt{\rho_\beta}\|_F^2 = \|Q_\gamma^A \sqrt{\rho_\beta}\|_F^2 = \delta_\gamma. \tag{4.79}$$

Claim 106. *Let c_1, c_2, λ be universal constants. Let $X \subseteq \Lambda$. For every unitary U_X supported on X, we have*

$$\|Q_\gamma^A U_X \sqrt{\rho_\beta}\|_F^2 \leq \exp\left(\lambda |X|\right) \delta_\gamma^{\frac{c_2}{c_2 + \beta}}. \tag{4.80}$$

Let us see a simple application of the claim. It allows us to control the variance of A even after

local operations are applied to it. More precisely,

$$\begin{aligned}\|AU_X\sqrt{\rho_\beta}\|_F^2 &= \|AP_\gamma^A U_X\sqrt{\rho_\beta}\|_F^2 + \|A(\mathbb{1}-P_\gamma^A)U_X\sqrt{\rho_\beta}\|_F^2 \\ &\leq \gamma^2 + \|A\|^2\cdot\|(\mathbb{1}-P_\gamma^A)U_X\sqrt{\rho_\beta}\|_F^2.\end{aligned} \tag{4.81}$$

By Claim 106, the expression on the second line is upper bounded by $\gamma^2+\|A\|^2 e^{\Theta(1)|X|}\delta_\gamma^{\Theta(1)/\beta}$. This upper bound on $\|AU_X\sqrt{\rho_\beta}\|_F$ suffices to provide an inverse-polynomial *lower bound* on the variance of A^2, since we can lower bound δ_γ for an appropriate choice of γ. However we now show how one can polynomially improve upon this upper bound (thereby the lower bound on variance) using the following claim. This claim, along the lines of Claim 106, also shows that local unitaries U_X do not change the desired expectation values.

Claim 107. *Let $X\subseteq\Lambda$. For every unitary U_X supported on X, we have*[8]

$$\left\|AQ_\gamma^A U_X\sqrt{\rho_\beta}\right\|_F^2 \leq \frac{1}{\gamma}\cdot\exp\left(\Theta(1)\cdot|X|\right)\delta_\gamma^{\Theta(1)/\beta} + \Theta(1)\cdot|X|^6\cdot\langle A^2\rangle. \tag{4.82}$$

Proof of both the Claims 106, 107 appear in Section 4.10.2. An immediate corollary of this claim is the following, that improves upon Eq. (4.81).

Corollary 108. *Let X be a subset of Λ of size $|X|=\mathcal{O}(1)$. For every unitary U_X supported on X, we have*

$$\|AU_X\sqrt{\rho_\beta}\|_F^2 \leq \gamma^2 + \frac{e^{\Theta(1)\cdot|X|}\delta_\gamma^{\Theta(1)/\beta}}{\gamma} + \Theta(1)|X|^6\langle A^2\rangle. \tag{4.83}$$

Proof. Similar to Eq. (4.81), we upper bound $\|AU_X\sqrt{\rho_\beta}\|_F^2$ as

$$\|AU_X\sqrt{\rho_\beta}\|_F^2 = \left\|AP_\gamma^A U_X\sqrt{\rho_\beta}\right\|_F^2 + \left\|AQ_\gamma^A U_X\sqrt{\rho_\beta}\right\|_F^2 \leq \gamma^2 + \left\|AQ_\gamma^A U_X\sqrt{\rho_\beta}\right\|_F^2, \tag{4.84}$$

since $Q_\gamma^A = \mathbb{1}-P_\gamma^A$. By combining this with Claim 107, the corollary follows. □

4.9.5 Step 4: Reduction to infinite temperature variance

In Section 4.9.3, we reduced the variance of a global operator A (which is related to $\widetilde{W}$ via Remark 103) to the quasi-local operators $\{A_{(i)}\}_{i\in\Lambda}$. We now argue that there is some i such that it is possible to bound the *finite* temperature variance of $A_{(i)}$ in terms of its variance at *infinite* temperature. This can be done by applying some local rotations given by a unitary U_{X_i} on the state. The intuition here is that if rotations are allowed, then the eigenvectors of A_{X_i} can be rearranged to yield largest possible variance with ρ_β. This turns out to be larger than the variance with η the infinite temperature state. Define the site i_0 as

$$i_0 := \arg\max_i \|A_{(i)}\sqrt{\eta}\|_F. \tag{4.85}$$

[8] Explicit $\mathcal{O}(1)$ constants that appear in this inequality are made clear in the proof.

Claim 109. *Let η be the infinite temperature Gibbs state. There exists a unitary $U_{X_{i_0}}$ supported on X_{i_0} such that*

$$\|U_{X_{i_0}}^\dagger A_{(i_0)} U_{X_{i_0}} \sqrt{\rho_\beta}\|_F \geq \frac{1}{2}\|A_{(i_0)}\sqrt{\eta}\|_F.$$

Proof of Claim 109. We first show the existence of a unitary $U_{X_{i_0}}$ satisfying

$$\|U_{X_{i_0}}^\dagger A_{(i_0)} U_{X_{i_0}} \sqrt{\rho_\beta}\|_F \geq \|A_{(i_0)}\sqrt{\eta}\|_F - 2\|A_{X_{i_0}} - A_{(i_0)}\|. \tag{4.86}$$

Recall that the notation A_{X_i} has been defined in Claim 105. For the proof of the above inequality, we start from the following,

$$\begin{aligned}\|U_{X_{i_0}}^\dagger A_{(i_0)} U_{X_{i_0}} \sqrt{\rho_\beta}\|_F &= \|U_{X_{i_0}}^\dagger \left[(A_{(i_0)} - A_{X_{i_0}}) + A_{X_{i_0}}\right] U_{X_{i_0}} \sqrt{\rho_\beta}\|_F \\ &\geq \|U_{X_{i_0}}^\dagger A_{X_{i_0}} U_{X_{i_0}} \sqrt{\rho_\beta}\|_F - \|A_{(i_0)} - A_{X_{i_0}}\| \end{aligned} \tag{4.87}$$

and lower-bound the norm of $\|U_{X_{i_0}}^\dagger A_{X_{i_0}} U_{X_{i_0}} \sqrt{\rho_\beta}\|_F$. For this, define

$$\rho_{\beta,X} := \mathrm{tr}_{X^{\mathrm{c}}}(\rho_\beta), \tag{4.88}$$

where $\mathrm{tr}_{X^{\mathrm{c}}}$ is the partial trace operation for the Hilbert space on X^{c}. We define the spectral decomposition of $A_{X_{i_0}}$ as

$$A_{X_{i_0}} = \sum_{s=1}^{\mathcal{D}_{X_{i_0}}} \varepsilon_s |\varepsilon_s\rangle\langle\varepsilon_s|, \tag{4.89}$$

where ε_s is ordered as $|\varepsilon_1| \geq |\varepsilon_2| \geq |\varepsilon_3| \geq \cdots$ and $\mathcal{D}_{X_{i_0}}$ is the dimension of the Hilbert space on X_{i_0}. Additionally, define the spectral decomposition of $\rho_{\beta,X_{i_0}}$ as

$$\rho_{\beta,X_{i_0}} = \sum_{s=1}^{\mathcal{D}_{X_{i_0}}} p_s |\mu_s\rangle\langle\mu_s|, \tag{4.90}$$

where p_s is ordered as $p_1 \geq p_2 \geq p_3 \geq \cdots$ and $|\mu_s\rangle$ is the sth eigenstate of $\rho_{\beta,X_{i_0}}$. We now choose the unitary operator $U_{X_{i_0}}$ such that

$$U_{X_{i_0}} |\mu_s\rangle = |\varepsilon_s\rangle \quad \text{for} \quad s = 1, 2, \ldots, \mathcal{D}_{X_{i_0}} \tag{4.91}$$

We then obtain

$$U_{X_{i_0}} \rho_{\beta,X_{i_0}} U_{X_{i_0}}^\dagger = \sum_{s=1}^{\mathcal{D}_{X_{i_0}}} p_s |\varepsilon_s\rangle\langle\varepsilon_s|. \tag{4.92}$$

This implies

$$\begin{aligned}\|U_{X_{i_0}}^\dagger A_{X_{i_0}} U_{X_{i_0}}\sqrt{\rho_\beta}\|_F^2 &= \mathrm{tr}[U_{X_{i_0}}^\dagger A_{X_{i_0}}^2 U_{X_{i_0}}\rho_\beta] \\ &= \mathrm{tr}[A_{X_{i_0}}^2 U_{X_{i_0}}\rho_\beta U_{X_{i_0}}^\dagger] = \mathrm{tr}_{X_{i_0}}[A_{X_{i_0}}^2 U_{X_{i_0}}\rho_{\beta,X_{i_0}} U_{X_{i_0}}^\dagger] \\ &= \sum_{s=1}^{\mathcal{D}_{X_{i_0}}} p_s\varepsilon_s^2 \geq \frac{1}{\mathcal{D}_{X_{i_0}}}\sum_{s=1}^{\mathcal{D}_{X_{i_0}}}\varepsilon_s^2 = \|A_{X_{i_0}}\sqrt{\eta}\|_F^2, \end{aligned} \tag{4.93}$$

where the inequality used the fact that p_s, ε_s are given in descending order. Then, the minimization problem of $\sum_s p_s\varepsilon_s$ for p_{ss} with the constraint $p_1 \geq p_2 \geq p_3 \geq \cdots$ has a solution of $p_1 = p_2 = \cdots = p_{\mathcal{D}_{X_{i_0}}}$. Using the lower bound

$$\|A_{X_{i_0}}\sqrt{\eta}\|_F = \|(A_{X_{i_0}} - A_{(i_0)} + A_{(i_0)})\sqrt{\eta}\|_F \geq \|A_{(i_0)}\sqrt{\eta}\|_F - \|A_{X_{i_0}} - A_{(i_0)}\|, \tag{4.94}$$

we can reduce inequality (4.93) to

$$\|U_{X_{i_0}}^\dagger A_{X_{i_0}} U_{X_{i_0}}\sqrt{\rho_\beta}\|_F \geq \|A_{(i_0)}\sqrt{\eta}\|_F - \|A_{X_{i_0}} - A_{(i_0)}\|. \tag{4.95}$$

By combining the inequalities (4.87) and (4.95), we obtain the inequality (4.86) as follows:

$$\begin{aligned}\|U_{X_{i_0}}^\dagger A_{(i_0)} U_{X_{i_0}}\sqrt{\rho_\beta}\|_F &\geq \|U_{X_{i_0}}^\dagger A_{X_{i_0}} U_{X_{i_0}}\sqrt{\rho_\beta}\|_F - \|A_{(i_0)} - A_{X_{i_0}}\| \\ &\geq \|A_{(i_0)}\sqrt{\eta}\|_F - 2\|A_{(i_0)} - A_{X_{i_0}}\| \ ,\end{aligned}$$

To arrive at the final bound we use Claim 105. Fix

$$R = \left\lceil \left(\frac{2}{a_2}\log\frac{8\cdot 4^{\frac{1}{\tau}}\cdot a_1\cdot\left(1+a_2^{-\frac{1}{\tau}}\right)}{\tau^{\frac{2}{\tau}}\|A_{(i_0)}\sqrt{\eta}\|_F}\right)^{\frac{1}{\tau}} \right\rceil$$

in Claim 105. We get

$$\|A_{(i_0)} - A_{X_{i_0}}\| \leq \frac{1}{4}\|A_{(i_0)}\sqrt{\eta}\|_F \tag{4.96}$$

Substituting $a_2 = \mathcal{O}(1/\beta), a_1 = \mathcal{O}(1), \tau = \mathcal{O}(1)$, we find that we can ensure the condition (4.96) for

$$R = \mathrm{diam}(X_{i_0}) = \left(\beta\log\left(\frac{1}{\|A_{(i_0)}\sqrt{\eta}\|_F}\right)\right)^{\Theta(1)}. \tag{4.97}$$

□

The previous claim demonstrates that by applying local unitaries, the finite temperature variance of $A_{(i_0)}$ can be related to $\|A_{(i_0)}\sqrt{\eta}\|_F$. Here, we use the discussion in Section 4.9.3 and Section 4.9.4 to prove that the rotated local variance $\|U_{X_{i_0}}^\dagger A_{(i_0)} U_{X_{i_0}}\sqrt{\rho_\beta}\|_F$ in Claim 109 is close to the variance of $A_{(i)}$ and hence, can be related to the variance of the global operator A. By establishing this, we

can obtain a lower bound on the variance of A (an in turn $\widetilde{W}$) in terms of the infinite temperature variance $\max_{i\in\Lambda}\|A_{(i)}\sqrt{\eta}\|_F$. This is stated in the following theorem:

Theorem 110. *Let $\beta > 0$, H be a κ-local Hamiltonian on the lattice Λ and $\rho_\beta = \frac{e^{-\beta H}}{\mathrm{tr}(e^{-\beta H})}$. Let A be a (τ, a_1, a_2, ζ)-quasi-local operator (see Eq. (4.18) and Remark 103) where $a_2 = \mathcal{O}(1/\beta), a_1 = \mathcal{O}(1)$ are constants and we assume $\zeta = 1$, $\tau \leq 1$ and $\mathrm{tr}[A\rho_\beta] = 0$. We have*

$$\langle A^2\rangle = \mathrm{tr}(A^2\rho_\beta) \geq \left(\max_{i\in\Lambda}\mathrm{tr}[A_{(i)}^2\eta]\right)^{\beta^{\Theta(1)}}.$$

We remark that the theorem statement above hides several terms that depend on the lattice, such as the lattice dimension, the degree of the graph and the locality of Hamiltonian (which we have fixed to be a constant). Additionally, the assumptions $a_2 = \mathcal{O}(1/\beta), a_1 = \mathcal{O}(1)$ are suitably made for the later application of this bound in Section 4.9.6, but we note that this theorem also holds for other choices of a_1, a_2, with small modifications to the proof.

Proof. Let $U_{X_{i_0}}$ be the unitary as chosen in Claim 109. Using Theorem 104, we obtain the following upper bound for $U_{X_{i_0}}^\dagger A_{(i_0)} U_{X_{i_0}}$

$$\begin{aligned}\|U_{X_{i_0}}^\dagger A_{(i_0)} U_{X_{i_0}}\sqrt{\rho_\beta}\|_F^2 &\leq \left(\|AU_{X_{i_0}}\sqrt{\rho_\beta}\|_F + \int d\mu(U_{i_0})\|AU_{i_0}U_{X_{i_0}}\sqrt{\rho_\beta}\|_F\right)^2\\ &\leq 2\|AU_{X_{i_0}}\sqrt{\rho_\beta}\|_F^2 + 2\int d\mu(U_{i_0})\|AU_{i_0}U_{X_{i_0}}\sqrt{\rho_\beta}\|_F^2.\end{aligned}$$

Now, we can use Corollary 108 to upper bound $\|AU_{X_{i_0}}\sqrt{\rho_\beta}\|_F^2$ and $\|AU_{i_0}U_{X_{i_0}}\sqrt{\rho_\beta}\|_F$ in the right hand side, which yields

$$\|U_{X_{i_0}}^\dagger A_{(i_0)} U_{X_{i_0}}\sqrt{\rho_\beta}\|_F^2 \leq 4\gamma^2 + \frac{e^{\Theta(1)|X_{i_0}|}\delta_\gamma^{\Theta(1)/\beta}}{\gamma} + \Theta(1)|X_{i_0}|^6\langle A^2\rangle. \tag{4.98}$$

Using Claim 109, we have

$$\|U_{X_{i_0}}^\dagger A_{(i_0)} U_{X_{i_0}}\sqrt{\rho_\beta}\|_F \geq \frac{1}{2}\|A_{(i_0)}\sqrt{\eta}\|_F. \tag{4.99}$$

Putting together the upper bound in Eq. (4.98) and the lower bound in Eq. (4.99), we have

$$4\gamma^2 + \frac{e^{\Theta(1)|X_{i_0}|}\delta_\gamma^{\Theta(1)/\beta}}{\gamma} + \Theta(1)|X_{i_0}|^6\langle A^2\rangle \geq \frac{1}{4}\|A_{(i_0)}\sqrt{\eta}\|_F^2. \tag{4.100}$$

By choosing as $\gamma^2 = \|A_{(i_0)}\sqrt{\eta}\|_F^2/32 =: \gamma_0^2$, we obtain

$$\frac{e^{\Theta(1)|X_{i_0}|}\delta_{\gamma_0}^{\Theta(1)/\beta}}{\gamma_0} + \Theta(1)|X_{i_0}|^6\langle A^2\rangle \geq \gamma_0^2. \tag{4.101}$$

This inequality implies that either

$$\delta_{\gamma_0} \geq \left(\gamma_0^3 e^{-\Theta(1)|X_{i_0}|}\right)^{\beta\cdot\Theta(1)}$$

or

$$\langle A^2 \rangle \geq \frac{\Theta(1)\gamma_0^2}{|X_{i_0}|^6}.$$

Combining with Eq. (4.78), we conclude that

$$\langle A^2 \rangle \geq \min\left\{\gamma_0^2 \cdot \left(\gamma_0^3 e^{-\Theta(1)|X_{i_0}|}\right)^{\beta\cdot\Theta(1)}, \frac{\Theta(1)\gamma_0^2}{|X_{i_0}|^6}\right\}.$$

Eq. (4.97) ensures that

$$|X_{i_0}| = \Theta(1)R^D = \beta^{\Theta(1)} \log\left(\frac{1}{\|A_{(i_0)}\sqrt{\eta}\|_F}\right)^{\Theta(1)},$$

where we have used the assumption that lattice dimension D is $\Theta(1)$. Plugging in this expression for $|X_{i_0}|$ with the choice of γ_0, we find

$$\begin{aligned}
\mathrm{tr}(A^2\rho_\beta) = \langle A^2 \rangle &\geq \min\left\{\|A_{(i_0)}\sqrt{\eta}\|_F^{\beta\cdot\Theta(1)} \cdot e^{-\beta\Theta(1)|X_{i_0}|}, \frac{\Theta(1)\|A_{(i_0)}\sqrt{\eta}\|_F^{\Theta(1)}}{|X_{i_0}|^6}\right\} \\
&\geq \min\left\{\|A_{(i_0)}\sqrt{\eta}\|_F^{\beta^{\Theta(1)}}, \frac{\Theta(1)\|A_{(i_0)}\sqrt{\eta}\|_F^{\Theta(1)}}{\beta^{\Theta(1)}\log(\frac{1}{\|A_{(i_0)}\sqrt{\eta}\|_F})^{\Theta(1)}}\right\} \\
&\geq \min\left\{\|A_{(i_0)}\sqrt{\eta}\|_F^{\beta^{\Theta(1)}}, \frac{\Theta(1)\|A_{(i_0)}\sqrt{\eta}\|_F^{\Theta(1)}}{\beta^{\Theta(1)}}\right\} \\
&\geq \|A_{(i_0)}\sqrt{\eta}\|_F^{\beta^{\Theta(1)}}.
\end{aligned}$$

Since we chose i_0 in Eq. (4.85) such that $\|A_{(i_0)}\sqrt{\eta}\|_F = \max_i \|A_{(i)}\sqrt{\eta}\|_F$, this proves the theorem. □

4.9.6 Final step: Putting things together

We are now ready to apply the results of the past steps to prove Theorem 98. The bound in Theorem 110 in the previous section is stated for a general quasi-local observable A. As discussed in Remark 103, a special choice of A is when it is proportional to $\widetilde{W} - \mathrm{tr}[\rho_\beta \widetilde{W}]\mathbb{1}$ where recall that for an arbitrary $v \in \mathbb{R}^m$, $W = \sum_i v_i E_i$ and $\widetilde{W}$ are the operators defined in Lemma 100. In Section 4.10.4, we show that $\widetilde{W}$ is a $\big(1/D, \mathcal{O}(1), \mathcal{O}(1/\beta), c_*\beta^{2D+1}\left(\max_{i\in\Lambda}|v_i|\right)\big)$-quasi-local operator, for $c_* = \mathcal{O}(1)$. Thus, the following operator

$$A^* = \frac{\beta^{-2D-1}}{c_* \max_{i\in\Lambda}|v_i|}(\widetilde{W} - \mathrm{tr}[\rho_\beta \widetilde{W}]\mathbb{1}),$$

is $(\mathcal{O}(1), \mathcal{O}(1), \mathcal{O}(1/\beta), 1)$-quasi-local and satisfies $\mathrm{tr}[A^*\rho_\beta] = 0$. We now apply Theorem 110 to the operator A^* to prove Theorem 98. We need to estimate $\max_i \mathrm{tr}[A^{*2}_{(i)}\eta]$. Consider the following equality obtained from the definition of A^*:

$$\max_{i\in\Lambda} \mathrm{tr}[(A^*_{(i)})^2\eta] = \frac{\beta^{-4D-2}}{c_*^2\left(\max_{i\in\Lambda}|v_i|\right)^2}\left(\max_{i\in\Lambda}\mathrm{tr}[(\widetilde{W}_{(i)})^2\eta]\right).$$

The following lemma is shown in Section 4.10.5.

Lemma 111. *It holds that*

$$\max_{i\in\Lambda}(\mathrm{tr}[(\widetilde{W}_{(i)})^2\eta]) = \frac{\Theta(1)}{(\beta\log(\beta)+1)^{2D+2}}\left(\max_{i\in\Lambda} v_i^2\right).$$

This implies

$$\max_{i\in\Lambda}\mathrm{tr}[(A^*_{(i)})^2\eta] = \frac{\Theta(1)}{\beta^{4D+2}\,(\beta\log(\beta)+1)^{2D+2}\,(\max_{i\in\Lambda}|v_i|)^2}\left(\max_{i\in\Lambda} v_i^2\right) = \frac{1}{\beta^{\Theta(1)}}.$$

Using this lower bound in Theorem 110, we find

$$\begin{aligned}
\langle(\widetilde{W})^2\rangle - \left(\langle\widetilde{W}\rangle\right)^2 &= c_*^2\beta^{4D+2}\left(\max_{i\in\Lambda}|v_i|\right)^2\left(\langle(A^*)^2\rangle - (\langle A^*\rangle)^2\right)\\
&= \beta^{\Theta(1)}\left(\max_{i\in\Lambda}|v_i|\right)^2\cdot\left(\max_i \mathrm{tr}[(A^*_{(i)})^2\eta]\right)^{\beta^{\Theta(1)}}\\
&= \beta^{\Theta(1)}\cdot\left(\frac{1}{\beta^{\Theta(1)}}\right)^{\beta^{\Theta(1)}}\cdot\left(\max_{i\in\Lambda}|v_i|\right)^2\\
&\overset{(1)}{=} \beta^{\Theta(1)}\cdot e^{-\beta^{\Theta(1)}}\left(\max_{i\in\Lambda}|v_i|\right)^2 \geq \beta^{\Theta(1)}\cdot\frac{e^{-\beta^{\Theta(1)}}}{m}\left(\sum_i v_i^2\right),
\end{aligned}$$

where we used $\beta^{-\Theta(1)} \geq e^{-\Theta(\beta)}$ in (1). Putting together the bound above with Eq. (4.55), we find that for every $v\in\mathbb{R}^m$,

$$v^\top\cdot\left(\nabla^2\log Z_\beta(\lambda)\right)\cdot v \geq \beta^2\,\mathrm{Var}[\widetilde{W}] \geq \beta^{\Theta(1)}\cdot\frac{e^{-\beta^{\Theta(1)}}}{m}\sum_{i=1}^m v_i^2.$$

This establishes Theorem 98.

4.10 Deferred proofs

4.10.1 Fourier transform of $\tanh(\beta\omega/2)/(\beta\omega/2)$

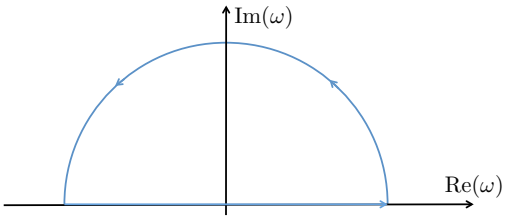
We here derive the Fourier transform of

$$\tilde{f}_\beta(\omega) = \frac{\tanh(\beta\omega/2)}{\beta\omega/2},$$

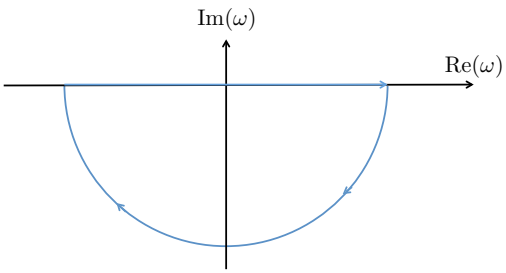
which is

$$f_\beta(t) := \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega t}\tilde{f}_\beta(\omega)d\omega.$$

For the calculation of the Fourier transform, we first consider the case of $t>0$. By defining C^+



(a) Integral path C^+ of ω for $t > 0$



(b) Integral path C^- of ω for $t < 0$

Figure 4-3: Cauchy's integral theorem for the calculation of the Fourier transform.

as a integral path as in Fig. 4-3 (a), we obtain

$$\begin{aligned}\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega t}\tilde{f}_\beta(\omega)d\omega &= \frac{1}{2\pi}\int_{C^+} e^{i\omega t}\tilde{f}_\beta(\omega)d\omega \\ &= i\sum_{m=0}^{\infty}\mathrm{Res}_{\omega=i\pi+2im\pi}[e^{i\omega t}\tilde{f}_\beta(\omega)]. \end{aligned} \tag{4.102}$$

Note that the singular points of $[e^{i\omega t}\tilde{f}_\beta(\omega)]$ are given by $\beta\omega = i\pi(2m+1)$ with m integers. We can calculate the residue as

$$\mathrm{Res}_{\beta\omega=i\pi+2im\pi}[e^{i\omega t}\tilde{f}_\beta(\omega)] = \frac{4e^{-(2m+1)\pi t/\beta}}{\beta\pi}\frac{-i}{2m+1} \tag{4.103}$$

We thus obtain

$$f_\beta(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega t}\tilde{f}_\beta(\omega)d\omega = \frac{4}{\beta\pi}\sum_{m=0}^{\infty}\frac{e^{-(2m+1)\pi t/\beta}}{2m+1}. \tag{4.104}$$

for $t > 0$.

We can perform the same calculation for $t < 0$. In this case, we define C^- as a integral path as in Fig. 4-3 (b), and obtain

$$f_\beta(t) = \frac{1}{2\pi}\int_{C^-} e^{i\omega t}\tilde{f}_\beta(\omega)d\omega = -i\sum_{m=0}^{\infty}\mathrm{Res}_{\omega=-i\pi-2im\pi}[e^{i\omega t}\tilde{f}_\beta(\omega)]. \tag{4.105}$$

By using

$$\mathrm{Res}_{\omega=-i\pi-2im\pi}[e^{i\omega t}\tilde{f}_\beta(\omega)] = \frac{4e^{(2m+1)\pi t/\beta}}{\beta\pi}\frac{i}{2m+1}, \tag{4.106}$$

we have

$$f_\beta(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega t}\tilde{f}_\beta(\omega)d\omega = \frac{4}{\beta\pi}\sum_{m=0}^{\infty}\frac{e^{(2m+1)\pi t/\beta}}{2m+1}. \tag{4.107}$$

for $t < 0$. By combining the above expressions for $f_\beta(t)$, we arrive at

$$f_\beta(t) = \frac{4}{\beta\pi}\sum_{m=0}^{\infty}\frac{e^{-(2m+1)\pi|t|/\beta}}{2m+1}. \tag{4.108}$$

The summation is calculated as

$$\sum_{m=0}^{\infty}\frac{e^{-(2m+1)x}}{2m+1} = \int_x^{\infty}\sum_{m=0}^{\infty}e^{-(2m+1)x'}dx' = \int_x^{\infty}\frac{1}{e^{x'}-e^{-x'}}dx' = \frac{1}{2}\log\frac{e^x+1}{e^x-1} \tag{4.109}$$

for $x > 0$, which yields

$$f_\beta(t) = \frac{2}{\beta\pi}\log\frac{e^{\pi|t|/\beta}+1}{e^{\pi|t|/\beta}-1}. \tag{4.110}$$

Since

$$\log\frac{e^{\pi|t|/\beta}+1}{e^{\pi|t|/\beta}-1} \le \frac{2}{e^{\pi|t|/\beta}-1},$$

$f_\beta(t)$ shows an exponential decay in $|t|$.

4.10.2 Proof of Claims 106 and 107

Proof of Claim 106. Recall that the goal is to show that for every $X \subseteq \Lambda$ and arbitrary unitaries U_X,

$$\|Q_\gamma^A U_X\sqrt{\rho_\beta}\|_F^2 \le c_1 e^{\lambda|X|}\delta_\gamma^{\frac{c_2}{c_2+\beta}}, \tag{4.111}$$

where $Q_\gamma^A = \mathbb{1} - P_\gamma^A$ and P_γ^A was defined in Eq. (4.76) and $H' := H - \frac{1}{\beta}\log Z_\beta$, which allows us to write $\rho_\beta = e^{-\beta H'}$. To prove this inequality, we start from the following expression:

$$\|Q_\gamma^A U_X\sqrt{\rho_\beta}\|_F^2 = \left\|\sum_{m\in\mathbb{Z}} Q_\gamma^A U_X P_m^{H'}\sqrt{\rho_\beta}\right\|_F^2 = \sum_{m\in\mathbb{Z}}\left\|Q_\gamma^A U_X P_m^{H'}\sqrt{\rho_\beta}\right\|_F^2 \tag{4.112}$$

with

$$P_m^{H'} := \sum_{j:\mathcal{E}_j\in(m,m+1]} |j\rangle\langle j|, \tag{4.113}$$

where $|j\rangle$ is the eigenvector of the Hamiltonian H' with $\mathcal{E}_j$ the corresponding eigenvalue. Note that $\sum_{m\in\mathbb{Z}} P_m^{H'} = \mathbb{1}$ and we have $P_m^{H'} = 0$ for $m \notin [-\|H'\|, \|H'\|]$. For some $\Delta \in \mathbb{N}$ which we pick later,

we now decompose $\left\|Q_\gamma^A U_X P_m^{H'} \sqrt{\rho_\beta}\right\|_F^2$ as a sum of the following quantities

$$\left\|Q_\gamma^A U_X P_m^{H'} \sqrt{\rho_\beta}\right\|_F^2 = \left\|Q_\gamma^A \big(P_{>m+\Delta}^{H'} + P_{<m-\Delta}^{H'} + P_{[m-\Delta,m+\Delta]}^{H'}\big) U_X P_m^{H'} \sqrt{\rho_\beta}\right\|_F^2, \tag{4.114}$$

where

$$P_{>m+\Delta}^{H'} := \sum_{m'>m+\Delta} P_{m'}^{H'}, \quad P_{<m+\Delta}^{H'} := \sum_{m'<m+\Delta} P_{m'}^{H'}, \quad P_{[m-\Delta,m+\Delta]}^{H'} := \sum_{m-\Delta\le m'\le m+\Delta} P_{m'}^{H'}. \tag{4.115}$$

Summing over all $m \in \mathbb{Z}$ in Eq. (4.114) and using Eq. (4.112) followed by the triangle inequality gives us the following inequality

$$\begin{aligned}
&\|Q_\gamma^A U_X \sqrt{\rho_\beta}\|_F^2 \\
&\le 2 \underbrace{\sum_{m\in\mathbb{Z}} \left\|Q_\gamma^A P_{[m-\Delta,m+\Delta]}^{H'} U_X P_m^{H'} \sqrt{\rho_\beta}\right\|_F^2}_{:=(1)} + 2 \underbrace{\sum_{m\in\mathbb{Z}} \left\|Q_\gamma^A \big(P_{>m+\Delta}^{H'} + P_{<m-\Delta}^{H'}\big) U_X P_m^{H'} \sqrt{\rho_\beta}\right\|_F^2}_{:=(2)}.
\end{aligned} \tag{4.116}$$

We first bound (1) in Eq. (4.116). Note that for every m,

$$\begin{aligned}
\left\|Q_\gamma^A P_{[m-\Delta,m+\Delta]}^{H'} U_X P_m^{H'} \sqrt{\rho_\beta}\right\|_F^2 &\le \|P_m^{H'} \sqrt{\rho_\beta}\|^2 \cdot \left\|Q_\gamma^A P_{[m-\Delta,m+\Delta]}^{H'}\right\|_F^2 \\
&\le e^{-\beta m} \left\|Q_\gamma^A P_{[m-\Delta,m+\Delta]}^{H'}\right\|_F^2,
\end{aligned} \tag{4.117}$$

where the first inequality used Eq. (4.15). The expression in the last line can be upper bounded as

$$\begin{aligned}
e^{-\beta m} \left\|Q_\gamma^A P_{[m-\Delta,m+\Delta]}^{H'}\right\|_F^2 &= e^{-\beta m} \mathrm{tr}\left[Q_\gamma^A P_{[m-\Delta,m+\Delta]}^{H'}\right] \\
&\le e^{-\beta m} e^{\beta(m+\Delta+1)} \mathrm{tr}\left[Q_\gamma^A P_{[m-\Delta,m+\Delta]}^{H'} \rho_\beta P_{[m-\Delta,m+\Delta]}^{H'}\right] \\
&= e^{\beta(\Delta+1)} \left\|Q_\gamma^A P_{[m-\Delta,m+\Delta]}^{H'} \sqrt{\rho_\beta}\right\|_F^2.
\end{aligned}$$

where the inequality follows from

$$e^{-\beta(m+\Delta+1)} P_{[m-\Delta,m+\Delta]}^{H'} \preceq P_{[m-\Delta,m+\Delta]}^{H'} \rho_\beta P_{[m-\Delta,m+\Delta]}^{H'}.$$

Thus we conclude, from Equation (4.117), that

$$\begin{aligned}
\left\|Q_\gamma^A P_{[m-\Delta,m+\Delta]}^{H'} U_X P_m^{H'} \sqrt{\rho_\beta}\right\|_F^2 &\le e^{\beta(\Delta+1)} \left\|Q_\gamma^A P_{[m-\Delta,m+\Delta]}^{H'} \sqrt{\rho_\beta}\right\|_F^2 \\
&= e^{\beta(\Delta+1)} \sum_{m'\in[m-\Delta,m+\Delta]} \left\|Q_\gamma^A P_{m'}^{H'} \sqrt{\rho_\beta}\right\|_F^2.
\end{aligned}$$

So the first term (1) in Eq. (4.116) can be bounded by

$$\begin{aligned}\sum_{m\in\mathbb{Z}}\left\|Q_\gamma^A P^{H'}_{[m-\Delta,m+\Delta]}U_X P_m^{H'}\sqrt{\rho_\beta}\right\|_F^2 &\le e^{\beta(\Delta+1)}\sum_{m\in\mathbb{Z}}\sum_{m'\in[m-\Delta,m+\Delta]}\left\|Q_\gamma^A P^{H'}_{m'}\sqrt{\rho_\beta}\right\|_F^2\\ &\overset{(1)}{=} e^{\beta(\Delta+1)}\cdot(2\Delta+1)\sum_{m'\in\mathbb{Z}}\left\|Q_\gamma^A P^{H'}_{m'}\sqrt{\rho_\beta}\right\|_F^2\\ &= (2\Delta+1)e^{\beta(\Delta+1)}\left\|Q_\gamma^A\sqrt{\rho_\beta}\right\|_F^2 = (2\Delta+1)e^{\beta(\Delta+1)}\delta_\gamma, \quad (4.118)\end{aligned}$$

where in (1) we use the fact that each m' appears $(2\Delta+1)$ times in the summation $\sum_{m\in\mathbb{Z}}\sum_{m'\in[m-\Delta,m+\Delta]}$.

We now move on to upper bound (2) in Eq. (4.116) as follows. We have

$$\left\|Q_\gamma^A\big(P^{H'}_{>m+\Delta}+P^{H'}_{<m-\Delta}\big)U_X P_m^{H'}\sqrt{\rho_\beta}\right\|_F^2 \le \left\|\big(P^{H'}_{>m+\Delta}+P^{H'}_{<m-\Delta}\big)U_X P_m^{H'}\right\|^2\cdot\|P_m^{H'}\sqrt{\rho_\beta}\|_F^2, \quad (4.119)$$

where we use $\|Q_\gamma^A\|\le 1$. Using Lemma 85, we obtain

$$\left\|\big(P^{H'}_{>m+\Delta}+P^{H'}_{<m-\Delta}\big)U_X P_m^{H'}\right\| \le Ce^{-\lambda(\Delta-|X|)}, \quad (4.120)$$

where C and λ are universal constants. Plugging Eq. (4.120) into Eq. (4.119), we get

$$\left\|Q_\gamma^A\big(P^{H'}_{>m+\Delta}+P^{H'}_{<m-\Delta}\big)U_X P_m^{H'}\sqrt{\rho_\beta}\right\|_F^2 \le C^2e^{-2\lambda(\Delta-|X|)}\|P_m^{H'}\sqrt{\rho_\beta}\|_F^2. \quad (4.121)$$

With this, we can bound (2) in Eq. (4.116) by

$$\sum_{m\in\mathbb{Z}}\left\|Q_\gamma^A\big(P^{H'}_{>m+\Delta}+P^{H'}_{<m-\Delta}\big)U_X P_m^{H'}\sqrt{\rho_\beta}\right\|_F^2 \le \sum_{m\in\mathbb{Z}}C^2e^{-2\lambda(\Delta-|X|)}\|P_m^{H'}\sqrt{\rho_\beta}\|_F^2 = C^2e^{-2\lambda(\Delta-|X|)}, \quad (4.122)$$

where the equality used the fact that $\sum_{m\in\mathbb{Z}}\|P_m^{H'}\sqrt{\rho_\beta}\|_F^2=\mathrm{tr}(\rho_\beta)=1$. Putting together Eq. (4.118) and (4.122) into Eq. (4.116), we finally obtain the upper bound of

$$\|Q_\gamma^A U_X\sqrt{\rho_\beta}\|_F^2 \le (4\Delta+2)e^{\beta(\Delta+1)}\delta_\gamma + 2C^2e^{-2\lambda(\Delta-|X|)}. \quad (4.123)$$

We let $\Delta = c\beta^{-1}\log(1/\delta_\gamma)$, which gives

$$\|Q_\gamma^A U_X\sqrt{\rho_\beta}\|_F^2 \le c_1e^{\lambda|X|}\delta_\gamma^{\frac{c_2}{c_2+\beta}}, \quad (4.124)$$

for some universal constants c_1, c_2. This proves the claim statement. □

We now proceed to prove Claim 107.

Proof of Claim 107. Recall that the aim is to prove that for every $X\subseteq\Lambda$ and unitary U_X we have

$$\left\|AQ_\gamma^A U_X\sqrt{\rho_\beta}\right\|_F^2 \le \frac{e^{\Theta(1)|X|}\delta_\gamma^{\Theta(1)/\beta}}{\gamma}+\frac{\Theta(1)|X|^5}{\gamma^5}\langle A^2\rangle$$

We let c_5, λ_1, τ_1 be $\Theta(1)$ constants as defined in Lemma 86 and $c_1, c_2, \lambda = \mathcal{O}(1)$ be constants given by Claim 106. For the proof, we first decompose Q^A_γ as

$$Q^A_\gamma = \sum_{s=1}^{\infty} P^A_s, \quad P^A_s := P^A_{(s\gamma,(s+1)\gamma]} + P^A_{[-(s+1)\gamma,-s\gamma)}, \quad P^A_0 := P^A_{[-\gamma,\gamma]}, \tag{4.125}$$

where $P^A_{[a,b]}$ is defined as $P^A_{[a,b]} := \sum_{a\le\omega\le b} \Pi_\omega$ (where Π_ω is the subspace spanned by the eigenvectors of A with eigenvalue ω). Using this notation, observe that $\|AQ^A_\gamma U_X \sqrt{\rho_\beta}\|_F$ can be bounded by

$$\begin{aligned}\left\|AQ^A_\gamma U_X \sqrt{\rho_\beta}\right\|_F^2 = \sum_{s=1}^{\infty} \left\|AP^A_s U_X \sqrt{\rho_\beta}\right\|_F^2 \le& \sum_{s=1}^{\infty} \left\|AP^A_s\right\|^2 \cdot \left\|P^A_s U_X \sqrt{\rho_\beta}\right\|_F^2 \\ \le& \gamma^2 \sum_{s=1}^{\infty} (s+1)^2 \left\|P^A_s U_X \sqrt{\rho_\beta}\right\|_F^2,\end{aligned} \tag{4.126}$$

where we use $\|AP^A_s\| \le \gamma(s+1)$ from the definition (4.125) of P^A_s. The norm $\left\|P^A_s U_X \sqrt{\rho_\beta}\right\|_F$ is bounded from above by

$$\left\|P^A_s U_X \sqrt{\rho_\beta}\right\|_F = \left\|P^A_s U_X \sum_{s'=0}^{\infty} P^A_{s'} \sqrt{\rho_\beta}\right\|_F \le \sum_{s'=0}^{\infty} \left\|P^A_s U_X P^A_{s'} \sqrt{\rho_\beta}\right\|_F, \tag{4.127}$$

where we use $\sum_{s'=0}^{\infty} P^A_{s'} = \mathbb{1}$ in the first equation and in the second inequality we use the triangle inequality for the Frobenius norm. Using Lemma 86 we additionally have

$$\|P^A_s U_X P^A_{s'}\| \le c_5 |X| e^{-(\lambda_1 \gamma |s-s'|/|X|)^{1/\tau_1}} \quad \text{for every } s, s' \ge 0, \tag{4.128}$$

where c_5, λ_1 are given in Lemma 86. Using this, we have

$$\begin{aligned}\left\|P^A_s U_X P^A_0 \sqrt{\rho_\beta}\right\|_F &= \left\|P^A_s U_X P^A_0 \sqrt{\rho_\beta}\right\|_F^{1/2} \cdot \left\|P^A_s U_X P^A_0 \sqrt{\rho_\beta}\right\|_F^{1/2} \\ &\overset{(1)}{\le} \left(2c_1 e^{\lambda|X|} \delta_\gamma^{\frac{c_2}{c_2+\beta}}\right)^{1/2} \cdot \left\|P^A_s U_X P^A_0 \sqrt{\rho_\beta}\right\|_F^{1/2} \\ &\overset{(2)}{\le} \left(2c_1 e^{\lambda|X|} \delta_\gamma^{\frac{c_2}{c_2+\beta}}\right)^{1/2} \cdot \left\|P^A_s U_X P^A_0\right\|^{1/2} \cdot \left\|\sqrt{\rho_\beta}\right\|_F^{1/2} \\ &\overset{(3)}{\le} \left(2c_1 e^{\lambda|X|} \delta_\gamma^{\frac{c_2}{c_2+\beta}}\right)^{1/2} \cdot \left(c_5 |X| e^{-(\lambda_1\gamma|s|/|X|)^{1/\tau_1}}\right)^{1/2} \cdot 1 \\ &\overset{(4)}{=} (\delta'_\gamma)^{1/2} \cdot \left(c_5 |X| e^{-(\lambda_1\gamma|s|/|X|)^{1/\tau_1}}\right)^{1/2},\end{aligned} \tag{4.129}$$

where inequality (1) uses $\left\|P^A_s U_X P^A_0 \sqrt{\rho_\beta}\right\|_F \le 2c_1 \delta_\gamma^{\frac{c_2}{c_2+\beta}}$ [9], inequality (2) uses Eq. (4.15), inequality (3) uses Eq. (4.128) and the fact that $\|\sqrt{\rho_\beta}\|_F = \mathrm{tr}(\rho_\beta) = 1$ and equality (4) defines

[9] Since $Q^A_\gamma \le \mathbb{1}$ and $P^A_0 = \mathbb{1} - Q^A_\gamma$, we have

$$\left\|P^A_s U_X P^A_0 \sqrt{\rho_\beta}\right\|_F \le \left\|Q^A_\gamma U_X (\mathbb{1} - Q^A_\gamma) \sqrt{\rho_\beta}\right\|_F \le \left\|Q^A_\gamma U_X \sqrt{\rho_\beta}\right\|_F + \left\|Q^A_\gamma U_X Q^A_\gamma \sqrt{\rho_\beta}\right\|_F \le \left\|Q^A_\gamma U_X \sqrt{\rho_\beta}\right\|_F + \delta_\gamma \le 2c_1 e^{\lambda|X|} \delta_\gamma^{\frac{c_2}{c_2+\beta}},$$

$\delta'_\gamma := 2c_1 e^{\lambda|X|}\delta_\gamma^{\frac{c_2}{c_2+\beta}}$. Using Eq. (4.129), we obtain the following

$$\begin{aligned}\sum_{s'=0}^{\infty}\left\|P_s^A U_X P_{s'}^A\sqrt{\rho_\beta}\right\|_F &\leq \left\|P_s^A U_X P_0^A\sqrt{\rho_\beta}\right\|_F + \sum_{s'=1}^{\infty}\left\|P_s^A U_X P_{s'}^A\right\|\cdot\left\|P_{s'}^A\sqrt{\rho_\beta}\right\|_F \\ &\leq {\delta'}_\gamma^{1/2} c_5^{1/2}|X|^{1/2} e^{-(\lambda_1\gamma s/|X|)^{1/\tau_1}/2} \\ &\quad + \sum_{s'=1}^{\infty} c_5|X| e^{-(\lambda_1\gamma|s-s'|/|X|)^{1/\tau_1}}\left\|P_{s'}^A\sqrt{\rho_\beta}\right\|_F, \end{aligned} \tag{4.130}$$

where the first term in the inequality was obtained from Eq. (4.129) and the second term was obtained from Eq. (4.128). We now upper-bound the summation in the second term of Eq. (4.130) by using the Cauchy–Schwarz inequality as follows:

$$\begin{aligned}&\sum_{s'=1}^{\infty} e^{-(\lambda_1\gamma|s-s'|/|X|)^{1/\tau_1}}\left\|P_{s'}^A\sqrt{\rho_\beta}\right\|_F \\ &= \sum_{s'=1}^{\infty} e^{-(\lambda_1\gamma|s-s'|/|X|)^{1/\tau_1}/2}\cdot\left(e^{-(\lambda_1\gamma|s-s'|/|X|)^{1/\tau_1}/2}\left\|P_{s'}^A\sqrt{\rho_\beta}\right\|_F\right) \\ &\leq \left(\sum_{s'=1}^{\infty} e^{-(\lambda_1\gamma|s-s'|/|X|)^{1/\tau_1}}\right)^{1/2}\left(\sum_{s'=1}^{\infty} e^{-(\lambda_1\gamma|s-s'|/|X|)^{1/\tau_1}}\left\|P_{s'}^A\sqrt{\rho_\beta}\right\|_F^2\right)^{1/2} \\ &\overset{(1)}{\leq} \xi^{1/2}\left(\sum_{s'=1}^{\infty} p_{s'}^A e^{-(\lambda_1\gamma|s-s'|/|X|)^{1/\tau_1}}\right)^{1/2}, \end{aligned} \tag{4.131}$$

with $\xi = 2 + \frac{2\tau_1|X|}{\lambda_1\gamma}(2\tau_1)^{\tau_1}$, where $p_{s'} := \left\|P_{s'}^A\sqrt{\rho_\beta}\right\|_F^2$ and we used Fact 78 in inequality (1). Note that $\sum_{s'=1}^{\infty} p_{s'} = \|Q_\gamma^A\sqrt{\rho_\beta}\|_F^2$ because of Eq. (4.125). We can obtain the following upper bound by

where the first inequality used $P_s^A \preceq Q_\gamma^A$ $(s \geq 1)$ and the last inequality used $\|Q_\gamma^A U_X\sqrt{\rho_\beta}\|_F \leq c_1 e^{\lambda|X|}\delta_\gamma^{\frac{c_2}{c_2+\beta}}$ from Claim 106.

combining the equations Eq. (4.127), (4.130) and (4.131):

$$
\begin{aligned}
&\left\|P_s^A U_X \sqrt{\rho_\beta}\right\|_F^2 \\
&\leq \Big(\sum_{s'=0}^{\infty} \left\|P_s^A U_X P_{s'}^A \sqrt{\rho_\beta}\right\|_F \Big)^2 \\
&\leq \Big({\delta'}_\gamma^{1/2} c_5^{1/2} |X|^{1/2} e^{-(\lambda_1 \gamma s/|X|)^{1/\tau_1}/2} + \sum_{s'=1}^{\infty} c_5 |X| e^{-(\lambda_1 \gamma |s-s'|/|X|)^{1/\tau_1}} \left\|P_{s'}^A \sqrt{\rho_\beta}\right\|_F \Big)^2 \\
&\leq \Big({\delta'}_\gamma^{1/2} c_5^{1/2} |X|^{1/2} e^{-(\lambda_1 \gamma s/|X|)^{1/\tau_1}/2} + c_5 |X| \xi^{1/2} \cdot \Big(\sum_{s'=1}^{\infty} p_{s'}^A e^{-(\lambda_1 \gamma |s-s'|/|X|)^{1/\tau_1}} \Big)^{1/2} \Big)^2 \\
&\leq \underbrace{2 c_5 {\delta'}_\gamma |X| e^{-(\lambda_1 \gamma s/|X|)^{1/\tau_1}}}_{:=f_1(s)} + \underbrace{2 c_5^2 |X|^2 \xi \cdot \Big(\sum_{s'=1}^{\infty} p_{s'}^A e^{-(\lambda_1 \gamma |s-s'|/|X|)^{1/\tau_1}} \Big)}_{:=f_2(s)}.
\end{aligned}
$$

Recall that the goal of this claim was to upper bound Eq. (4.126), which we can rewrite now as

$$
\left\|A Q_\gamma^A U_X \sqrt{\rho_\beta}\right\|_F^2 \leq \gamma^2 \sum_{s=1}^{\infty} (s+1)^2 \left\|P_s^A U_X \sqrt{\rho_\beta}\right\|_F^2 \leq \gamma^2 \sum_{s=1}^{\infty} (s+1)^2 f_1(s) + \gamma^2 \sum_{s=1}^{\infty} (s+1)^2 f_2(s). \quad (4.132)
$$

We bound each of these terms separately. In order to bound the first term, observe that

$$
\begin{aligned}
\gamma^2 \sum_{s=1}^{\infty} (s+1)^2 f_1(s) &= 2\gamma^2 c_5 {\delta'}_\gamma |X| \sum_{s=1}^{\infty} (s+1)^2 e^{-(\lambda_1 \gamma s/|X|)^{1/\tau_1}} \\
&\overset{(1)}{\leq} 2\gamma^2 c_5 {\delta'}_\gamma |X| \cdot 8\tau_1 \cdot \left(\frac{(3\tau_1)^{\tau_1} |X|}{\lambda_1 \gamma} \right)^3 \leq \frac{16 c_5 {\delta'}_\gamma |X|^4 \tau_1 (3\tau_1)^{3\tau_1}}{\lambda_1^3 \gamma},
\end{aligned}
$$

where inequality (1) uses Fact 78. We now bound the second term in Eq. (4.132) as follows

$$
\begin{aligned}
\gamma^2 \sum_{s=1}^{\infty} (s+1)^2 f_2(s) &= 2\gamma^2 c_5^2 |X|^2 \xi \sum_{s=1}^{\infty} (s+1)^2 \Big(\sum_{s'=1}^{\infty} p_{s'}^A e^{-(\lambda_1 \gamma |s-s'|/|X|)^{1/\tau_1}} \Big). \\
&= 2\gamma^2 c_5^2 |X|^2 \xi \sum_{s'=1}^{\infty} p_{s'}^A \Big(\sum_{s=1}^{\infty} (s+1)^2 e^{-(\lambda_1 \gamma |s-s'|/|X|)^{1/\tau_1}} \Big) \\
&\leq 2\gamma^2 c_5^2 |X|^2 \xi \sum_{s'=1}^{\infty} p_{s'}^A (2s')^2 \Big(\sum_{s=1}^{\infty} (1+|s-s'|)^2 e^{-(\lambda_1 \gamma |s-s'|/|X|)^{1/\tau_1}} \Big) \\
&\overset{(1)}{\leq} 2\gamma^2 c_5^2 |X|^2 \xi \cdot \xi' \sum_{s'=1}^{\infty} p_{s'}^A (2s')^2,
\end{aligned} \quad (4.133)
$$

with $\xi' = 2 + 16 \left(\frac{(3\tau_1)^{\tau_1} |X|}{\lambda_1 \gamma} \right)^3$, where inequality (1) follows from Fact 78. Further upper bound this

expression by simplifying the pre-factors, we get

$$\begin{aligned}\gamma^2\sum_{s=1}^{\infty}(s+1)^2 f_2(s) &\le 8\gamma^2 c_5^2|X|^2\xi\xi'\sum_{s'=1}^{\infty}p^A_{s'}(s')^2\\ &= 8c_5^2|X|^2\xi\xi'\sum_{s'=1}^{\infty}(\gamma s')^2 p^A_{s'}\\ &= 8c_5^2|X|^2\xi\xi'\sum_{s'=1}^{\infty}\big\|(\gamma s')P^A_{s'}\sqrt{\rho_\beta}\big\|_F^2\\ &\overset{(2)}{\le} 8c_5^2|X|^2\xi\xi'\sum_{s'=0}^{\infty}\big\|P^A_{s'}A\sqrt{\rho_\beta}\big\|_F^2 = 8c_5^2|X|^2\xi\xi'\langle A^2\rangle,\end{aligned} \tag{4.134}$$

In inequality (2), we used $(\gamma s')P^A_{s'} \preceq AP^A_{s'}$ from the definition (4.125) of $P^A_{s'}$. Note that $|X|^2\xi\xi'$ is bounded from above by $\mathcal{O}(|X|^6)$). By combining the above inequalities altogether, we prove Eq. (4.82). □

4.10.3 Derivation of the sub-exponential concentration

Recall that the goal in this section is to prove the following lemma.

Lemma 112 (Restatement of Lemma 86)**.** *Let A be a $(\tau, a_1, a_2, 1)$-quasi-local operator with $\tau < 1$, as given in Eq. (4.18). For an arbitrary operator O_X supported on a subset $X \subseteq \Lambda$ with $|X| = k_0$ and $\|O_X\| = 1$, we have*

$$\|P^A_{\ge x+y}O_X P^A_{\le x}\| \le c_5\cdot k_0 \exp\Big(-(\lambda_1 y/k_0)^{1/\tau_1}\Big), \tag{4.135}$$

where $\tau_1 := \frac{2}{\tau} - 1$ and c_5 and λ_1 are constants which only depend on a_1 and a_2. In particular, the a_2 dependence of c_5 and λ_1 is given by $c_5 \propto a_2^{2/\tau}$ and $\lambda_1 \propto a_2^{-2/\tau}$ respectively.

Before proving this lemma, let us elaborate upon the method. Recall that

$$P^A_{\le x} = \sum_{\omega\le x}\Pi_\omega, \quad P^A_{>y} = \sum_{\omega>y}\Pi_\omega, \tag{4.136}$$

where Π_ω is the projector onto the eigenvalue ω eigenspace of A. One way to prove the upper bound in the estimation of the norm (4.135) is to utilize the technique in Ref. [AKL16] (i.e., Lemma 85). The argument proceeds by considering

$$\begin{aligned}\|P^A_{\ge x+y}O_X P^A_{\le x}\| &= \|P^A_{\ge x+y}e^{-\nu A}e^{\nu A}O_X e^{-\nu A}e^{\nu A}P^A_{\le x}\|\\ &\le \|P^A_{\ge x+y}e^{-\nu A}\|\cdot\|e^{\nu A}O_X e^{-\nu A}\|\cdot\|e^{\nu A}P^A_{\le x}\|\\ &\le e^{-\nu y}\|e^{\nu A}O_X e^{-\nu A}\|,\end{aligned} \tag{4.137}$$

which reduces the problem to estimation of the norm $\|e^{\nu A}O_X e^{-\nu A}\|$. Additionally, by definition

of A in Theorem 110 we have

$$A = \sum_{\ell=1}^{n} g_\ell \bar{A}_\ell, \tag{4.138}$$

where $\bar{A}_\ell$ is κ-local and g_ℓ is sub-exponentially decaying function for ℓ (as made precise in Eq. (4.18)), namely $g_\ell \propto \exp(-\mathcal{O}(\ell^\tau))$. In this case, for $\nu = \Theta(1)$ in (4.137), the norm of the imaginary time evolution can be finitely bounded only in the case $D = 1$ [Kuw16]. That is, the norm $\|e^{\nu A} O_X e^{-\nu A}\|$ diverges to infinity for $\tau < 1$. However, our main contribution in this section is that we are able to prove the lemma statement *without* going through the inequalities in (4.137) (which in turn used earlier results of [Kuw16, AKL16]). We now give more details.

Proof of Lemma 112

In order to estimate the norm, we need to take a different route from (4.137). Recall that A is a $(\tau, a_1, a_2, 1)$-quasi local operator. Let I be any interval of the real line and P_I^A be the projector onto the eigenspace of A with eigenvalues in I. Using the operator inequality

$$P_{\geq z}^A (A - \omega \mathbb{1})^m \succeq (z - \omega)^m P_{\geq z}^A,$$

we obtain

$$\|(A - \omega\mathbb{1})^m O_X P_I^A\| \geq \|P_{\geq z}^A (A - \omega)^m O_X P_I^A\| \geq (z - \omega)^m \|P_{\geq z}^A O_X P_I^A\|, \tag{4.139}$$

hence

$$\|P_{\geq z}^A O_X P_I^A\| \leq \frac{\|(A - \omega)^m O_X P_I^A\|}{(z - \omega)^m}. \tag{4.140}$$

Our strategy to establish Eq. (4.135) will be to expand

$$\|P_{\geq x+y}^A O_X P_{\leq x}^A\| \leq \sum_{j=0}^{\infty} \|P_{\geq x+y}^A O_X P_{I_j}^A\|, \tag{4.141}$$

for carefully chosen intervals $I_j := (x - a_1(j+1), x - a_1 j]$ (the term a_1 is as given in the statement of Lemma 112). Towards this, let us fix an arbitrary ω and an interval $I := (\omega - a_1, \omega]$, and prove an upper bound on $\|P_{\geq \omega+\theta}^A O_X P_I^A\|$ (for all θ). We show the following claim.

Claim 113. *Let A be a $(\tau, a_1, a_2, 1)$-quasi local operator as defined in (4.18). There is a constant c_6 such that*

$$\|P_{\geq \omega+\theta}^A O_X P_I^A\| \leq \frac{1}{\tau} \exp\left[-[\theta/(ec_6 k_0)]^{1/\tau_1} + 1\right]. \tag{4.142}$$

The claim is proved in subsection 4.10.3. Let us use the claim to establish the lemma. In the inequality (4.141), we need to estimate $\|P_{\geq x+y}^A O_X P_{I_j}^A\|$ with $I_j := (x - (j+1)a_1, x - j a_1]$. Setting

$\omega = x - ja_1$ and $\theta = y + ja_1$ in Claim 113, we have

$$\|P^A_{\geq x+y} O_X P^A_{I_j}\| \leq \frac{1}{\tau} \exp\left\{ - \left(\frac{y + a_1 j}{e c_6 k_0} \right)^{1/\tau_1} + 1 \right\}. \tag{4.143}$$

In order to complete the bound on Equation (4.141), we need to take summation with respect to j. We have

$$\sum_{j=0}^{\infty} \|P^A_{\geq x+y} O_X P^A_{I_j}\| \leq \sum_{j=0}^{\infty} \frac{1}{\tau} \exp\left\{ - \left(\frac{y + a_1 j}{e c_6 k_0} \right)^{1/\tau_1} + 1 \right\} \leq \frac{1}{\tau} e^{-\frac{1}{2}\left(\frac{y}{e c_6 k_0}\right)^{1/\tau_1}} \left(1 + \frac{e c_6 k_0 \tau_1}{a_1} (2\tau_1)^{1/\tau_1} \right), \tag{4.144}$$

where in last inequality we used Fact 78 (3) with $c = (e c_6 k_0 / a_1)^{-1/\tau_1}$, $p = 1/\tau_1$ and $a = y/a_1$. This gives the form of (4.135) and completes the proof.

Proof of Claim 113

From Equation (4.140), it suffices to upper bound $\|(A - \omega)^m O_X P^A_I\|$. Abbreviate $\tilde{A} := A - \omega \mathbb{1}$. Introduce the multi-commutator

$$\mathrm{ad}^s_{\tilde{A}}(O_X) := \underbrace{[\tilde{A}, \ldots [\tilde{A}, [\tilde{A}, O_X]] \ldots]}_{s \text{ times}}.$$

Consider the following identity,

$$\tilde{A}^m O_X P^A_I = \sum_{s=0}^{m} \binom{m}{s} \mathrm{ad}^s_{\tilde{A}}(O_X) \tilde{A}^{m-s} P^A_I. \tag{4.145}$$

This shows that

$$\|\tilde{A}^m O_X P^A_I\| \leq \sum_{s=0}^{m} \binom{m}{s} \|\mathrm{ad}^s_{\tilde{A}}(O_X)\| \cdot \|\tilde{A}^{m-s} P^A_I\| \leq \sum_{s=0}^{m} \binom{m}{s} a_1^{m-s} \|\mathrm{ad}^s_{\tilde{A}}(O_X)\|, \tag{4.146}$$

where we use $\|\tilde{A}^{m-s} P^A_I\| = \|(A - \omega)^{m-s} P^A_I\| \leq a_1^{m-s}$ for $I = (\omega - a_1, \omega]$. The remaining task is to estimate the upper bound of $\|\mathrm{ad}^s_{\tilde{A}}(O_X)\| = \|\mathrm{ad}^s_A(O_X)\|$. This is done in the following claim.

Claim 114. *Let A be a $(\tau, a_1, a_2, 1)$-quasi local operator as defined in (4.18). Then, for an arbitrary operator O_X which is supported on a subset X ($|X| = k_0$), the norm of the multi-commutator $\mathrm{ad}^s_A(O_X)$ is bounded from above by*

$$\|\mathrm{ad}^s_A(O_X)\| \leq \frac{(2a_1)^s (2k_0)^s e^s}{\tau} \cdot \left(\frac{2}{a_2 \tau} \right)^{\frac{2s}{\tau}} \cdot (s^{\tau_1})^s \quad \text{for} \quad s \leq m, \tag{4.147}$$

where the constants a_1 and a_2 have been defined in Eq. (4.18).

By applying the inequality (4.147) to (4.146), we obtain

$$\begin{aligned}\|\tilde{A}^m O_X P_I^A\| &\leq \sum_{s=0}^{m} \binom{m}{s} a_1^{m-s} \frac{(2a_1)^s (2k_0)^s e^s}{\tau} \cdot \left(\frac{2}{a_2\tau}\right)^{\frac{2s}{\tau}} \cdot (s^{\tau_1})^s \\ &\leq \sum_{s=0}^{m} \binom{m}{s} (2a_1)^m \frac{(2ek_0)^m}{\tau} \cdot \left(\frac{2}{a_2\tau}\right)^{\frac{2m}{\tau}} \cdot (m^{\tau_1})^m \\ &= (4a_1)^m \frac{(2ek_0)^m}{\tau} \cdot \left(\frac{2}{a_2\tau}\right)^{\frac{2m}{\tau}} \cdot (m^{\tau_1})^m = \frac{1}{\tau}\left[8ea_1k_0[2/(a_2\tau)]^{2/\tau} m^{\tau_1}\right]^m .\end{aligned}$$

Therefore, setting $z = \omega + \theta$ in the inequality (4.140), we obtain

$$\|P^A_{\geq \omega+\theta} O_X P_I^A\| \leq \frac{\|\tilde{A}^m O_X P_I^A\|}{\theta^m} \leq \frac{1}{\tau}\left[8ea_1k_0[2/(a_2\tau)]^{2/\tau} \frac{m^{\tau_1}}{\theta}\right]^m \tag{4.148}$$

$$\leq \frac{1}{\tau}\left(\frac{c_6 k_0 m^{\tau_1}}{\theta}\right)^m , \tag{4.149}$$

where $c_6 := 8ea_1[2/(a_2\tau)]^{2/\tau}$. Let us choose $m = \tilde{m}$ with $\tilde{m}$ the minimum integer such that

$$\frac{c_6 k_0 \tilde{m}^{\tau_1}}{\theta} \leq 1/e. \tag{4.150}$$

The above condition is satisfied by $\tilde{m}^{\tau_1} \leq \theta/(ec_6k_0)$, which implies

$$\tilde{m} = \left\lfloor [\theta/(ec_6k_0)]^{1/\tau_1} \right\rfloor , \tag{4.151}$$

where $\lfloor\cdot\rfloor$ is the floor function. From this choice, the claim concludes.

Proof of Claim 114

As stated in Claim 114, we let A be a $(\tau, a_1, a_2, 1)$-quasi local operator as defined in (4.18). Recall that we need to show, for an arbitrary operator O_X which is supported on k_0 sites, the norm of the multi-commutator $\mathrm{ad}_A^s(O_X)$ is bounded by

$$\|\mathrm{ad}_A^s(O_X)\| \leq \frac{(2a_1)^s (2k_0)^s e^s}{\tau} \cdot \left(\frac{2}{a_2\tau}\right)^{\frac{2s}{\tau}} \cdot (s^{\tau_1})^s \quad \text{for} \quad s \leq m.$$

We start from the following expansion:

$$\mathrm{ad}_A^s(O_X) = \sum_{k_1,k_2,\ldots,k_s} g_{k_1} g_{k_2} \cdots g_{k_s} [\bar{A}_{k_s}, [\bar{A}_{k_{s-1}}, \cdots [\bar{A}_{k_1}, O_X] \cdots].$$

By using Lemma 3 in Ref. [KMS16] and setting $\zeta = 1$ (see Definition 80) we obtain

$$\|[[\bar{A}_{k_s}, [\bar{A}_{k_{s-1}}, \cdots [\bar{A}_{k_1}, O_X] \cdots]\| \leq 2^s k_0 (k_0 + k_1)(k_0 + k_1 + k_2) \cdots (k_0 + k_1 + k_2 + \cdots + k_{s-1}), \tag{4.152}$$

where we used that A is a $(\tau, a_1, a_2, 1)$-quasi local operator. Recall that we set $\|O_X\| = 1$ and $|X| = k_0$. The norm of $\mathrm{ad}_A^s(O_X)$ is bounded from above by

$$\begin{aligned}
&\|\mathrm{ad}_A^s(O_X)\| \\
&\leq \sum_{k_1,k_2,\ldots,k_s=1}^{\infty} 2^s g_{k_1} g_{k_2} \cdots g_{k_s} k_0 (k_0+k_1)(k_0+k_1+k_2)\cdots(k_0+k_1+k_2+\cdots+k_{s-1}) \\
&= \sum_{K\geq s} \sum_{\substack{k_1+k_2+\ldots+k_s=K \\ k_1\geq 1, k_2\geq 1,\ldots,k_s\geq 1}} 2^s g_{k_1} g_{k_2} \cdots g_{k_s} k_0 (k_0+k_1)(k_0+k_1+k_2)\cdots(k_0+k_1+k_2+\cdots+k_{s-1}),
\end{aligned} \tag{4.153}$$

where the summation over K starts from s because each of $\{k_j\}_{j=1}^s$ is larger than 1. Now, using the inequality $\log[g_k/a_1] \leq -a_2 k^\tau$ with $\tau \leq 1$, we have $\sum_{j=1}^s \log(g_{k_j}/a_1) \leq \log(g_{k_1+k_2+\cdots+k_s}/a_1)$. This follows from $\sum_{j=1}^s k_j^\tau \geq (k_1+k_2+\cdots+k_s)^\tau$. Thus, using $k_1+k_2+\cdots+k_s = K$, the summand in the inequality (4.153) is upper-bounded by

$$g_{k_1} g_{k_2} \cdots g_{k_s} k_0 (k_0+k_1)(k_0+k_1+k_2)\cdots(k_0+k_1+k_2+\cdots+k_{s-1}) \leq a_1^s (g_K/a_1) k_0 (k_0+K)^{s-1}, \tag{4.154}$$

where we use the inequality $k_1+k_2+\cdots+k_j \leq K$ for $j = 1, 2, \ldots, s-1$. By combining the two inequalities (4.153) and (4.154), we obtain

$$\begin{aligned}
\|\mathrm{ad}_A^s(O_X)\| &\leq \sum_{K\geq s} \sum_{\substack{k_1+k_2+\ldots+k_s=K \\ k_1\geq 1, k_2\geq 1,\ldots,k_s\geq 1}} (2a_1)^s (g_K/a_1) k_0 (k_0+K)^{s-1} \\
&\overset{(1)}{\leq} \sum_{K\geq s} \left(\!\!\binom{s}{K-s}\!\!\right) (2a_1)^s (g_K/a_1) k_0 (k_0+K)^{s-1} \\
&= \sum_{K\geq s} \binom{K-1}{s-1} (2a_1)^s (g_K/a_1) k_0 (k_0+K)^{s-1} \\
&\overset{(2)}{\leq} (2a_1)^s \frac{(2k_0)^s}{2} \sum_{K\geq s} \frac{e^s K^s}{s^s} (g_K/a_1) (K)^{s-1} \\
&\overset{(3)}{\leq} \frac{(2a_1)^s (2k_0)^s e^s}{s^s} \cdot \frac{1}{2} \sum_{K\geq s} K^{2s-1} e^{-a_2 K^\tau} \leq \frac{(2a_1)^s (2k_0)^s e^s}{s^s} \cdot \frac{1}{2} \sum_{K\geq 0} K^{2s-1} e^{-a_2 K^\tau} \\
&\overset{(4)}{\leq} \frac{(2a_1)^s (2k_0)^s e^s}{s^s \tau} \cdot \left(\frac{2s}{a_2 \tau}\right)^{\frac{2s}{\tau}} = \frac{(2a_1)^s (2k_0)^s e^s}{\tau} \cdot \left(\frac{2}{a_2\tau}\right)^{\frac{2s}{\tau}} \cdot \left(s^{\frac{2}{\tau}-1}\right)^s,
\end{aligned}$$

where in (1), $\left(\!\!\binom{\ }{\ }\!\!\right)$ denotes the multi-combination, namely $\left(\!\!\binom{n}{m}\!\!\right) = \binom{n+m-1}{n-1}$, in (2) we upper bound $\binom{K-1}{s-1} \leq \frac{e^s K^s}{s^s}$, $k_0 + K \leq 2k_0 K$, in (3) we use the sub-exponential form of g_K in Eq. (4.18) and in (4) we use Fact 78. Since $\tau_1 = \frac{2}{\tau} - 1$, this proves the statement.

4.10.4 Quasi-locality of $\widetilde{W}$

We here aim to obtain (τ, a_1, a_2, ζ)-quasi-locality of the operator $\widetilde{W}$, where $\{\tau, a_1, a_2, \zeta\}$ defined in Definition 80. This is used in Section 4.9.6. In particular, we will show that

$$(\tau, a_1, a_2, \zeta) = \left(1/D, \mathcal{O}(1), \mathcal{O}(1/\beta), \mathcal{O}(\beta^{2D+1})\left(\max_{j\in\Lambda} v_j\right)\right)$$

suffices to prove the quasi-locality of $\widetilde{W}$. Recall the definition of $\widetilde{W}$:

$$\widetilde{W} = \int_{-\infty}^{\infty} f_\beta(t)\ e^{-iHt}\ W\ e^{iHt} dt,$$

where

$$f_\beta(t) = \frac{2}{\beta\pi} \log \frac{e^{\pi|t|/\beta}+1}{e^{\pi|t|/\beta}-1}$$

and

$$W = \sum_{i\in\Lambda} v_i E_i.$$

We write

$$\widetilde{W} = \sum_i v_i \int_{-\infty}^{\infty} f_\beta(t)\ e^{-iHt}\ E_i\ e^{iHt} dt.$$

Abbreviate

$$\tilde{E}_i(t) := e^{-iHt}\ E_i\ e^{iHt}$$

and recall that $\tilde{E}_i = \int_\infty^\infty f_\beta(t)\tilde{E}_i(t)$. Moreover, (with some abuse of notation) let $B(r,i) \subseteq \Lambda$ denote the ball of radius r such that: the centre of $B(r,i)$ coincides with the the center of the smallest ball containing E_i. We assume that r ranges in the set $\{m_i, m_i+1, \ldots, n_i\}$, where m_i is the radius of the smallest ball containing E_i and n_i is the number such that $B(n_i, i) = \Lambda$. Define

$$\tilde{E}_i^r(t) := \mathrm{tr}_{B(r,i)^c}[\tilde{E}_i(t)] \otimes \frac{\mathbb{1}_{B(r,i)^c}}{\mathrm{tr}[\mathbb{1}_{B(r,i)^c}]}, \quad \tilde{E}_i^0(t) = 0,$$

i.e., $\tilde{E}_i^r(t)$ traces out all the qudits in $\tilde{E}_i(t)$ that are at outside the $B(r,i)$-ball around $\tilde{E}_i^r$. From [BHV06], we have

$$\|\tilde{E}_i(t) - \tilde{E}_i^r(t)\| \leq \|E_i\| \min\left\{1, c_3 r^{D-1} e^{-c_4(r - m_i - v_{\mathrm{LR}}|t|)}\right\}$$

which in particular implies

$$\|\tilde{E}_i^r(t) - \tilde{E}_i^{r-1}(t)\| \leq 2\min\left\{1, c_3 e^{c_4 m_i} r^{D-1} e^{-c_4(r - v_{\mathrm{LR}}|t|)}\right\},$$

where we use $\|E_i\| = 1$, v_{LR} is the Lieb-Robinson velocity (as defined in Fact 81) and c_3, c_4 are constants. We note that the $2\min\{1,\cdot\}$ is derived from the trivial upper bound $\|\tilde{E}_i^r(t) - \tilde{E}_i^{r+1}(t)\| \leq 2$.

This allows us to write the following quasi-local expression:

$$\tilde{E}_i(t) = \sum_{r=m_i}^{n_i} \left(\tilde{E}_i^r(t) - \tilde{E}_i^{r-1}(t) \right).$$

Using this, we can now write the quasi-local representation of $\tilde{E}_i$ as follows.

$$\int_{-\infty}^{\infty} f_\beta(t) \tilde{E}_i(t) dt = \int_{-\infty}^{\infty} f_\beta(t) \sum_{r=m_i}^{n_i} \left(\tilde{E}_i^r(t) - \tilde{E}_i^{r-1}(t) \right).$$

To see that it is quasi-local, observe that the term with radius r has norm

$$\begin{aligned}
&\int_{-\infty}^{\infty} f_\beta(t) \left\| \tilde{E}_i^r(t) - \tilde{E}_i^{r-1}(t) \right\| \\
&\leq 2c_3 e^{c_4 m_i} r^{D-1} \cdot \left(e^{-c_4 r} \int_{|t| \leq r/(2v_{\mathrm{LR}})} e^{c_4 v_{\mathrm{LR}} |t|} f_\beta(t) dt + \int_{|t| > r/(2v_{\mathrm{LR}})} f_\beta(t) dt \right) \\
&\leq 2c_3 e^{c_4 m_i} r^{D-1} \cdot \left(e^{-c_4 r/2} \int_{-\infty}^{\infty} f_\beta(t) dt + \int_{|t| > r/(2v_{\mathrm{LR}})} \frac{2\beta}{\pi |t|} e^{-\pi |t|/\beta} dt \right) \\
&\leq 2c_3 e^{c_4 m_i} r^{D-1} \left(e^{-c_4 r/2} + \frac{4\beta v_{\mathrm{LR}}}{\pi r} \frac{e^{-\pi r/(2\beta v_{\mathrm{LR}})}}{\pi/\beta} \right) \\
&= 2c_3 e^{c_4 m_i} r^{D-1} \left(e^{-c_4 r/2} + \frac{4\beta^2 v_{\mathrm{LR}}}{\pi^2} r^{-1} e^{-\pi r/(2\beta v_{\mathrm{LR}})} \right) \leq c_5 r^{D-1} e^{-c_4' r},
\end{aligned}$$

with $c_4' = \min(c_4/2, \pi/(2\beta v_{\mathrm{LR}}))$, where $c_5 = \mathcal{O}(\beta^2)$ is a constant which does not depend on r, and we use $\int_{-\infty}^{\infty} f_\beta(t) = \tilde{f}_\beta(0) = 1$ and $f_\beta(t) \leq 2\beta/(\pi|t|) e^{-\pi|t|/\beta}$. Note that $c_4' = \mathcal{O}(1/\beta)$. Define

$$a_{B(r,i)} := e^{c_4' r/2} \int_{-\infty}^{\infty} f_\beta(t) \left(\tilde{E}_i^r(t) - \tilde{E}_i^{r-1}(t) \right).$$

Here, the operator $a_{B(r,i)}$ is supported on the subset $B(r,i)$. Then, from $|B(r,i)| = \mathcal{O}(r^D)$, the quasi-local representation of $\widetilde{W}$ is given as

$$\widetilde{W} = \sum_{i \in \Lambda} v_i \sum_{r=m_i}^{n_i} e^{-c_4' r/2} a_{B(r,i)} = \sum_{i \in \Lambda} \sum_{r=m_i}^{n_i} e^{-\mathcal{O}(|B(r,i)|^{\frac{1}{D}})} v_i a_{B(r,i)},$$

with $e^{-\mathcal{O}(|B(r,i)|^{\frac{1}{D}})}$ decaying sub-exponentially with rate $\tau = 1/D$, for all $i \in \Lambda$. We also obtain the parameter ζ in Eq. (4.18), which can be calculated for a fixed r. For an arbitrary r, we have

$$\begin{aligned}
\sum_{r,j: B(r,j) \ni i} v_j \| a_{B(r,j)} \| \leq c_5 r^{D-1} \sum_{j: B(r,j) \ni i} v_j e^{-c_4' r/2} &\leq \left(\max_{j \in \Lambda} v_j \right) c_5 c_B r^{2D-1} e^{-c_4' r/2} \\
&\leq \Theta(\beta^{2D+1}) \left(\max_{j \in \Lambda} v_j \right),
\end{aligned}$$

where we define c_B such that $|B(r,j)| \leq c_B r^D$ and we used $c_5 = \mathcal{O}(\beta^2)$. This completes the representation and shows that $\widetilde{W}$ is a $\big(1/D, \mathcal{O}(1), \mathcal{O}(1/\beta), \mathcal{O}(\beta^{2D+1})\,(\max_{j\in\Lambda} v_j)\big)$-quasi-local.

4.10.5 Proof of Lemma 111

Recall that the goal in this section is to prove that for $\widetilde{W}$ defined in Lemma 100 we have

$$\max_{i\in\Lambda} \mathrm{tr}[(\widetilde{W}_{(i)})^2\eta] = \frac{\Theta(1)}{(\beta\log(\beta)+1)^{2D+2}}\left(\max_{i\in\Lambda} v_i^2\right),$$

where η is the maximally mixed state. In this direction, we will now prove that

$$\max_{i\in\Lambda} \|\widetilde{W}_{(i)}\sqrt{\eta}\|_F \geq \frac{c_7}{(\beta\log(\beta)+1)^{D+1}} \max_{i\in\Lambda}(|v_i|), \tag{4.155}$$

for a constant $c_7 = \mathcal{O}(1)$. For convenience, let us define $\mathrm{argmax}_{i\in\Lambda}|v_i| = i_+$, or equivalently $|v_{i_+}| = \max_{i\in\Lambda}|v_i|$. In the following, we prove the inequality (4.155) for $\|\widetilde{W}_{(i+)}\sqrt{\eta}\|_F$ instead of $\max_{i\in\Lambda}\|\widetilde{W}_{(i)}\sqrt{\eta}\|_F$. By using the inequality $\max_{i\in\Lambda}\|\widetilde{W}_{(i)}\|_F \geq \|\widetilde{W}_{(i_+)}\|_F$, we obtain the main statement. We denote the ball region $B(r,i_+)$ by B_r for the simplicity, where r is fixed later. Let us consider $\widetilde{W}[B_r]$ which is defined as follows:

$$\widetilde{W}[B_r] := \int_{-\infty}^{\infty} f_\beta(t) e^{-iHt} W[B_r] e^{iHt} dt, \quad W[B_r] := \sum_{i\in B_r} v_i E_i. \tag{4.156}$$

Since $\widetilde{W}[B_r]$ is obtained from $W[B_r]$ in an equivalent manner as $\widetilde{W}$ is obtained from W, the following claim follows along the same lines as Theorem 102. We skip the very similar proof.

Claim 115. *It holds that*

$$\|\widetilde{W}[B_r]\|_F^2 \geq \frac{\mathcal{D}_\Lambda}{c_5[\beta\log(r)+1]^2}\sum_{i\in B_r} v_i^2,$$

where c_5 is a constant of $\mathcal{O}(1)$.

Since the new operator $\widetilde{W}[B_r]$ well approximates the property of $\widetilde{W}$ around the site i_+, as long as r is sufficiently large, we expect that $\widetilde{W}[B_r]_{(i_+)}$ and $\widetilde{W}_{(i_+)}$ are close to each other. The claim below makes this intuition rigorous:

Claim 116. *It holds that*

$$\|\widetilde{W}_{(i_+)} - \widetilde{W}[B_r]_{(i_+)}\| \leq c_1|v_{i_+}|\beta^D e^{-c_2 r/\beta}, \tag{4.157}$$

where c_1, c_2 are constants of $\mathcal{O}(1)$.

This claim implies that the contribution of all the terms in $\widetilde{W}_{(i_+)}$ which are not included in the

B_r ball around i_+ decays exponentially with r. Hence,

$$\begin{aligned}\|\widetilde{W}_{(i_+)}\|_F = \|\widetilde{W}_{(i_+)} - \widetilde{W}[B_r]_{(i_+)} + \widetilde{W}[B_r]_{(i_+)}\|_F &\geq \|\widetilde{W}[B_r]_{(i_+)}\|_F - \|\widetilde{W}_{(i_+)} - \widetilde{W}[B_r]_{(i_+)}\|_F \\ &\geq \|\widetilde{W}[B_r]_{(i_+)}\|_F - \sqrt{\mathcal{D}_\Lambda}\|\widetilde{W}_{(i_+)} - \widetilde{W}[B_r]_{(i_+)}\| \\ &\geq \|\widetilde{W}[B_r]_{(i_+)}\|_F - \sqrt{\mathcal{D}_\Lambda}c_1|v_{i_+}|\beta^D e^{-c_2 r/\beta}, \quad (4.158)\end{aligned}$$

where we use $\|\widetilde{W}_{(i_+)} - \widetilde{W}[B_r]_{(i_+)}\|_F \leq \sqrt{\mathcal{D}_\Lambda}\|\widetilde{W}_{(i_+)} - \widetilde{W}[B_r]_{(i_+)}\|$ in the second inequality. Second, we consider the approximation of $\widetilde{W}[B_r]$ by $\widetilde{W}[B_r, B_{r'}]$ which are supported on $B_{r'}$:

$$\widetilde{W}[B_r, B_{r'}] := \mathrm{tr}_{B^{\mathrm{c}}_{r'}}(\widetilde{W}[B_r]) \otimes \frac{\mathbb{1}_{B^{\mathrm{c}}_{r'}}}{d^{|B^{\mathrm{c}}_{r'}|}}. \quad (4.159)$$

Because of the quasi-locality of $\widetilde{W}$, we expect $\widetilde{W}[B_r, B_{r'}] \approx \widetilde{W}[B_r]$ for $r' \gg r$. This is shown in the following lemma:

Claim 117. *The norm difference between $\widetilde{W}[B_r, B_{r'}]$ and $\widetilde{W}[B_r]$ is upper-bounded as*

$$\|\widetilde{W}[B_r] - \widetilde{W}[B_r, B_{r'}]\| \leq c_3|v_{i_+}|r^D\beta e^{-c_4|r'-r|/\beta} \quad (4.160)$$

and

$$\|\widetilde{W}[B_r]_{(i_+)} - \widetilde{W}[B_r, B_{r'}]_{(i_+)}\| \leq 2c_3|v_{i_+}|r^D\beta e^{-c_4|r'-r|/\beta}, \quad (4.161)$$

where c_3, c_4 are constants of $\mathcal{O}(1)$.

The claim reduces the inequality (4.158) to

$$\begin{aligned}\|\widetilde{W}_{(i_+)}\|_F &\geq \|\widetilde{W}[B_r]_{(i_+)}\|_F - \sqrt{\mathcal{D}_\Lambda}c_1|v_{i_+}|\beta^D e^{-c_2 r/\beta} \\ &\geq \|\widetilde{W}[B_r, B_{r'}]_{(i_+)}\|_F - \sqrt{\mathcal{D}_\Lambda}\|\widetilde{W}[B_r]_{(i_+)} - \widetilde{W}[B_r, B_{r'}]_{(i_+)}\| - \sqrt{\mathcal{D}_\Lambda}c_1|v_{i_+}|\beta^D e^{-c_2 r/\beta} \\ &\geq \|\widetilde{W}[B_r, B_{r'}]_{(i_+)}\|_F - \sqrt{\mathcal{D}_\Lambda}c_1|v_{i_+}|\beta^D e^{-c_2 r/\beta} - 2\sqrt{\mathcal{D}_\Lambda}c_3|v_{i_+}|r^D\beta e^{-c_4(r'-r)/\beta}. \quad (4.162)\end{aligned}$$

Next, we relate the norm of $\widetilde{W}[B_r, B_{r'}]_{(i_+)}$ to that of $\widetilde{W}[B_r, B_{r'}]$ using Claim 87. By recalling that $\widetilde{W}[B_r, B_{r'}]_{(i_+)}$ is supported on $B_{r'}$, this gives

$$\|\widetilde{W}[B_r, B_{r'}]_{(i_+)}\|_F \geq \frac{1}{|B_{r'}|}\|\widetilde{W}[B_r, B_{r'}]\|_F, \quad (4.163)$$

which reduces the inequality (4.162) to

$$\begin{aligned}\|\widetilde{W}_{(i_+)}\|_F \geq& \frac{1}{|B_{r'}|}\|\widetilde{W}[B_r, B_{r'}]\|_F - \sqrt{\mathcal{D}_\Lambda}c_1|v_{i_+}|\beta^D e^{-c_2 r/\beta} - 2\sqrt{\mathcal{D}_\Lambda}c_3|v_{i_+}|r^D\beta e^{-c_4(r'-r)/\beta} \\ \geq& \frac{1}{|B_{r'}|}\|\widetilde{W}[B_r]\|_F - \sqrt{\mathcal{D}_\Lambda}c_1|v_{i_+}|\beta^D e^{-c_2 r/\beta} - (2 + 1/|B_{r'}|)\sqrt{\mathcal{D}_\Lambda}c_3|v_{i_+}|r^D\beta e^{-c_4(r'-r)/\beta}, \\ & \quad (4.164)\end{aligned}$$

where in the second inequality we apply Claim 117 to $\|\widetilde{W}[B_r, B_{r'}]\|_F$. Finally, we use the lower

bound given in Claim 115 and the inequality $\sum_{i\in B_r} v_i^2 \geq v_{i_+}^2$ (since $i_+ \in B_r$) to obtain

$$\|\widetilde{W}[B_r]\|_F^2 \geq \frac{\mathcal{D}_\Lambda}{c_5[\beta\log(r)+1]^2} v_{i_+}^2.$$

This reduces the inequality (4.164) to the following:

$$\frac{\|\widetilde{W}_{(i_+)}\|_F}{\sqrt{\mathcal{D}_\Lambda}} \geq \frac{|v_{i_+}|}{c_8\,(r')^D\sqrt{c_5}[\beta\log(r)+1]} - c_1|v_{i_+}|\beta^D e^{-c_2 r/\beta} - 3c_3|v_{i_+}|r^D\beta e^{-c_4(r'-r)/\beta},$$

where we used $|B_{r'}| \leq c_8\,(r')^D$, for some constant c_8. By choosing $r' = 2r$ and $r = \Theta(1)\cdot D\beta\log(\beta)+1$, we have

$$\frac{\|\widetilde{W}_{(i_+)}\|_F}{\sqrt{\mathcal{D}_\Lambda}} = \|\widetilde{W}_{(i_+)}\sqrt{\eta}\|_F \geq \frac{c_7|v_{i_+}|}{(\beta\log(\beta)+1)^{D+1}}, \tag{4.165}$$

for some constant c_7. This completes the proof. □

Proof of Claims 116, 117

Proof of Claim 116. Recall that the goal is to prove

$$\|\widetilde{W}_{(i_+)} - \widetilde{W}[B_r]_{(i_+)}\| \leq c_1|v_{i_+}|\beta^D e^{-c_2 r/\beta}$$

for constants c_1, c_2. We start from the integral representation of $\widetilde{W}_{(i_+)}$:

$$\widetilde{W}_{(i_+)} = \widetilde{W} - \int d\mu(U_{i_+})U_{i_+}^\dagger\widetilde{W}U_{i_+}, \tag{4.166}$$

where $\mu(U_{i_+})$ is the Haar measure for unitary operator U_{i_+} which acts on the i_+th site. This yields

$$\widetilde{W}_{(i_+)} - \widetilde{W}[B_r]_{(i_+)} = \widetilde{W}[B_r^{\rm c}] - \int d\mu(U_{i_+})U_{i_+}^\dagger\widetilde{W}[B_r^{\rm c}]U_{i_+}. \tag{4.167}$$

We thus obtain

$$\begin{aligned}
\|\widetilde{W}_{(i_+)} - \widetilde{W}[B_r]_{(i_+)}\| &\leq \sup_{U_{i_+}}\|[U_{i_+},\widetilde{W}[B_r^{\rm c}]]\| \\
&\leq \int_{-\infty}^{\infty} f_\beta(t)\sum_{j\in B_r^{\rm c}}|v_j|\sup_{U_{(i)}}\|[U_{i_+},e^{-iHt}E_j e^{iHt}]\|dt \\
&\leq f\cdot|v_{i_+}|\sum_{j\in B_r^{\rm c}}\int_{-\infty}^{\infty} f_\beta(t)\min(e^{-c(\mathrm{dist}(i_+,j)-m_j-v_{\rm LR}t)},1)dt,
\end{aligned} \tag{4.168}$$

where we use $|v_j| \leq |v_{i_+}|$ and the Lieb-Robinson bound (Fact 81) for the last inequality, and m_j is the radius of the support for E_j (see also Sec. 4.10.4). Because the function $f_\beta(t)$ decays as $e^{-\mathcal{O}(t/\beta)}$

and $\mathrm{dist}(i_+, j) \geq r$ for $j \in B_r^{\mathrm{c}}$, we have

$$|v_{i_+}| \sum_{j \in B_r^{\mathrm{c}}} f e^{cm_j} \int_{-\infty}^{\infty} f_\beta(t) \min(e^{-c(\mathrm{dist}(i_+,j) - v_{\mathrm{LR}} t)}, 1) dt \leq c_1 |v_{i_+}| \beta^D e^{-c_2 r/\beta}. \tag{4.169}$$

This completes the proof. □

Proof of Claim 117. Recall that we wanted to show

$$\|\widetilde{W}[B_r] - \widetilde{W}[B_r, B_{r'}]\| \leq c_3 |v_{i_+}| r^D \beta e^{-c_4 |r'-r|/\beta}.$$

In order to prove this, we also utilize the integral representation of $\widetilde{W}[B_r, B_{r'}]$:

$$\widetilde{W}[B_r, B_{r'}] := \int d\mu(U_{B_{r'}^{\mathrm{c}}}) U_{B_{r'}^{\mathrm{c}}}^\dagger \widetilde{W}[B_r] U_{B_{r'}^{\mathrm{c}}}, \tag{4.170}$$

which yields an upper bound of $\|\widetilde{W}[B_r] - \widetilde{W}[B_r, B_{r'}]\|$ as

$$\|\widetilde{W}[B_r] - \widetilde{W}[B_r, B_{r'}]\| \leq \int d\mu(U_{B_{r'}^{\mathrm{c}}}) \|[\widetilde{W}[B_r], U_{B_{r'}^{\mathrm{c}}}]\|. \tag{4.171}$$

From the definition (4.156) of $\widetilde{W}[B_r]$ and the Lieb-Robinson bound (Fact 81), we obtain

$$\begin{aligned}
\int d\mu(U_{B_{r'}^{\mathrm{c}}}) \|[\widetilde{W}[B_r], U_{B_{r'}^{\mathrm{c}}}]\| &\leq \int d\mu(U_{B_{r'}^{\mathrm{c}}}) \int_{-\infty}^{\infty} f_\beta(t) \sum_{j \in B_r} |v_j| \cdot \|[e^{-iHt} E_j e^{iHt}, U_{B_{r'}^{\mathrm{c}}}]\| \\
&\leq f \max_{j \in B_r} (\mathrm{Supp}(E_j)) |v_{i_+}| \int_{-\infty}^{\infty} f_\beta(t) \sum_{j \in B_r} \min(e^{-c(r'-r-m_j - v_{\mathrm{LR}} t)}, 1) dt \\
&\leq c_3' |v_{i_+}| \cdot |B_r| \cdot \beta e^{-c_4 |r'-r|/\beta},
\end{aligned} \tag{4.172}$$

with c_3' a constant of $\mathcal{O}(1)$, where $\mathrm{Supp}(E_j) \propto m_j^D = \mathcal{O}(1)$ is the support of E_j. Since $|B_r| \propto r^D$, we obtain the main inequality (4.160). Now, since

$$\widetilde{W}[B_r]_{(i_+)} = \widetilde{W}[B_r] - \int d\mu(U_{i_+}) U_{i_+}^\dagger \widetilde{W}[B_r] U_{i_+}$$

and

$$\widetilde{W}[B_r, B_{r'}]_{(i_+)} = \widetilde{W}[B_r, B_{r'}] - \int d\mu(U_{i_+}) U_{i_+}^\dagger \widetilde{W}[B_r, B_{r'}] U_{i_+},$$

we obtain the second inequality (4.161) due to

$$\|\widetilde{W}[B_r]_{(i_+)} - \widetilde{W}[B_r, B_{r'}]_{(i_+)}\| \leq 2 \|\widetilde{W}[B_r] - \widetilde{W}[B_r, B_{r'}]\|.$$

This completes the proof. □

Chapter 5

Thermal phase transition versus computational hardness

Chapter summary: In this work, we study two related problems regarding quantum many-body systems: developing approximation algorithms for the partition function of these systems, and characterization of the thermal phase transition.

(1) We extend the scope of a recent approach due to Barvinok for solving classical counting problems [Bar16a] to quantum many-body systems. This allows us to find a deterministic quasi-polynomial time classical algorithm that estimates the partition function of quantum systems at temperatures above the phase transition point. We also find an algorithm for the anisotropic XXZ model in the *ferromagnetic* regime at any temperature over arbitrary graphs. Previously, a randomized algorithm was known only for the ferromagnetic XY model [BG17] using different techniques.

(2) We show that the partition function of a geometrically-local Hamiltonian does not have any complex zeros near the real axis above a constant temperature which depends only on the geometric properties of the Hamiltonian.

(3) We prove that in a system of n qudits at temperatures above the phase transition point, where the complex zeros of the partition function are far from the real axis, the correlations between two observables whose distance is $\Omega(\log(n))$ decay exponentially. We can improve the factor of $\log(n)$ to a constant when the Hamiltonian has commuting terms (which includes classical Hamiltonians) or is on a 1D chain. These results build on a work of Dobrushin and Shlosman on translationally-invariant classical spin systems [DS87].

This chapter is based on:

[HMS20] Aram W. Harrow, Saeed Mehraban, and Mehdi Soleimanifar. Classical algorithms, correlation decay, and complex zeros of partition functions of quantum many-body systems. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC), pages 378–386*, 2020.

5.1 Introduction

In thermal equilibrium, a quantum system characterized by a local Hamiltonian H is in the Gibbs (or thermal) state $\rho = \exp(-\beta H)/Z_\beta(H)$, where β is the inverse of temperature and $Z_\beta(H) =$

$\mathrm{tr}[\exp(-\beta H)]$ is the *partition function* of the system. Many useful statistical properties of the system including the free energy and entropy can be obtained from the partition function and its derivatives. However, exactly evaluating the partition function is known to be #P-hard. This motivates devising efficient algorithms that *approximately* evaluate this quantity.

Our starting point for finding such approximation algorithms is based on the observation that the phenomenon of the thermal phase transition is an obstacle for finding *efficient* algorithms. Consider a quantum many-body system that consists of n qudits interacting according to a local Hamiltonian H. As the temperature of this system increases, meaning $\beta \to 0$, the Gibbs state ρ approaches the maximally mixed state $\mathbb{1}/d^n$. Thus, in this case, finding the partition function is trivial since $Z_{\beta=0}(H) = d^n$. On the other hand, this problem becomes significantly harder at lower temperatures. In particular, as $\beta \to \infty$, the Gibbs state approaches the ground space of the Hamiltonian H and the free energy $F_\beta(H) = -1/\beta \log Z_\beta(H)$ approaches the ground energy which is known to be QMA-hard to estimate. Hence, we see that the computational hardness of estimating the partition function (or equivalently the free energy) depends on the inverse temperature β and goes through a transition from being trivial to QMA-hard as β increases.

In statistical physics, however, another transition occurs as β increases, namely, the transition in the *phase* of the system. At the thermal phase transition point, certain physical properties of the system undergo an abrupt change. Does the computational hardness of estimating the partition function also undergo an abrupt change at the same transition point? This question has been studied in *classical Ising* or *hard-core model*, and the answer is known to be affirmative. For these systems, there are efficient algorithms for estimating the partition function when $\beta < \beta_c$ [Wei06, SST14] whereas by a result of Sly and Sun [SS12, Sly10] the same problem is NP-hard under a randomized reduction for $\beta > \beta_c$.

Hence, it appears that the thermal phase transition poses a barrier to obtaining efficient algorithms, and we need a framework for characterizing this phenomenon. There are at least two methods for such purpose. One, which is the basis of our algorithm, stems from analyzing the locus of the complex zeros of the partition function. Another seemingly different method involves the decay of correlations in the Gibbs state of the system. In this work, we study the interface between these two methods and their algorithmic implications for interacting quantum systems.

5.1.1 Our main results

Approximation algorithm from the complex zeros of quantum partition functions

In general, the partition function can be written as $Z_\beta(H) = \sum_k \exp(-\beta E_k)$, where each E_k is an eigenvalue of the Hamiltonian H. If β is real, the terms $\exp(-\beta E_k)$ are all strictly positive, and hence the partition function $Z_\beta(H)$ is strictly positive itself. However, this changes when β is allowed to be complex. In that case, the terms $\exp(-\beta E_k)$ acquire complex phases that when added together might cancel each other and make the partition function zero. We call the solutions of $Z_\beta(H) = 0$ for $\beta \in \mathbb{C}$, the *complex zeros* of the partition function.

The significance of these zeros becomes more clear if one looks at the free energy $F_\beta(H)$. The zeros of $Z_\beta(H)$ are the *singularities* of $\log Z_\beta(H) = -\beta F_\beta(H)$. Since $Z_\beta(H) \neq 0$ when β is real, we see that all these singularities are located in the complex plane and the free energy is analytic near the real axis. As the number of particles n grows, the number and location of these points change. Perhaps rather surprisingly, some of these singularities approach the real axis in the limit of a large number of particles, $n \to \infty$. The first point on the positive real axis where these zeros converge in

the large n limit is called the critical inverse temperature and denoted by β_c (see Figure 5-1). This critical temperature separates different phases of matter and important quantities such as the free energy become non-analytic in the vicinity of β_c. The study of these complex zeros in connection with the phase transition phenomenon in classical Ising models was initiated by Lee and Yang [LY52] and later extended by Fisher [Fis65]. This approach is one of the few rigorous methods available in the theory of phase transitions.

One can go beyond partition functions and consider complex roots of high-degree polynomials that appear in combinatorics such as estimating the permanent of a matrix. Recently, there has been a surge of interest in studying these complex zeros in theoretical computer science due to their algorithmic applications. In particular, a new approach based on the truncated Taylor expansion introduced by Barvinok [Bar16a] directly connects the locus of the complex zeros to approximation algorithms for counting problems. In this work, we extend the scope of this method by applying it to quantum many-body systems.

We first state the condition on the location of zeros that we use in our approximation algorithm. Under this condition, it is guaranteed that the inverse temperature β at which the partition function is estimated is connected to $\beta = 0$ by a path in the complex plane that avoids the complex zeros along its way with a significant margin. Even though this algorithm works for any such path, we restrict our attention to the physically-relevant case when this zero-free region contains the real β-axis. Hence, we define:

Definition 118. *The δ-neighborhood of the interval $[0, \beta]$ for some $\beta \in \mathbb{R}^+$ is a region of the complex plane defined as $\Omega_{\delta,\beta} = \{z \in \mathbb{C} : \exists z' \in [0, \beta], |z - z'| \leq \delta\}$ (see Figure 5-1 for an example of such a region).*

Definition 119 (Analyticity condition, informal version of Conditions 1' and 1)**.** *For a system of n particles with a local Hamiltonian H, we define:*

1. *A δ-neighborhood $\Omega_{\delta,\beta}$ of the interval $[0, \beta]$ (see Definition 118) is called zero-free if δ is some constant and $\forall \beta' \in \Omega_{\delta,\beta}$ the partition function $Z_{\beta'}(H) \neq 0$ and moreover, $|\log Z_{\beta'}(H)| \leq O(n)$.*

2. *Equivalently, the free energy $F_\beta(H)$ is called δ-analytic along $[0, \beta]$ if $\Omega_{\delta,\beta}$ is a zero-free region.*

While the condition $|\log Z_\beta(H)| \leq O(n)$ is satisfied on the real β axis, it may not hold in the complex plane close to the zeros of $Z_\beta(H)$. If the partition function is a polynomial, which is the case for the classical Ising model, this condition follows when β is constantly far from the complex zeros. In general, though, we need to include this as an independent assumption.

We now state our first result which shows that the framework of the truncated Taylor expansion [Bar16a] can be naturally extended to also estimate quantum partition functions.

Theorem 120 (Informal version of Theorem 136)**.** *There is a deterministic classical algorithm that takes a local Hamiltonian H and a number ε as inputs, runs in time $n^{O(\log(n/\varepsilon))}$, and outputs a value within ε-multiplicative error of the partition function $Z_\beta(H)$ at inverse temperature β as long as the free energy is δ-analytic along the $[0, \beta]$ line (see Definition 119) for some $\delta = O(1)$.*

Given the result of Theorem 120, the main challenge is to find the zero-free region or the critical point β_c for a Hamiltonian H. This can be achieved in certain systems such as the *classical* Ising model by using their specific structure [LSS19a, BS17a]. In general, though, it is a hard problem to exactly find this region given an arbitrary Hamiltonian. One can compare this with when a

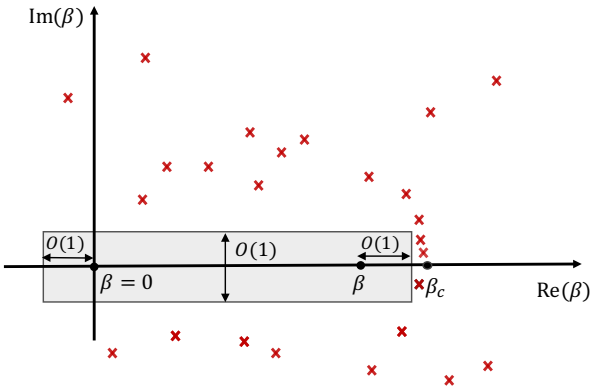


Figure 5-1: The location of complex zeros of the partition function, the critical point β_c, and the zero-free region near the real axis (as in Definition 119). The free energy is analytic in this region.

1D quantum system is assumed to have a constant spectral gap. Under this condition, there is an efficient algorithm for estimating the ground energy. However, it has been shown that validating this condition, i.e. determining if a Hamiltonian is gapped or not, is undecidable in the worst case [CPGW15].

In our next result, we show that for any geometrically-local Hamiltonian H, there exists a zero-free *disk* of radius β_0 around $\beta = 0$ for some constant $\beta_0 \leq \beta_c$ which depends only on the geometric parameters of H. Here, we say a Hamiltonian is geometrically local if the local terms in H act on neighboring qudits that are located on a D-dimensional lattice $\Lambda \subset \mathbb{Z}^D$.

Theorem 121 (Informal version of Theorem 137)**.** *There exists a real constant β_0 such that for all $\beta \in \mathbb{C}$ with $|\beta| \leq \beta_0$, the partition function $Z_\beta(\Lambda)$ of a geometrically-local Hamiltonian H does not vanish, and furthermore, $\big|\log|Z_\beta(\Lambda)|\big| \leq O(n)$.*

As mentioned earlier, without focusing on a specific family of Hamiltonians, improving the zero-free region in Theorem 121 seems implausible. Hence, an alternative approach is to show the absence of zeros at a given β by assuming the validity of other conditions such as the decay of correlations. This along with other results in this direction is the subject of the next section.

The decay of correlations in the Gibbs state

Another signature of the thermal phase transition is the appearance of long-range order in the system. In the example of a magnetic system, below the phase transition in the ferromagnetic phase (also called the *ordered* phase), distant spins are correlated and point in the same direction, whereas in the paramagnetic phase (also known as the *disordered* phase), the correlations between disjoint parts of the system decay exponentially with their distance. More precisely, we define the exponential decay of correlations as

Definition 122 (Correlation decay condition, informal version of Condition 2)**.** *The Gibbs state $\rho_\beta(H)$ of a geometrically-local Hamiltonian H at inverse temperature β exhibits an exponential decay of correlations if for any two disjoint observables O_1 and O_2 and any region B such that $\mathrm{supp}(O_1), \mathrm{supp}(O_2) \subset B$ there exist constants ξ and c such that*

$$\big|\operatorname{tr}[\rho_\beta(H)O_1O_2] - \operatorname{tr}[\rho_\beta(H)O_1]\operatorname{tr}[\rho_\beta(H)O_2]\big| \leq c|B|\|O_1\|\|O_2\|e^{-\mathrm{dist}(O_1,O_2)/\xi}. \tag{5.1}$$

What is the relation between the decay of correlations and the complex zeros of the partition function? Note that the former involves correlations in the system at a *real* temperature while the latter concerns the *complex* temperature features of the partition function. Could it be that these two apparently distinct characterizations are indeed equivalent?

Besides its physical significance, the correlation decay property is also crucially used in many approximation algorithms both in classical [Wei06] and quantum [BK16] settings. Hence, we see that there are different approaches for estimating the partition function: using the absence of complex zeros versus relying on the decay of correlations. A third approach is to use Markov chain Monte Carlo (MCMC) sampling algorithms for this purpose. An important question is whether the range of temperatures that these widely different approaches cover is the same.

The equivalence between the exponential decay of correlations (fast mixing in space) and fast convergence of MCMC algorithms (fast mixing in time) has been established before for classical systems [DSVW04, Wei04]. Here, we consider the equivalence between the absence of complex zeros and the exponential decay of correlations. This question has been recently raised in [LSS19a] where one direction of this equivalence was proved for the special case of the classical Ising model [LSS19a].

We study this question in the context of quantum many-body systems, but as a special case, our results apply to classical systems. We build on a work of Dobrushin and Shlosman [DS87] who proved this equivalence for arbitrary translationally-invariant classical systems. We are not aware of any application of the proof techniques in this paper or [DS87] that appear in the recent results on counting problems. Hence, the methods developed here might be of independent interests.

In our next result, we show that the absence of complex zeros around some real β implies the exponential decay of correlations at that β.

Theorem 123 (Informal version of results in Section 5.5)**.** *Let $\rho_\beta(H)$ be the Gibbs state of a geometrically-local Hamiltonian at inverse temperature β in the zero-free region $\Omega_{\delta,\beta}$ (given in Definition 119) for some constant δ . This state has the decay of correlation property as in Definition 122 in any of the following cases:*

(i.) The distance between the observables O_1 and O_2 is at least $\Omega(\log n)$[1],

(ii.) The Hamiltonian H is the sum of mutually commuting local terms, or

(iii.) The Hamiltonian H is defined on a $1D$ chain.

The class of commuting Hamiltonians includes important examples such as stabilizer Hamiltonians like the Toric code, Color code, or Levin-Wen model [LW05].

Proving the converse of Theorem 123 turns out to be more challenging. Nevertheless, we can give evidence for this direction by generalizing the result of [DS87] to classical systems that are not translationally invariant, and also quantizing certain steps in the proof. By fully establishing this equivalence, one could rigorously confirm the physical intuition that a quantum system enters the disordered phase at the point where the free energy becomes analytic.

Theorem 124 (Informal version of Theorem 153)**.** *Let H be a geometrically-local Hamiltonian of a classical spin system, i.e. the local terms H_i are all diagonal in the same product basis. For this system, the exponential decay of correlations given in Definition 122 implies the absence of zeros near the real axis as in Definition 119.*

[1] In other words, we bound the correlation between O_1 and O_2 by $cn\|O_1\|\|O_2\|e^{-\mathrm{dist}(O_1,O_2)/\xi}$

Two-local Hamiltonians and Lee-Yang zeros

For our last result, we switch gears and focus on a specific family of 2-local Hamiltonians. We again use the idea of extrapolation, but this time, our extrapolation parameter instead of β is the strength of the external magnetic field applied to the system in the z-direction. The physical motivation is that when the system is subject to a large external field in a specific direction (the z-direction in our case), all spins align themselves in that direction, and estimating the properties of the system becomes trivial. On the other hand, as we move to smaller fields, the other interaction terms between the particles gain significance, making the problem non-trivial. Our result is an approximation algorithm for the quantum XXZ model with the following Hamiltonian:

Definition 125. *The anisotropic* XXZ *Hamiltonian on an interaction graph $G = (V, E)$ is given by*

$$H(\mu) = -\sum_{(i,j)\in E} \left(J_{ij}(X_iX_j + Y_iY_j) + J_{ij}^{zz} Z_iZ_j\right) - \mu \sum_{i\in V} Z_i. \tag{5.2}$$

We find an approximation algorithm for this model. This is stated in the following theorem. This model is outside the family of ferromagnetic systems considered previously in [BG17] and to the best of our knowledge no efficient algorithm was previously known for estimating its partition function.

Theorem 126 (Informal version of Theorem 169)**.** *There is a deterministic algorithm that runs in $n^{O(\log(n/\varepsilon))}$ time and outputs an ε-multiplicative approximation to the partition function of the anisotropic* XXZ *model (see Definition 125) in the ferromagnetic regime, i.e. when $J_{ij}^{zz} \geq |J_{ij}|$, and μ is an arbitrary constant.*

5.1.2 Technical contribution and sketch of our proofs

Sketch of the proof for Theorem 123 The technique used in the proof of Theorem 123 is inspired by the extrapolation idea of Theorem 120 and also the proof of the similar statement for the classical systems due to [DS87].

One major issue that appears in the proof of this Theorem and Theorem 124 is the handling of entangled boundary conditions. To address this, we consider the Gibbs state after a subset of spins have been measured. This means we work with partition functions of the form $\mathrm{tr}[\exp(-\beta H)N]$ for some positive semi-definite operator N. We then define a function $f(\beta)$ that measures the correlation between disjoint observables O_1 and O_2. This function is defined in a slightly different way than the *covariance* form in (5.1) and is tuned to have specific properties. In particular, we show that at $\beta = 0$, the value of this function is zero, i.e. $f(0) = 0$. This is expected intuitively since the system is in the maximally mixed state at $\beta = 0$ and particles are distributed independently at random. However, we further show that the low order derivatives of this function up to $O(\mathrm{dist}(O_1, O_2))$ are all zero at $\beta = 0$, i.e.

$$\frac{d^k f(\beta)}{d\beta^k}\Big|_{\beta=0} = 0, \quad \text{for } k = 0, 1, \ldots, O(\mathrm{dist}(O_1, O_2)). \tag{5.3}$$

Hence, this function looks very flat around the origin. Additionally, we prove that $f(\beta)$ is an analytic function in the zero-free region. Finally, we show that this together with the constraints on

the derivatives imply that the value of $f(\beta)$, which shows how correlated O_1 and O_2 are, remains exponentially small when moving from the origin to a constant β.

This gives us an upper bound $\propto n \exp(-\mathrm{dist}(O_1, O_2)/\xi)$ on the amount of correlation. The extra factor of n makes this bound exponentially small when $\mathrm{dist}(O_1, O_2) = \Omega(\log n)$.

Remark 127. *Even with the extra factor of n, our bound remains useful for algorithmic applications such as in [BK16]. There one needs to split the system into computationally tractable smaller pieces and solve the problem for those pieces locally. The error of this strategy can be bounded using the exponential decay of correlations. To keep this error less than $1/\operatorname{poly}(n)$, one needs to choose the distances to be $O(\log n)$ which is the regime that our result covers.*

In classical systems, one can remove the constraint $\mathrm{dist}(O_1, O_2) = \Omega(\log n)$ by using the Markov property of the Gibbs states. This property is known not to (exactly) hold in the quantum case. We can get around this issue in certain instances. This includes when the Hamiltonian consists of commuting terms or when it is defined on a 1D chain. In both cases, using either the commutativity of local terms or quantum belief propagation [Has07b] (refer to Proposition 131 in the body for the precise statement), we show that by removing the interaction terms acting on particles that are far from the observables O_1 and O_2, the correlations between O_1 and O_2 do not change by much. Hence, the system size reduces to the number of particles in the vicinity of the two observables. This number replaces the prefactor n we had before and is negligible compared to the exponential factor $\exp(-\mathrm{dist}(O_1, O_2)/\xi)$. Thus, for these systems, the decay of correlations holds even when $\mathrm{dist}(O_1, O_2)$ is a constant. In higher dimensions, using quantum belief propagation results in an error proportional to the size of the boundary which restricts its application for our purpose.

Sketch of the proofs for Theorem 121 and Theorem 124 We first introduce a core idea which plays a central role in the proofs of both Theorem 137 and Theorem 123. For ease of notation, we denote the partition function of a geometrically-local Hamiltonian H defined over a D-dimensional lattice $\Lambda \subset \mathbb{Z}^D$ by $Z_\beta(\Lambda)$. The particles are located on the vertices of this lattice.

In Theorem 121, our goal is to show that $Z_\beta(\Lambda) \neq 0$ inside a disk of radius β_0, i.e. for $\beta \in \mathbb{C}$ where $|\beta| \leq \beta_0$ for some constant β_0. We consider a series of sublattices $\emptyset = \Lambda_0 \subset \Lambda_1 \subset \Lambda_2 \subset \cdots \subset \Lambda_n = \Lambda$ such that each sublattice Λ_i has one fewer vertex than Λ_{i+1}. By convention, we let $Z_\beta(\emptyset) = 1$. As long as the sublattice Λ_i has only a constant number of particles, we can always ensure $Z_\beta(\Lambda_i) \neq 0$ by choosing β to be a sufficiently small constant. One might worry that by adding more particles, the partition function vanishes.

Our main contribution is to prove this does not happen. We do so by showing that the partition function after involving new particles does not become smaller than a constant fraction of the partition function before adding the particles. In other words, we show there exists a constant $c > 1$ such that

$$|Z_\beta(\Lambda_{i+1})| \geq c^{-1} |Z_\beta(\Lambda_i)|, \quad i \in \{1, 2, \ldots, n-1\}. \tag{5.4}$$

By repeatedly applying this bound, we obtain the following exponentially small (yet sufficiently large for our purposes) lower bound on the partition function of the whole system

$$|Z_\beta(\Lambda)| \geq c^{-n} |Z_\beta(\Lambda_1)|. \tag{5.5}$$

This leads to the bound given in Theorem 121. This lower bound is obtained using a method

known as the *cluster expansion*. Since the local terms in the Hamiltonian of a quantum system do not necessarily commute, applying this method becomes quite technical. The cluster expansion that we use is due to Hastings [Has06, KGK+14], which represents the operator $\exp(H)$ as sum of products of local terms H_i. This allows us to express $Z_\beta(\Lambda_{i+1})$ in terms of $Z_\beta(\Lambda_i)$ plus some small correction terms that account for the interaction terms acting on the added particle. This by itself does not lead to a bound on the partition function. Our technical contribution is to use an inductive proof to connect such a decomposition to the lower bound (5.4) (see the proof of Theorem 137 in the body for details).

A similar strategy is used in the proof of Theorem 124 which is based on the result of [DS87] for translationally-invariant classical systems. We essentially show a similar bound to (5.4) on how much the partition function can shrink after adding new particles. Here, instead of cluster expansions, we use the exponential decay of correlations to show such a lower bound. However, notice that the decay of correlations is a property of the system at a *real* β, whereas we want to bound the absolute value of the partition function at some *complex* β. There are multiple steps in the proof before we can get around this issue.

One crucial step is to reduce the proof of the analyticity of the free energy to a condition that roughly speaking (see Proposition 155 for the details) states that changing the value of a spin in the system only causes a small relative change in the partition function of the system even for complex β [DS87]. This is proved by isolating the effect of this spin flip from the rest of the system using the decay of correlations. This requires removing the imaginary part of β for all the interactions in the vicinity of the flipped spin and bounding the resulting error.

This overall approach involves a subtle use of the boundary conditions in the spin system. In the quantum case, this means applying local projectors (or more generally a positive operator) to the Gibbs state before evaluating the partition function. These projectors can in general be entangled which makes using this proof technique more challenging for quantum systems.

Sketch of the proof for Theorem 120 The basis of our algorithm in Theorem 120 is the following observation. It is computationally easy to find the partition function and its derivatives at $\beta = 0$. Note that in a system of n qudits, $Z_{\beta=0}(H) = d^n$ and its derivatives are

$$\frac{d^k Z_\beta(H)}{d\beta^k}\Big|_{\beta=0} = (-1)^k \operatorname{tr}[H^k]. \tag{5.6}$$

Since the local Hamiltonian H equals $\sum_{i=1}^m H_i$ for some $m = \text{poly}(n)$, its kth power H^k is also the sum of $n^{O(k)}$ many local terms, i.e.

$$H^k = \sum_{j=1}^{n^{O(k)}} H_j^{(k)}, \tag{5.7}$$

where $H_j^{(k)}$ is a product of k local terms H_i. Each of the new terms $H_j^{(k)}$ acts on a region that is at most k times larger than the support of the original terms H_i which is still some constant. We can find $\operatorname{tr}[H^k]$ by adding $n^{O(k)}$ many terms like $\operatorname{tr}[H_j^{(k)}]$, which allows us to compute the derivatives (5.6) in time bounded by $n^{O(k)}$.

How can the solution at $\beta = 0$ be used to estimate the one at some non-zero β? We use a technique due to Barvinok [Bar16b, Bar15] that has been applied to similar counting problems.

The idea is to *extrapolate* this solution at $\beta = 0$ to find $Z_\beta(H)$ at some non-zero β where the problem is non-trivial. The extrapolation is done simply by using a *truncated Taylor expansion* of $\log Z_\beta(H)$ at $\beta = 0$. Since our goal is to find the partition function with some ε-multiplicative error, it is sufficient to estimate $\log Z_\beta(H)$ within ε-additive error.

The main barrier to the reliability of this algorithm is establishing the fast convergence of the Taylor expansion. Such a Taylor expansion is only valid when $\log Z_\beta(H)$ remains a complex-analytic function, meaning the extrapolation is done along a path contained in the zero-free region. This is precisely the condition stated in Definition 119. Under this assumption, the Taylor theorem along with the bound $|\log Z_\beta(H)| \leq O(n)$ that we get from being in the zero-free region give

$$\left| \log Z_\beta(H) - \sum_{k=0}^{K-1} \frac{1}{k!} \frac{d^k \log Z_\beta(H)}{d\beta^k}\Big|_{\beta=0} \right| \leq c_1 n e^{-c_2 K} \tag{5.8}$$

for some constants c_1, c_2 (see Proposition 135 in the body for details). The running time of computing the terms in this expansion is dominated by that of finding the derivatives which, as mentioned earlier, takes time $n^{O(K)}$. To get an additive error of ε for $\log Z_\beta(H)$, it suffices to choose $K = O(\log(n/\varepsilon))$ resulting in a quasi-polynomial time algorithm.

The running time of this algorithm depends exponentially on the distance between the zeros and the extrapolation path. This allows us to clearly see why our algorithm fails beyond the phase transition point. If we try to extrapolate to $\beta \geq \beta_c$, we need to find a zero-free region that avoids the "armor" of zeros that are concentrated around the real axis at β_c. This results in a zero-free region with a vanishing width. Hence, the running time blows up, which matches our expectation from the NP hardness result above β_c [SS12].

Sketch of the proof for Theorem 126 Thus far we have only considered complex zeros of the partition function as a function of β. These are often called Fisher zeros [Fis65]. One can, however, fix β and consider the partition function as a function of other parameters in the Hamiltonian. When that parameter is the strength of the external magnetic field denoted by μ, these zeros are called Lee-Yang zeros [LY52]. In a pioneering result, Lee and Yang showed that for ferromagnetic systems, the locus of these zeros can be exactly determined and they are all on the imaginary axis in the complex μ-plane.

A generalization of this theorem has been proved for a class of 2-local quantum systems including the anisotropic Heisenberg model [SF71]. The result follows by mapping the quantum system to a classical spin system and applying a Lee-Yang type argument to the classical model.

Knowing the location of the complex zeros, we use the extrapolation algorithm to estimate the solution at a constant μ by finding the low-order derivatives of the partition function at $\mu = 0$. We can apply this to the quantum XXZ model given in (5.2).

5.1.3 Previous work

Classical statistical physics and combinatorial counting

The Gibbs distribution and partition function appear naturally in combinatorial optimization, statistical physics, and machine learning. In particular, the classical *Ising model* has been studied extensively within these areas. These studies have cultivated in various probabilistic and deterministic approximation algorithms for this model and its variants. In the following, we summarize some

of these results.

Most notable and the first rigorously proven efficient algorithm for the Ising model is the result of Jerrum and Sinclair [JS93] that uses a Markov chain Monte Carlo (MCMC) sampling algorithm to estimate the partition function in the ferromagnetic regime on arbitrary graphs. More generally, it has been shown that one can set up Markov chains for sampling from the Gibbs distribution that mix rapidly *if* and *only if* the correlations decay exponentially. This is known as the equivalence of mixing in time and mixing in space [DSVW04, Wei04].

Another approach uses the decay of correlations in the Gibbs distribution. This property essentially allows one to decompose the interaction graph of the system into smaller computationally tractable pieces, and then combine the results of the computation on those pieces to find the overall partition function. In contrast to the MCMC approach, algorithms based on the decay of correlations can be deterministic. This approach, for instance, has lead to efficient deterministic algorithms for the hard-core model up to the hardness threshold [Wei06] and the antiferromagnetic Ising model [SST14].

There is a recent conceptually different approach to estimating the partition function, which is the basis of this work. This approach views the partition function as a high-dimensional polynomial and uses the truncated Taylor expansion to extend the solution at a computationally easy point to a non-trivial regime of parameters. Since its introduction [Bar16a], this method has been used to obtain deterministic algorithms for various interesting problems such as the ferromagnetic and antiferromagnetic Ising models [LSS19b, PR18] on bounded graphs.

The question of the relation between the analyticity of the free energy and the decay of correlations was recently considered in [LSS19a] where the authors show that the correlation decay implies the absence of Fisher zeros near the real axis. A more general statement has been proved by Dobrushin and Shlosman [DS87] for translationally-invariant classical systems.

Quantum many-body systems

The problem of estimating the partition function and correlation decay in quantum systems has also been studied in the past. We review some of these results here.

There are various results (e.g., [PW09, CS17]) that estimate the partition function by sampling from the Gibbs state using a quantum computer (also known as quantum Gibbs sampling). The best known bound on the running time of these algorithms is exponential in the number of particles. This running time can be reduced if we assume other conditions. For example, [KBa16] shows that a strong form of the decay of correlations implies an efficient quantum Gibbs sampler for commuting Hamiltonians. If in addition to the decay of correlations, we add the decay of quantum conditional mutual information, then this result can be extended to non-commuting Hamiltonians [BK16]. Turning these quantum algorithms into classical ones results in an $n^{\text{polylog}(n)}$ running time. Although we cannot directly compare these results with our algorithm due to different conditions that are imposed, the $n^{O(\log n)}$ running time that we achieve outperforms that of these algorithms.

Considering the success of approximation schemes for the classical statistical problems, it is desirable to import those results to evaluate the thermal properties of interacting quantum many-body systems. This indeed can be done for some models like the quantum transverse field Ising model [Bra15] or the quantum XY model [BG17] in the ferromagnetic regime using what is called the *quantum-to-classical mapping*. However, this approach only works for a restricted set of Hamiltonians known as stoquastic Hamiltonians in which all off-diagonal matrix elements are real and non-positive. This set is known to be restricted in many ways. For example, estimating the ground

state energy of a general quantum Hamiltonian is QMA-complete but the problem is in AM when we restrict to stoquastic Hamiltonians.

Establishing the decay of correlations in the Gibbs state has also been studied in quantum settings. In particular, it has been shown that the Gibbs state has this property in the 1D translationally invariant case [Ara69] or above some constant temperature in higher dimensions [KGK+14]. Thus, in these regimes, there exist efficient representations for the state of the system using a tensor network ansatz like matrix product states or projected entanglement pair states [Has06, KGK+14, MSVC15]. However, this does not necessarily imply an efficient algorithm that finds and faithfully manipulates these tensor networks.

The decay of conditional mutual information is another property of the Gibbs state that has been rigorously proved for 1D systems [KBa19b] and conjectured for higher dimensions. This result has been used to find algorithmic schemes for preparing the Gibbs state on a quantum computer [BK16] or estimating the free energy in 1D [Kim17, KS18]. A recent result of [KKBa20] uses cluster expansions along with a technique very similar to the one we use in Theorem 123 (i.e. showing the low-order derivatives of the correlation function are zero) to establish the decay of conditional mutual information above some constant temperature.

5.1.4 Discussion and open questions

Our work raises many questions that we leave for future work. Here we mention some of them.

1. Perhaps the most immediate problem is to fully establish (or refute) the connection between the decay of correlations and the absence of zeros. There are at least two directions to pursue.

   (a) It would be interesting to prove the exponential decay of correlations in the zero-free region of non-commuting Hamiltonians in higher dimensions. Currently we can only show this when the distance of the observables is $\Omega(\log n)$. It seems for this to work, the region of applicability of certain tools such as quantum belief propagation needs to be extended to the complex regime.

   (b) Establishing the absence of zeros in quantum systems when the correlations decay exponentially is also open. A first step might be to prove this for commuting Hamiltonians or 1D chains. In Section 5.6, we have already extended some parts of the proof of this statement for the classical systems to commuting Hamiltonians, but it seems to complete the proof, a more careful analysis of the entangled boundary conditions is required.

2. While we focus on the covariance form of the correlations (5.1), one can also consider quantum conditional mutual information (qCMI) as a measure of correlations. Using the absence of zeros to prove the decay of qCMI is another interesting question. This would extend the result of [KKBa20] to lower temperatures down to the phase transition point. Since the approach of [KKBa20] resembles some of the techniques we use, this looks like a promising direction.

3. Is there some range of temperatures or Hamiltonian parameters that a quantum computer cannot efficiently sample from the Gibbs state but the extrapolation technique still works? At least, when the parameter of interest is temperature, this depends on the fate of the previous questions we mentioned, i.e. showing that the decay of correlations and qCMI are necessary for the absence of zeros. The result of [BK16] implies an efficient quantum sampler under the

same conditions. Are there other parameters besides temperature for which one can show a separation between these notions?

4. Is it possible to improve the lower bound we obtained for the critical point β_c in Theorem 121 without using other conditions such as the decay of correlations? In general, what is the computational hardness of determining the thermal phase transition point β_c?

5. Can the running time of our algorithm be improved for specific systems to polynomial time? This has been achieved for the classical Ising model [LSS19b, PR18] by relating the derivatives of the partition function to combinatorial objects that can be efficiently counted.

6. Can we use the extrapolation idea to avoid the *sign problem*? The easy regime, which includes the starting point of the extrapolation, could be a regime of parameters where the Hamiltonian is sign-free and MCMC algorithms yield a good estimate, whereas the end point is where the sign problem exists. A candidate parameter for extrapolation is the chemical potential. There are important physical systems such as lattice gauge theories for which at zero chemical potential the partition function is sign-free while there is a severe sign problem for non-zero chemical potentials.

7. Barvinok's approach has been used to obtain approximation algorithms for other problems related to quantum computing [EM17, MB19, BGM21]. Are there other relevent applications for this method?

5.2 Preliminaries and notation

5.2.1 Local and geometrically-local Hamiltonians

Consider a D-dimensional lattice $\Lambda \subset \mathbb{Z}^D$ containing n sites with a d-dimensional particle (qudit) on each site. The Hilbert space is $\mathcal{H} = \bigotimes_{i\in\Lambda} \mathcal{H}_i$ where $\mathcal{H}_i$ is the local Hilbert space of site i. For a region $A \subseteq \Lambda$, we denote its size by $|A|$ and its complement by $\bar{A}$. The diameter of A is defined to be $\mathrm{diam}(A) = \sup\{\mathrm{dist}(x, y) : x, y \in A\}$. The interaction of these particles is described by a local Hamiltonian H that has the following form:

$$H = \sum_{X\subset\Lambda} H_X. \tag{5.9}$$

Each term H_X is a Hermitian operator with operator norm at most h that is acting non-trivially only on the sites in X. We denote this by writing $\mathrm{supp}(H_X) = X$. The local terms H_X do not necessarily commute with each other. Similarly, we define $H_A = \sum_{X\subseteq A} H_X$ to be the Hamiltonian restricted to a region $A \subseteq \Lambda$. We denote the number of local terms in the Hamiltonian by m and often also write $H = \sum_{i=1}^{m} H_i$. The 1-norm of an operator O is denoted by $\|O\|_1$ and its operator norm by $\|O\|$.

In order to impose geometric locality on the interactions between the particles, we consider the interactions that satisfy the following condition.

Definition 128 (Geometrically-local Hamiltonians)**.** *A Hamiltonian $H = \sum_{X\subset\Lambda} H_X$ such that $|\mathrm{supp}(H_X)| = 0$ when $\mathrm{diam}(X) > R$ or $|X| > \kappa$ is called a (κ, R)-local Hamiltonian. We call κ the locality and R the range of H. We use the words geometrically-local and (κ, R)-local interchangeably when κ, R are kept constant.*

This should be contrasted with the case where $|\operatorname{supp}(H_X)| = 0$ when $|X| > \kappa$ but there is no restriction on $\operatorname{diam}(X)$. In order to distinguish between these two, we use the terms *geometrically-local* versus *local* throughout this paper. We also focus mostly on geometrically-local Hamiltonians with a *finite range* R, but most of our results also apply to Hamiltonians with interactions that decay fast enough, for example, with some exponential rate.

Remark 129. *In general, the locality κ of a geometrically-local Hamiltonian on a D-dimensional lattice Λ can be bounded as $\kappa \leq O(R^D)$, which is the size of a ball of diameter R. Nevertheless, we treat both κ and R as independent parameters in this paper.*

For the Hamiltonians we consider, the sum $\| \sum_{X \cap \{x_0\} \neq \emptyset} H_X \|$ is bounded from above by a constant like $O(hR^{\kappa D})$ for any $x_0 \in \Lambda$, but in general, this is a loose bound and we introduce the *growth constant* as an independent parameter such that:

Definition 130 (Growth constant)**.** *Given a geometrically-local Hamiltonian H, the growth constant g is defined such that $|\sum_{X \cap \{x_0\} \neq \emptyset} H_X| \leq gh$ for all sites $x_0 \in \Lambda$.*

Given a (κ, R)-local Hamiltonian H, we denote the boundary of a region $A \subseteq \Lambda$ by ∂A and define it as $\partial A = \{v \in \Lambda \setminus A : \exists v' \in A,\ \operatorname{dist}(v - v') \leq R\}$. Defined this way, the boundary of A is a subset of $\bar{A}$.

For local Hamiltonians with $\kappa = 2$, we define an *interaction graph* which is an undirected graph $G = (V, E)$ with a qudit on each vertex $v \in V$ and an edge (i, j) between any two vertices i, j that are acted on by a local term in the Hamiltonian. For qubits, $d = 2$ and such a Hamiltonian is of the following form:

$$H = - \sum_{\substack{(i,j) \in E \\ a,b \in \{x,y,z\}}} J_{ij}^{ab} \sigma_a \otimes \sigma_b - \sum_{\substack{i \in V \\ a \in \{x,y,z\}}} h_i^a \sigma_a, \tag{5.10}$$

where $J_{ij}^{ab}, h_i^a \in \mathbb{R}$ are the interaction coefficients and $\sigma_a \in \{X, Y, Z, \mathbb{1}\}$ are Pauli matrices.

5.2.2 Quantum thermal state and partition function

The free energy of state ρ at inverse temperature β is defined as

$$F_\beta(\rho) = \operatorname{tr}(H\rho) - \frac{1}{\beta} S(\rho),$$

where $S(\rho) = -\operatorname{tr}(\rho \log \rho)$ is the von Neumann entropy of ρ (here and throughout this paper, we assume log denotes the natural logarithm). In thermal equilibrium, the free energy of the system is minimized. Using the non-negativity of the relative entropy $S(\rho \| \frac{e^{-\beta H}}{Z(\beta)}) \geq 0$, one can see that

$$\begin{aligned} \min_\rho F_\beta(\rho) &= \min_\rho \operatorname{tr}(H\rho) - \frac{1}{\beta} S(\rho) \\ &= -\frac{1}{\beta} \log(Z_\beta(\Lambda)), \end{aligned} \tag{5.11}$$

where $Z_\beta(H) = \mathrm{tr}[\exp(-\beta H)]$ is the partition function of the system at inverse temperature β. When dealing with spin systems on a lattice, we often denote the partition function of the system by $Z_\beta(\Lambda)$ rather than $Z_\beta(H)$.

Furthermore, the state that achieves this minimization, known as the Gibbs (or thermal) state, is given by

$$\rho_\beta(H) = \frac{\exp(-\beta H)}{Z_\beta(H)}. \tag{5.12}$$

We often need to consider the Gibbs state after some measurement has been performed on a local region of the lattice. The post-selected state $\rho_\beta(H|N)$ associated with a positive operator N is given by

$$\rho_\beta(H|N) = \frac{\sqrt{N}\exp(-\beta H)\sqrt{N}}{\mathrm{tr}[\exp(-\beta H)N]}. \tag{5.13}$$

5.2.3 Quantum belief propagation

Suppose certain local terms in Hamiltonian H are removed and consider the Gibbs state before and after this change. We would like to relate these Gibbs states by applying a local operator on the old state to obtain the new one. This has been addressed before in [Has07b] under the name *quantum belief propagation*. We only mention this result without the proof and refer the reader to [Has07b, KBa19b] for the derivation and more details.

Proposition 131 (Quantum belief propagation)**.** *Let H be a geometrically-local Hamiltonian on lattice Λ. Consider a sublattice $C \subset \Lambda$. We denote the terms in H acting on both C and $\bar{C}$ by $H_{\partial C}$. There exists a quasi-local operator η such that*

$$e^{-\beta H} = \eta e^{-\beta(H - H_{\partial C})}\eta^\dagger, \tag{5.14}$$

where $\|\eta\| \leq \exp(\beta/2\|H_{\partial C}\|)$. Moreover, there exists a truncation of η denoted by η_ℓ supported non-trivially only on ∂C and sites within distance ℓ from ∂C such that for some positive constants α_1, α_2,

$$\left\|\eta - \eta_\ell\right\| \leq e^{\alpha_1|\partial C| - \alpha_2 \ell}. \tag{5.15}$$

5.2.4 Tools from complex analysis

Given a function that is analytic in a region of the complex plane, i.e. it is complex differentiable, we are interested in approximating the function in that region with a low-degree polynomial. Conventional methods in complex analysis allow us to achieve this using a Taylor expansion around a point inside that region.

Definition 132 (Taylor expansion of analytical functions)**.** *Given a complex function $f(z)$ that is analytic in a region A, the Taylor expansion of $f(z)$ around a point $z_0 \in A$ is a power series $\sum_{k=0}^{\infty} a_k (z - z_0)^k$, where for $\forall k = 0, 1, \ldots$*

$$a_k = \frac{1}{k!}\frac{d^k f(z_0)}{dz^k} = \frac{1}{2\pi i}\oint_C \frac{f(w)}{(w - z_0)^{k+1}} dw \tag{5.16}$$

for an arbitrary contour C around z_0 inside the region A.

In Section 5.7, we map the partition function of a quantum system to that of a classical system. As we increase the precision of the mapping, we get a family of classical systems with increasing size that in the limit of an infinite number of particles have the same partition function as the quantum system. The following theorem guarantees that the zero-free region of the classical ensemble coincides with that of the original quantum system.

Theorem 133 (Multivariate Hurwitz's theorem)**.** *If a sequence of multivariate functions $f_1, f_2, f_3, \ldots$ are analytic and non-vanishing on a connected open set $D \subset \mathbb{C}^n$ and converge uniformly on compact subsets of D to f, then f is either non-vanishing on D or is identically zero.*

The proof can be found in standard complex analysis textbooks [Gam03].

5.3 Algorithm for estimating the partition function

In this section, we provide more details about the approximation algorithm that we presented in Section 5.1.

Definition 134. *An approximation algorithm for the partition function $Z_\beta(H)$ takes as input the description of the local Hamiltonian H, the inverse temperature β, and a parameter ε and gives an estimate $\tilde{Z}_\beta(H)$ with ε-multiplicative error, i.e.*

$$\left|\tilde{Z}_\beta(H) - Z_\beta(H)\right| \leq \varepsilon Z_\beta(H). \tag{5.17}$$

This is, up to unimportant constants, equivalent to finding an ε-additive error for $\log Z_\beta(H)$ or $F_\beta(H)$.

We now make a connection between analyticity of functions and approximation algorithms precise. Similar propositions were first proved by [Bar16a] for bounded degree polynomials.

Suppose we want to estimate the value of a complex function $f(z)$. We consider two cases. One is when there is an upper bound on the absolute value of the function in the region that $f(z)$ is analytic. The other is when the given function is $f(z) = \log(g(z))$ for a polynomial $g(z)$ of degree n. The latter is used in Section 5.7.2 when studying the XXZ model. We need the former version since as we will see in Theorem 136, the partition function of quantum (or even some classical) systems is not always a polynomial in $\exp(\beta J)$.

Proposition 135 (Truncated Taylor series for bounded functions and polynomials)**.** *We denote a disk of radius b centered at the origin in the complex plane by Δ_b, that is $\Delta_b = \{z \in \mathbb{C} : |z| \leq b\}$.*

(1) Let $f(z)$ be a complex function that is analytic and bounded as $|f(z)| \leq M$ when $z \in \Delta_b$ for a constant $b > 1$. Then the error of approximating $f(z)$ by a truncated Taylor series of order K for all $|z| \leq 1$ is bounded by

$$\left|f(z) - \sum_{k=0}^{K} a_k z^k\right| \leq \frac{M}{b^K(b-1)}, \quad |z| \leq 1. \tag{5.18}$$

(2) *Assume b is fixed and there is a deterministic algorithm that finds the coefficients a_k in time $O(N^k)$ for some parameter N. Then there exists a deterministic algorithm with running time $N^{O(\log(M/\varepsilon))}$ that outputs an ε-additive approximation for $f(z)$.*

(3) *[cf. [Bar16a]] Let $f(z) = \log(g(z))$ for some polynomial $g(z)$ of degree N that does not vanish when $z \in \Delta_b$. The error of approximating $f(z)$ by a truncated Taylor series of order K for $|z| \le 1$ is bounded by $\frac{N}{K+1}\frac{1}{b^K(b-1)}$.*

(4) *[cf. [Bar16a]] Assuming b is fixed, there exists a deterministic algorithm with running time $N^{O(\log(N/\varepsilon))}$ that outputs an ε-additive approximation for $\log(g(z))$.*

Proof. The proof of (1) is a basic result in complex analysis based on the Cauchy integral theorem for analytic functions. Let C' be a circle of radius b that contains both z and $z = 0$. We have

$$\begin{aligned} f(z) =& \frac{1}{2\pi i}\oint_{C'} \frac{f(w)}{w-z}dw = \frac{1}{2\pi i}\oint_{C'} \frac{f(w)}{w}\left(1-\frac{z}{w}\right)^{-1} dw \\ &= \frac{1}{2\pi i}\oint_{C'} \frac{f(w)}{w}\left(\sum_{k=0}^{K}\left(\frac{z}{w}\right)^k + \left(\frac{z}{w}\right)^{K+1}\left(1-\frac{z}{w}\right)^{-1}\right) dw \\ &= \sum_{k=0}^{K}\frac{f^{(k)}(0)}{k!}z^k + \frac{1}{2\pi i}\oint_{C'}\frac{f(w)}{w-z}\left(\frac{z}{w}\right)^{K+1} dw, \end{aligned}$$

in which we used Eq. (5.16) to get to the last line. We can now bound the remainder as

$$\begin{aligned} \left|f(z) - \sum_{k=0}^{K}\frac{f^{(k)}(0)}{k!}z^k\right| &\le \frac{1}{2\pi}\oint_{C'}\frac{|f(w)|}{|w-z|}\left(\left|\frac{z}{w}\right|\right)^{K+1} dw. \\ &\le M\frac{b}{b-1}\left(\frac{1}{b}\right)^{K+1}, \end{aligned} \tag{5.19}$$

where the last line follows from the fact that $|w - z| \ge b - 1$, $|z| \le 1$, and $|f(w)| \le M$ on C'.

The proof of part (3) is similar to that of (1). The degree N polynomial $g(z)$ has at most N complex roots $\{\zeta_k\}_{k=1}^N$ such that $|\zeta_k| \ge b$. Thus, $g(z) = g(0)\prod_{l=1}^N(1 - \frac{z}{\zeta_l})$ and

$$\log(g(z)) = \log(g(0)) + \sum_{l=1}^{N}\log\left(1-\frac{z}{\zeta_l}\right), \quad \forall z:\ |z| \le 1. \tag{5.20}$$

We can expand each term like $\log(1 - \frac{z}{\zeta_l})$ as

$$\log\left(1-\frac{z}{\zeta_l}\right) = -\sum_{k=1}^{K}\frac{z^k}{k\zeta_l^k} + q_\ell(z), \tag{5.21}$$

where similar to part (1), we see that $q_\ell(z)$ is a term that can be bounded by

$$|q_\ell(z)| \le \frac{1}{K+1}\frac{1}{b^K(b-1)}. \tag{5.22}$$

Hence, the remainder term in the Taylor expansion of $\log(g(z))$ up to order K is $q(z) = \sum_{\ell=1}^{N} q_\ell(z)$, which is bounded by $|q(z)| \leq \frac{N}{K+1} \frac{1}{b^K(b-1)}$ as claimed in part (3).

In order to find the algorithms of part (2) and (4), we need to evaluate the Taylor coefficients of $f(z)$ up to some degree K. Since we want an ε-additive approximation of $f(z)$, one can see from parts (1) and (2) that it is sufficient to keep the Taylor expansion until order $K = O(\log(\frac{M}{\varepsilon}))$ for part (2) and $K = O(\log(\frac{N}{\varepsilon}))$ for part (4). To be able to evaluate the derivatives $\frac{d^k f(z)}{dz^k}$, we express them in terms of the derivatives of $g(z)$, i.e. $\frac{d^k g(z)}{dz^k}$ [2]. This can be done by noticing that

$$\frac{d^k g(z)}{dz^k} = \sum_{\ell=0}^{k-1} \binom{k-1}{\ell} \frac{d^\ell g(z)}{dz^\ell} \frac{d^{k-\ell} f(z)}{dz^{k-\ell}}, \tag{5.23}$$

so if we have access to $\frac{d^k g(z)}{dz^k}$, we can find $\frac{d^k f(z)}{dz^k}$ by solving the system of equations in time $\text{poly}(k)$. The important step, however, is to estimate $\frac{d^k g(z)}{dz^k}$. This by assumption takes time $N^{O(k)}$ for the kth derivative. Thus, evaluating the Taylor expansion in parts (2) and (4) can be done in time $N^{O(\log(M/\varepsilon))}$ and $N^{O(\log(N/\varepsilon))}$, respectively. □

Theorem 136 (Extrapolation algorithm for estimating the partition function)**.** *There exists a deterministic classical algorithm that runs in time $n^{O(\log(n/\varepsilon))}$ and outputs an estimate within ε-multiplicative error of the partition function $Z_\beta(H)$ at some constant β in the zero-free region $\Omega_{\delta,\beta}$ (see Definition 119).*

Proof of Theorem 136. We apply the truncated Taylor expansion. To use that result, we first need to specify the zero-free region and then bound the running time of computing the kth derivative by $n^{O(k)}$.

We can without loss of generality assume that the zero-free region $\Omega_{\delta,\beta}$ is a rectangular region of constant width and size depicted in Figure 5-1. The result of Proposition 135, however, holds when the zero-free region is a disk of radius b. To match these domains, we can compose the partition function with a function $\phi(z)$ that maps a disk of radius b to the rectangular region $\Omega_{\delta,\beta}$ such that $\phi(0) = 0$ and $\phi(1) = \beta$ and b is constant depending on δ. It is shown in Lemma 2.2.3 of [Bar16a] that one can find such a $\phi(z)$ which is a constant degree polynomial. Hence, the composed partition function is non-zero and bounded on this disk and we can apply the bound (5.18) on the error of the truncated Taylor expansion.

As mentioned in Section 5.1.2, for a system of n qudits, we can compute the order k derivatives of $Z_\beta(H)$ in time $n^{O(k)}$. Similarly, we can evaluate the derivatives of $Z_\beta(H)$ composed with the constant-degree polynomial $\phi(z)$ using the same running time. Keeping only $k = O(\log(n/\varepsilon))$ many terms results in a quasi-polynomial algorithm with multiplicative error ε. □

5.4 Lower bound on the critical inverse temperature

In this section, we show that at high temperatures, there are no complex zeros near the real axis. More precisely, we prove that there exists a disk of constant radius β_0 centered at $\beta = 0$ that does not contain any zeros and the free energy is analytic inside it. The radius β_0 depends only on the geometric parameters of the Hamiltonian such as the growth constant.

[2] We are using the same definition $f(z) = \log(g(z))$ for the function in part (1) as well.

Theorem 137 (High temperature zeros)**.** *Let H be a gometrically-local Hamiltonian on qudits with range R, growth constant g, and local interactions with norm at most h (see Definition 128 and Definition 130). There exists a real constant $\beta_0 = 1/(5egh\kappa)$ such that for all $\beta \in \mathbb{C}$ with $|\beta| \leq \beta_0$, the partition function $Z_\beta(\Lambda)$ of H does not vanish and $\log(Z_\beta(\Lambda))$ is analytic and bounded by $\left|\log|Z_\beta(\Lambda)|\right| \leq (e^2 gh|\beta| + \log d)n$.*

This gives a lower bound $\beta_0 \leq \beta_c$ on the phase transition point β_c. Also, as outlined in Theorem 136, if we can establish an upper bound like $\left|\log|Z_\beta(\Lambda)|\right| \leq O(n)$ for small enough complex β, we can devise an approximation algorithm for the partition function. Hence we get

Corollary 138 (Approximation algorithm for the partition function at high temperatures)**.** *There exists a quasi-polynomial time algorithm with running time $n^{O(\log(n/\varepsilon))}$ that outputs an ε-multiplicative approximation to the partition function $Z_\beta(\Lambda)$ of a geometrically-local Hamiltonian H when $|\beta| \leq \beta_0$.*

Before getting to the proof of Theorem 137, we need to gather some facts and lemmas. Given a lattice $\Lambda \subset \mathbb{Z}^D$ with n sites, we consider a series of sublattices $\Lambda_0 \subset \Lambda_1 \subset \Lambda_2 \subset \Lambda_2 \subset \cdots \subset \Lambda_n = \Lambda$ such that each sublattice Λ_j has one fewer vertex than Λ_{j+1} and $\Lambda_0 = \emptyset$. The partition function of Λ_0 is assigned to be $Z_\beta(\emptyset) = 1$ for any complex β. Therefore, we can write

$$Z_\beta(\Lambda) = d^n \prod_{j=0}^{n-1} \left(\frac{1}{d}\frac{Z_\beta(\Lambda_{j+1})}{Z_\beta(\Lambda_j)}\right), \tag{5.24}$$

where the factors of d are added for later convenience and to account for the dimension of the removed sites. In order to show $|\log|Z_\beta(\Lambda)|| \leq O(n)$ for a $\beta \in \mathbb{C}$, we just need to bound the logarithm of each of the terms in Eq. (5.24) by a constant, i.e.

$$\left|\log\left|\frac{1}{d}\frac{Z_\beta(\Lambda_{j+1})}{Z_\beta(\Lambda_j)}\right|\right| \leq O(1). \tag{5.25}$$

This bound tells us how much the partition function changes after removing a single site from the lattice. We later prove this by induction on the number of sites. However, as shown in the following lemma, this inequality is always satisfied when β is *real*.

Lemma 139 (Site removal bound)**.** *The following bound holds for any $X \subseteq \Lambda$ and $\beta \in \mathbb{R}^+$:*

$$\left|\log\left|\frac{1}{d^{|X|}}\frac{Z_\beta(\Lambda)}{Z_\beta(\Lambda \setminus X)}\right|\right| \leq gh|\beta||X|. \tag{5.26}$$

Recall that h is the maximum norm of the local terms H_X in H and the growth constant g is chosen such that $|\sum_{X \cap \{x_0\} \neq \emptyset} H_X| \leq gh$ for all sites $x_0 \in \Lambda$.

Proof. Given any $X \subseteq \Lambda$, we partition the Hamiltonian H into two parts: $H_{\text{near}} = \sum_{X' \subset \Lambda : X' \cap X \neq \emptyset} H_{X'}$ and $H_{\text{far}} = H_{\Lambda \setminus X}$ where $H_{\Lambda \setminus X}$ corresponds to the terms in the Hamiltonian

acting on the remaining sublattice $\Lambda \setminus X$. We have

$$\begin{aligned} Z_\beta(\Lambda) = \mathrm{tr}_\Lambda \left[e^{-\beta(H_{\mathrm{far}}+H_{\mathrm{near}})}\right] &\leq \mathrm{tr}_\Lambda \left[e^{-\beta H_{\mathrm{far}}} e^{-\beta H_{\mathrm{near}}}\right] \\ &\leq \mathrm{tr}_\Lambda \left[e^{-\beta H_{\mathrm{far}}}\right] \left\| e^{-\beta H_{\mathrm{near}}} \right\| \\ &\leq d^{|X|}\, \mathrm{tr}_{\Lambda\setminus X} \left[e^{-\beta H_{\mathrm{far}}}\right] e^{\left\|\beta H_{\mathrm{near}}\right\|} \\ &\leq Z_\beta(\Lambda \setminus X) d^{|X|} e^{gh|\beta||X|}, \end{aligned} \tag{5.27}$$

where we used the Golden-Thompson inequality in the first line and the Hölder inequality to get to the second line. The factor $d^{|X|}$ is added since the original trace is over the Hilbert space of Λ and not $\Lambda \setminus X$. Similarly, one can show $d^{|X|} Z_\beta(\Lambda \setminus X) \leq Z_\beta(\Lambda) e^{gh|\beta||X|}$. These bounds together prove the lemma. □

Theorem 137 extends bound (5.26) to the case where β is a small complex number. We prove this in two steps.

First step: In contrast to the proof of Lemma 139, the Golden-Thompson inequality can no longer be used in the complex regime. Hence, to compare the partition function before and after removing a site x_0, we need to find another way of separating the contribution of the terms in the Hamiltonian that act on x_0. We achieve this using a *cluster expansion* for the partition function that expands the operator $\exp(-\beta H)$ into a sum of products of local terms in H. The idea of using cluster expansions to study high temperature properties of classical or quantum spin systems has been widely applied before [KP86, Dob96, Par82, Gre69]. Here, we use a particular version of that expansion which is tailored for our application. This was first introduced in [Has06] and later improved and generalized in [KGK+14]. In Section 5.4.1, we modify the result of [Has06, KGK+14] and adapt it for complex partition functions.

Second step: Our next step is to use the cluster expansion and show that in the partition function, the contribution of the sites acting on x_0 compared to the rest of the terms is bounded by a constant. We show this in Section 5.4.2 by induction on the number of sites. This is our main contribution and lets us prove the bound (5.26).

5.4.1 The cluster expansion for the partition function

When using the cluster expansion, we often need to consider products of local terms like $\prod_{j=1}^{\ell} H_{X_j}$, but since the local interaction terms H_{X_j} do not necessarily commute with each other, we set an ℓ-tuple $(X_1, \ldots, X_\ell)$ to indicate the order of multiplication. We also need to decompose the sequence $(X_1, \ldots, X_\ell)$ into the union of *connected* components. Let us define what we mean by connected more formally.

Definition 140 (Connected sets)**.** *Fix a site $x_0 \in \Lambda$. A collection of sublattices such as $\mathcal{X} = \{X_1, X_2, \ldots, X_k\}$ is called a connected set containing x_0 with size $|\mathcal{X}| = k$ if the following conditions hold*

i) All the sublattices $X_1, X_2, \ldots, X_k$ have bounded size and diameter. That is $1 \leq |X_i| \leq \kappa$ and $\mathrm{diam}(X_i) \leq R$.

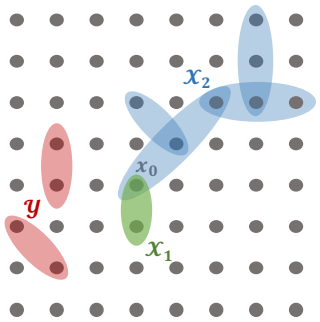


Figure 5-2: The sets $\mathcal{X}_1, \mathcal{X}_2$ are connected, contain x_0, and have size $|\mathcal{X}_1| = 1, |\mathcal{X}_2| = 4$. However, the set $\mathcal{Y}$ is not connected, does not include x_0, and has size $|\mathcal{Y}| = 2$.

ii) For any sublattice X_i in $\mathcal{X}$, a series of other members of $\mathcal{X}$ connect this set to the site x_0. See Figure 5-2 for an example. More precisely we have: for any $X_i \in \mathcal{X}$, there exists $I \subseteq [k]$ such that $i \in I$ and $\forall j \in I, \exists \ell \in I : X_j \neq X_\ell$ yet $X_j \cap X_l \neq \emptyset$, and moreover, $x_0 \in \cup_{j \in I} X_j$.

Although $\mathcal{X}$ consists of sublattice of Λ and not individual sites, in a slight abuse of notation, we specify a set $\mathcal{X}$ that contains the site x_0 by $x_0 \in \mathcal{X}$. We denote all the sites that a connected set $\mathcal{X}$ includes by $\mathrm{supp}(\mathcal{X})$.

Remark 141. *In Definition 140, we include an upper bound on the size and diameter of the subsets in $\mathcal{X}$, i.e. $|X_i| \leq \kappa, \mathrm{diam}(X_i) \leq R$. This is because for geometrically-local Hamiltonians, $\|H_X\| = 0$ for $|X_i| > \kappa$, $\mathrm{diam}(X_i) > R$, so we do not need to consider those sets.*

In the upcoming proofs, we need to have an upper bound on the number of the connected sets $\mathcal{X}$ that contains a specific site $x_0 \in \Lambda$. This is stated in the following lemma.

Lemma 142 (Cf. [KGK+14])**.** *The number of connected sets $\mathcal{X}$ of size $|\mathcal{X}|$ containing the site $x_0 \in \Lambda$ is upper bounded by $g^{|\mathcal{X}|}$ where g is the growth constant of the Hamiltonian H (see Definition 130). In particular, for a D-dimensional lattice and $\kappa = 2$, we have $g \leq 2eD$. We note that these bounds are in general loose and can be improved for specific Hamiltonians.*

The next lemma achieves the first step in our proof by setting up the cluster expansion for the partition function.

Lemma 143 (High temperature expansion)**.** *For any $x_0 \in \Lambda$, the partition function of the lattice $Z_\beta(\Lambda)$ admits the following decomposition for $|\beta| \leq \frac{1}{gh\kappa(e-1)}$:*

$$Z_\beta(\Lambda) = d \cdot Z_\beta(\Lambda \setminus X_0) + \sum_{\substack{\mathcal{X}: x_0 \in \mathcal{X} \\ \mathcal{X} \text{ is connected}}} W_\beta(\mathcal{X}) Z_\beta(\Lambda \setminus \mathrm{supp}(\mathcal{X})), \tag{5.28}$$

where $X_0 = \{x_0\}$ and we define $W_\beta(\mathcal{X})$ as

$$W_\beta(\mathcal{X}) = \sum_{p=|\mathcal{X}|}^{\infty} \frac{(-\beta)^p}{p!} \Big(\sum_{\substack{(X_1,\ldots,X_p) \\ \forall i \in [p]: X_i \in \mathcal{X} \\ \mathcal{X} = \cup_{i=1}^p \{X_i\}}} \mathrm{tr}_{\mathrm{supp}(\mathcal{X})} \Big[\prod_{j=1}^{p} H_{X_j} \Big] \Big). \tag{5.29}$$

The last sum in (5.29) is over all p-tuples $(X_2, X_2, \ldots, X_p)$ that one can form from members of $\mathcal{X}$ by repeating them at least once.

Proof. We start by Taylor expanding $\exp(-\beta H)$. The notation used here is set so that when Λ is replaced with a sublattice, a similar bound holds. We have

$$\begin{aligned} Z_\beta(\Lambda) &= \mathrm{tr}_\Lambda \big[\sum_{k=0}^{\infty} \frac{(-\beta)^k}{k!} (\sum_{X \subset \Lambda} H_X)^k \big] \\ &= \mathrm{tr}_\Lambda \big[\sum_{k=0}^{\infty} \frac{(-\beta)^k}{k!} (\sum_{X \subset \Lambda \setminus X_0} H_X + \sum_{X \subset \Lambda : X \cap X_0 \neq \emptyset} H_X)^k \big] \\ &= \mathrm{tr}_\Lambda \big[\sum_{k=0}^{\infty} \frac{(-\beta)^k}{k!} (\sum_{X \subset \Lambda \setminus X_0} H_X)^k \big] + \mathrm{tr}_\Lambda \big[\sum_{\ell=1}^{\infty} \frac{(-\beta)^\ell}{\ell!} \sum_{\substack{(X_1,\ldots,X_\ell) \\ \forall i \in [\ell] : X_i \subset \Lambda \\ \exists X_i : X_i \cap X_0 \neq \emptyset}} \prod_{j=1}^{\ell} H_{X_j} \big], \end{aligned} \tag{5.30}$$

where the trace is over the Hilbert space of Λ as usual. The first term in the last line is just the Taylor expansion of $d \cdot Z_\beta(\Lambda \setminus X_0)$. As in Eq. (5.27), the factor d is included because the original trace is over Λ and not $\Lambda \setminus X_0$. The last term, however, does not have a closed form, and involves summing over all the products of the local interaction terms H_{X_j} such that at least one of the terms has non-empty overlap with the site x_0. We can simplify this term by partitioning the sequence $(X_1, \ldots, X_l)$ into two parts. The first part forms a connected set $\mathcal{X}$ that contains the site x_0. The second part contains all X_i that do not intersect with this connected set $\mathcal{X}$. We then change the order of the summation in (5.30) by first summing over all X_i not connected to a fixed $\mathcal{X}$ and then varying the set $\mathcal{X}$. Define the coefficient $c_{p,q} = \binom{p+q}{p} \frac{(-\beta)^{p+q}}{(p+q)!}$, we have

$$\begin{aligned} &\sum_{\ell=1}^{\infty} \frac{(-\beta)^\ell}{\ell!} \sum_{\substack{(X_1,\ldots,X_l) \\ \forall i \in [\ell] : X_i \subset \Lambda \\ \exists X_i : X_i \cap X_0 \neq \emptyset}} \mathrm{tr}_\Lambda \big[\prod_{j=1}^{\ell} H_{X_j} \big] \\ &= \sum_{\substack{\mathcal{X} : x_0 \in \mathcal{X} \\ \mathcal{X} \text{ is connected}}} \sum_{p=|\mathcal{X}|, q=0}^{\infty} c_{p,q} \Big(\sum_{\substack{(X_1,\ldots,X_p) \\ \forall i \in [p] : X_i \in \mathcal{X} \\ \mathcal{X} = \cup_{i=1}^p \{X_i\}}} \mathrm{tr}_{\mathrm{supp}(\mathcal{X})} \big[\prod_{j=1}^{p} H_{X_j} \big] \sum_{\substack{(X_{p+1},\ldots,X_{p+q}) \\ X_{p+i} \cap \mathrm{supp}(\mathcal{X}) = \emptyset}} \mathrm{tr}_{\Lambda \setminus \mathrm{supp}(\mathcal{X})} \big[\prod_{j=p+1}^{p+q} H_{X_j} \big] \Big). \end{aligned}$$

The coefficient $\binom{p+q}{p}$ in $c_{p,q}$ counts the number of ways we can distribute our choices of $X_i \in \mathcal{X}$ inside the tuple $(X_1, \ldots, X_{p+q})$. The last sum in the right side term vanishes for $q = 0$. We can restate this sum in terms of the Taylor expansion of $Z_\beta(\Lambda \setminus \mathrm{supp}(\mathcal{X}))$. Using the fact that $c_{p,q} = \frac{(-\beta)^p}{p!} \frac{(-\beta)^q}{q!}$,

we get the following equality

$$\begin{aligned}
&\sum_{\ell=1}^{\infty} \frac{(-\beta)^{\ell}}{\ell!} \sum_{\substack{(X_1,\dots,X_\ell)\\ \forall i\in[\ell]:X_i\subset\Lambda\\ \exists X_i: X_i\cap X_0\neq\emptyset}} \operatorname{tr}_{\Lambda}\big[\prod_{j=1}^{\ell} H_{X_j}\big] \\
&= \sum_{\substack{\mathcal{X}:x_0\in\mathcal{X}\\ \mathcal{X}\text{ is connected}}} \sum_{p=|\mathcal{X}|}^{\infty} \frac{(-\beta)^{p}}{p!}\Big(\sum_{\substack{(X_1,\dots,X_p)\\ \forall i\in[p]:X_i\in\mathcal{X}\\ \mathcal{X}=\cup_{i=1}^{p}\{X_i\}}} \operatorname{tr}_{\operatorname{supp}(\mathcal{X})}\big[\prod_{j=1}^{p} H_{X_j}\big]\Big) Z_\beta\big(\Lambda\setminus\operatorname{supp}(\mathcal{X})\big) \\
&= \sum_{\substack{\mathcal{X}:x_0\in\mathcal{X}\\ \mathcal{X}\text{ is connected}}} W_\beta(\mathcal{X}) Z_\beta(\Lambda\setminus\operatorname{supp}(\mathcal{X})), \qquad (5.31)
\end{aligned}$$

which by plugging into Eq (5.30) gives us the expansion (5.28). Note that since we manipulated infinite series, we still need to prove the convergence of the expansion (5.28) for small enough complex β. We show the *absolute convergence* of this expansion by first bounding the infinite series $W_\beta(\mathcal{X})$ and then the expression (5.31). A similar expansion for a different purpose has been considered before in [Has06, KGK+14] where an upper bound for $W_\beta(\mathcal{X})$ is obtained. In particular, Lemma 5 in [KGK+14] implies [3]

$$|W_\beta(\mathcal{X})| \le d^{|\operatorname{supp}(\mathcal{X})|}\Big(e^{|\beta|h}-1\Big)^{|\mathcal{X}|}. \qquad (5.32)$$

By using the result of Lemma 142, we see that

$$\sum_{\substack{\mathcal{X}:x_0\in\mathcal{X}\\ \mathcal{X}\text{ is connected}}} \big|W_\beta(\mathcal{X})\big|\cdot\big|Z_\beta(\Lambda\setminus\operatorname{supp}(\mathcal{X}))\big| \le d^n e^{gh|\beta|n} \sum_{|\mathcal{X}|=1}^{\infty} (g\kappa)^{|\mathcal{X}|}\Big(e^{|\beta|h}-1\Big)^{|\mathcal{X}|}, \qquad (5.33)$$

in which we used the upper bound $|Z_\beta(\Lambda\setminus\operatorname{supp}\mathcal{X})| \le d^{n-|\operatorname{supp}(\mathcal{X})|}e^{gh|\beta|n}$ that can be shown using the Hölder inequality. This right-hand side of the inequality (5.33) is finitely bounded when

$$g\kappa(e^{|\beta|h}-1) \le 1, \qquad (5.34)$$

which along with the inequality $e^x \le 1+(e-1)x$ implies an upper bound on the size of the admissible β

$$|\beta| \le \frac{1}{gh\kappa(e-1)}. \qquad (5.35)$$

Hence, we get

$$\sum_{\substack{\mathcal{X}:x_0\in\mathcal{X}\\ \mathcal{X}\text{ is connected}}} \big|W_\beta(\mathcal{X})\big|\cdot\big|Z_\beta(\Lambda\setminus\operatorname{supp}(\mathcal{X}))\big| \le d^n e^{gh|\beta|n}\frac{g\kappa(e^{|\beta|h}-1)}{1-g\kappa(e^{|\beta|h}-1)}, \qquad (5.36)$$

[3] Note that compared to [KGK+14] we pick up the extra factor $d^{|\operatorname{supp}(\mathcal{X})|}$ when bounding $\big|\operatorname{tr}\big[\prod H_{X_j}\big]\big|$.

which for a fixed n, shows the absolute convergence of (5.28) and completes the proof of the lemma. □

Having this lemma, we can now proceed to the second step of our proof of Theorem 137.

5.4.2 A zero-free region at high temperatures

Proof of Theorem 137. As explained in the beginning of Section 5.4, to show the partition function does not vanish for small enough $|\beta|$, and moreover $|\log Z_\beta(\Lambda)| \le O(n)$, it is sufficient to prove the bound in (5.25). More specifically, for any site $x_0 \in \Lambda$ and $X_0 := \{x_0\}$, we prove

$$\left|\log\left|\frac{1}{d}\frac{Z_\beta(\Lambda)}{Z_\beta(\Lambda\setminus X_0)}\right|\right| \le e^2 gh|\beta|, \quad \forall|\beta| \le \beta_0 = \frac{1}{5egh\kappa} \tag{5.37}$$

The proof of this bound is by induction on the number of lattice sites n.

For the base of the induction, we assume $Z_\beta(\emptyset) = 1$ for all complex β. The induction hypothesis is the bound (5.37). Thus, our goal is to assume (5.37) for lattices of size $n-1$ and show that the same bound holds for lattices of size n. By using the "telescoping product" as in Eq. (5.25) along with the induction hypothesis, we obtain the following bound for all lattices of size at most $n-1$ including $\Lambda \setminus X_0$,

$$\left|\log\left|\frac{1}{d^{|\operatorname{supp}(X\setminus X_0)|}}\frac{Z_\beta(\Lambda\setminus X_0)}{Z_\beta(\Lambda\setminus X)}\right|\right| \le e^2 gh|\beta||\operatorname{supp}(X\setminus X_0)|, \quad |\beta| \le \beta_0, \tag{5.38}$$

where $X \subseteq \Lambda$ is an arbitrary non-empty set. According to the decomposition of $Z_\beta(\Lambda)$ obtained in Lemma 143, we have

$$\frac{1}{d}\frac{Z_\beta(\Lambda)}{Z_\beta(\Lambda\setminus X_0)} = 1 + \sum_{\substack{\mathcal{X}: x_0\in\mathcal{X}\\ \mathcal{X}\text{ is connected}}} W_\beta(\mathcal{X})\left(\frac{1}{d}\frac{Z_\beta(\Lambda\setminus\operatorname{supp}(\mathcal{X}))}{Z_\beta(\Lambda\setminus X_0)}\right). \tag{5.39}$$

Thus, we get

$$\begin{aligned}
\left|\log\left|\frac{1}{d}\frac{Z_\beta(\Lambda)}{Z_\beta(\Lambda\setminus X_0)}\right|\right| &= \left|\log\left|1 + \sum_{\substack{\mathcal{X}: x_0\in\mathcal{X}\\ \mathcal{X}\text{ is connected}}} W_\beta(\mathcal{X})\left(\frac{1}{d}\frac{Z_\beta(\Lambda\setminus\operatorname{supp}(\mathcal{X}))}{Z_\beta(\Lambda\setminus X_0)}\right)\right|\right| \\
&\le -\log\left(1 - \sum_{\substack{\mathcal{X}: x_0\in\mathcal{X}\\ \mathcal{X}\text{ is connected}}} |W_\beta(\mathcal{X})|\left|\frac{1}{d}\frac{Z_\beta(\Lambda\setminus\operatorname{supp}(\mathcal{X}))}{Z_\beta(\Lambda\setminus X_0)}\right|\right) \qquad (5.40)\\
&\le -\log\left(1 - \sum_{\substack{\mathcal{X}: x_0\in\mathcal{X}\\ \mathcal{X}\text{ is connected}}} \left(e^{|\beta|h}-1\right)^{|\mathcal{X}|} e^{ghe^2|\beta||\operatorname{supp}(\mathcal{X})|}\right), \qquad (5.41)
\end{aligned}$$

where we used the following inequality to get to Eq. (5.40): for all $\zeta \in \mathbb{C}, |\zeta| \le 1$, we have $\big|\log|1+\zeta|\big| \le -\log(1-|\zeta|)$. The last line (5.41) is obtained by plugging in the bound in (5.32) and the induction hypothesis (5.38).

It remains to show that Eq. (5.41) is bounded from above by $e^2gh|\beta|$. To get the desired upper bound on (5.41), it is sufficient to prove the following bound which we separately prove in Lemma 144:

$$\sum_{\substack{\mathcal{X}:x_0\in\mathcal{X}\\ \mathcal{X}\text{ is connected}}} \left(e^{|\beta|h}-1\right)^{|\mathcal{X}|} e^{ghe^2|\beta||\operatorname{supp}(\mathcal{X})|} \le e(e-1)gh|\beta|, \quad |\beta|\le\beta_0. \tag{5.42}$$

The reason this implies the claimed upper bound on (5.41) is that we have

$$\begin{aligned} -\log\left(1-\sum_{\substack{\mathcal{X}:x_0\in\mathcal{X}\\ \mathcal{X}\text{ is connected}}} \left(e^{|\beta|h}-1\right)^{|\mathcal{X}|} e^{ghe^2|\beta||\operatorname{supp}(\mathcal{X})|}\right) &\le -\log\left(1-e(e-1)gh|\beta|\right)\\ &\le e^2gh|\beta|. \end{aligned} \tag{5.43}$$

To get to the last line we used the inequality $-\log(1-\frac{e-1}{e}y)\le y,\ \forall y\in[0,1]$ with $y=e^2gh|\beta|$. Notice that $\beta_0=\frac{1}{5egh\kappa}$, which means $y=e^2gh|\beta|\le 1$ for $|\beta|\le\beta_0$.

This concludes the induction step and also the proof of the theorem. □

Lemma 144. *Consider the same setup as Theorem 137. The following bound holds for any $x_0\in\Lambda$:*

$$\sum_{\substack{\mathcal{X}:x_0\in\mathcal{X}\\ \mathcal{X}\ \textit{is connected}}} \left(e^{|\beta|h}-1\right)^{|\mathcal{X}|} e^{ghe^2|\beta||\operatorname{supp}(\mathcal{X})|} \le e(e-1)gh|\beta|, \quad |\beta|\le\beta_0. \tag{5.44}$$

Proof of Lemma 144. Since for a connected set $\mathcal{X}$, both its size $|\mathcal{X}|$ and the size of its support $|\operatorname{supp}(\mathcal{X})|$ show up in the summation, we need to take extra care in finding a proper upper bound. We achieve this for any lattice size $|\Lambda|=n$, this time by induction over the size of $|\mathcal{X}|$. It is not hard to check that the bound (5.44) holds for connected sets of size $|\mathcal{X}|=1$. Next, we assume by induction that (5.44) holds for connected sets $\mathcal{X}$ with $|\mathcal{X}|\le a-1$ for $a\ge 2$, and derive (5.44) for connected sets $|\mathcal{X}|\le a$. To this end, we begin with restating the sum in (5.44) in a different form. This includes adding the contribution of all connected sets $\mathcal{X}$ that contain a site x_0 in the following order.

First, we consider the contribution of a fixed set $X\subset\Lambda$ with size and diameter at most κ and R that contains x_0. We then sum over all the connected sets that include a site $x\in X$. It is not hard to see that by selecting all possible choices of X and performing the addition in this way, we *overcount* the number of connected sets $\mathcal{X}$ that contain x_0, and therefore get an upper bound on the original sum in (5.44). More formally, for any $x_0\in\Lambda$ and connected sets $\mathcal{X}$ of size at most $a>1$, we have

$$
\begin{aligned}
&\sum_{\substack{\mathcal{X}:x_0\in\mathcal{X}\\ \mathcal{X}\text{ is connected}\\ |\mathcal{X}|\le a}} \left(e^{|\beta|h}-1\right)^{|\mathcal{X}|} e^{ghe^2|\beta||\operatorname{supp}(\mathcal{X})|} \\
&\le \sum_{\substack{X:x_0\in X\\ |X|\le\kappa\\ \operatorname{diam}(X)\le R}} \left(\left(e^{|\beta|h}-1\right) e^{ghe^2|\beta||X|} \prod_{x\in X}\left(1+\sum_{\substack{\mathcal{X}':x\in\mathcal{X}'\\ \mathcal{X}'\text{ is connected}\\ |\mathcal{X}'|\le a-1}} \left(e^{|\beta|h}-1\right)^{|\mathcal{X}|} e^{ghe^2|\beta||\operatorname{supp}(\mathcal{X})|}\right)\right) \\
&\le \sum_{\substack{X:x_0\in X\\ |X|\le\kappa\\ \operatorname{diam}(X)\le R}} \left(\left(e^{|\beta|h}-1\right) e^{ghe^2|\beta||X|} \prod_{x\in X}\left(1+e(e-1)gh|\beta|\right)\right) \\
&\le g\left(e^{|\beta|h}-1\right)\left(e^{ghe^2|\beta|}\left(1+e(e-1)gh|\beta|\ \right)\right)^{\kappa} && (5.45)\\
&\le (e-1)g|\beta|he^{e(2e-1)gh|\beta|\kappa} && (5.46)\\
&\le e(e-1)g|\beta|h, && (5.47)
\end{aligned}
$$

where we used the induction hypothesis (which assumes (5.44) holds for the connected set $\mathcal{X}'$ with $|\mathcal{X}'|\le a-1$) to get from the second to the third line. Eq. (5.45) follows from the definition of the growth constant g which gives an upper bound on the number of sets X containing x_0 with size at most κ. To get to Eq. (5.46) and (5.47), we use the fact that $1+y\le e^y$, $e^y-1\le(e-1)y$ for $y\in[0,1]$ and $|\beta|\le\frac{1}{5eghe\kappa}$. □

5.5 Analyticity implies exponential decay of correlations

In this section, we show that the exponential decay of correlations is a *necessary* condition for the free energy to be analytic and bounded close to the real axis. Our bounds are stronger for commuting Hamiltonians on arbitrary lattices and non-commuting Hamiltonians on a 1D chain and slightly weaker for generic geometrically-local cases.

Similar to the rest of this paper, our general strategy uses extrapolation between different regimes of the inverse temperature parameter. We know that at $\beta=0$, the Gibbs state is just the maximally mixed state, so the decay of correlations property trivially holds. Additionally, we show that at $\beta=0$, the low-order derivatives of a function that encode the amount of correlation between two regions are zero. This combined with the absence of singularities coming from the analyticity condition puts an exponentially small bound on how fast this function (i.e. the correlations) can grow with β.

The proof is reminiscent of the one for classical systems first shown by [DS87]. As explained earlier, the essence of the proof is the following simple lemma from complex analysis.

Lemma 145 (cf. [DS87])**.** *Let $f(z_1,\dots,z_m)$ be a complex function that on a bounded connected open region $\Omega\subset\mathbb{C}^m$ is analytic and $|f(z_1,\dots,z_m)|\le M$. Let $k_1,\dots,k_m$ be non-negative integers*

summing to K. Suppose that $f(z_1, \dots, z_m)$ and its following derivatives are zero at some $\zeta_0 \in \Omega$:

$$\frac{d^K}{d^{k_1} z_i \dots d^{k_m} z_m} f(z_1, \dots, z_m)\Big|_{\zeta_0} = 0 \quad \text{if} \quad |\{i \in [m] : k_i \geq 1\}| \leq L - 1, \tag{5.48}$$

that is, unless we take the derivative with respect to at least L distinct variables z_i, this derivative is zero at ζ_0. Then, for any $(z_1, \dots, z_m) \in \Omega$, there exist constants c_1, c_2 depending on ζ_0 and $(z_1, \dots, z_m)$ such that $|f(z_1, \dots, z_m)| \leq c_1 M e^{-c_2 L}$.

Proof of Lemma 145. Without loss of generality, we can restrict ourselves to the single variable case, $m = 1$, by defining a path parameterized by $z \in [0, 1]$ that connects ζ to any point $(z_1, \dots, z_m)$ of interest. We denote the function on this path by $f(z)$. Region Ω in this case is just a region in the complex plane around $[0, 1]$ that has a small imaginary part such that $f(z)$ remains analytic and bounded.

Using conformal mapping similar to what we did in Theorem 136, we can map the unit disk onto Ω, which is the set of $z \in \mathbb{C}$ such that $|z| \leq 1$. Hence, without loss of generality, we assume $f(z)$ is analytic on the unit disk. It is also not hard to see that Eq. (5.48) implies the first L derivatives of $f(z)$ vanish at the origin. Thus, the Taylor expansion of $f(z)$ converges and we have

$$\forall z \in \Omega, \quad f(z) = \sum_{k>L} a_k z^k = z^L \sum_{k>L} a_k z^{k-L}, \tag{5.49}$$

but $\sum_{k>L} a_k z^{k-L}$ is itself an analytic function, so it is either a constant or attains its maximum absolute value on the boundary. It follows from $|f(z)| \leq M$ that in either case $|\sum_{k>L} a_k z^{k-L}| \leq M$. This implies $\forall |z| \leq 1,\ |f(z)| \leq M|z|^L$, which in turn proves the theorem. □

The connection between Lemma 145, the decay of correlations, and the analyticity condition becomes clear once we substitute our choice of function $f(z_1, \dots, z_m)$ and region Ω. We begin by defining Ω. Fixing our choice of function $f(z_1, \dots, z_m)$ is postponed until after we discuss the precise statement of the analyticity condition and the decay of correlations.

Region Ω corresponds to the region near the real β axis where the partition function does not vanish. Given a local Hamiltonian $H = \sum_{i=1}^m H_i$, we define complex variables $z_1, \dots, z_m$ such that each z_i roughly equals β plus some small *complex deviation*. Hence, instead of working with functions of βH such as $\exp(-\beta H)$, we consider functions of $\sum_{i=1}^m z_i H_i$ as in $\exp(-\sum_{i=1}^m z_i H_i)$. For a fixed inverse temperature β and maximum deviation δ, we denote the set of such tuples $(z_1, \dots, z_m)$ by $\Gamma_{\delta,\beta}$. By varying β from zero to some constant β and taking the union of corresponding $\Gamma_{\delta,\beta}$, the set $\Omega_{\delta,\beta}$ is obtained. More precisely, we have the following definition.

Definition 146 (The vicinity of the real β axis)**.** *Let $\Gamma_{\delta,\beta}$ be the set $\{(z_1, \dots, z_m) : \forall i \in [m], z_i \in \mathbb{C}, |z_i - \beta| \leq \delta\}$. We define $\Omega_{\delta,\beta}$ to be $\Omega_{\delta,\beta} = \bigcup_{\substack{\beta' \in \mathbb{R}^+ \\ \beta' \leq \beta}} \Gamma_{\delta,\beta'}$.*

We also define the perturbed Gibbs state as follows.

Definition 147 (Complex perturbed Gibbs state)**.** *The δ-perturbed Gibbs state of a local Hamiltonian $H = \sum_i^m H_i$ at inverse temperature β is defined as*

$$\rho_{\vec{z}}(H) = \frac{e^{-\sum_{i=1}^m z_i H_i}}{\mathrm{tr}[e^{-\sum_{i=1}^m z_i H_i}]}, \quad \vec{z} = (z_1, \dots, z_m) \in \Gamma_{\delta,\beta} \tag{5.50}$$

where $\Gamma_{\delta,\beta}$ is defined in Definition 146.

The analyticity condition we consider here is stronger than the ones derived in Section 5.4 in the high temperature regime or used in the approximation algorithm in Section 5.3. Previously we only included systems with open boundary conditions in our analysis, but here we also need to allow for other boundary conditions. This is not restricted to the quantum case, and Dobrushin and Shlosman use similar conditions in their proof for classical systems [DS87]. The precise statement of our condition is the following:

Condition 1 (Analyticity after measurement)**.** *The free energy of a geometrically-local Hamiltonian H is δ-analytic at β if for any local operator $N \geq 0$ with $\|N\| = 1$, there exists a constant c such that $\forall (z_1, \ldots, z_m) \in \Gamma_{\delta,\beta}$,*

$$\left| \log \left(\operatorname{tr} \left[e^{-\sum_{i=1}^m z_i H_i} N \right] \right) \right| \leq cn. \tag{5.51}$$

The free energy is said to be δ-analytic along $[0, \beta]$ if the bound (5.51) holds for $\forall (z_1, \ldots, z_m) \in \Omega_{\delta,\beta}$

To see the motivation for this condition, first note that when restricted to classical systems, this condition reduces to the one used in [DS87]. There the operator N sets the boundary conditions which fixes the value of certain spins in the system before computing the partition function, or more generally, finding the Gibbs distribution. A natural question then is how varying these boundary conditions affects the distribution. In particular, the *uniqueness of the Gibbs distribution* refers to the case that in the limit of a large number of particles, changing distant spins has a negligible effect on the distribution of spins on a finite region. Hence, a *unique* Gibbs distribution can be defined for such systems. This condition is not satisfied at all temperatures, and below the critical temperature, multiple Gibbs distributions exist. Thus, it seems natural to include the boundary conditions in the partition function when studying its complex zeros and the critical behavior of the system in general.

For quantum systems, one can think of fixing the boundary spin values by projecting them onto a specific state or more generally by post-selecting after a local measurement has been performed. Hence, $\operatorname{tr}\left[\exp\left(-\sum_{i=1}^m z_i H_i\right) N\right]$ is the partition function of the normalized Gibbs state after conditioning on the measurement outcome associated with N. Notice that, in principle, the state of the spins after post-selection can be entangled. As we will see, this causes technical difficulties in extending the classical results to the quantum regime.

Finally, we note that the validity of Condition 1 can be shown in the high temperature regime ($\beta \leq \beta_0$) by a slight modification of the argument in Section 5.4.1.

Our goal is to show that Condition 1 on the analyticity of the free energy implies the exponential decay of correlations. This condition is stated as follows.

Condition 2 (Exponential decay of correlations)**.** *The correlations in the Gibbs state $\rho_\beta(H)$ of a geometrically-local Hamiltonian decay exponentially if for any local Hermitian operators O_1 and O_2 and any region B such that $\operatorname{supp}(O_1), \operatorname{supp}(O_2) \subset B$, there exist constants ξ and c such that*

$$\left| \operatorname{tr}\left[\rho_\beta(H) O_1 O_2\right] - \operatorname{tr}\left[\rho_\beta(H) O_1\right] \operatorname{tr}\left[\rho_\beta(H) O_2\right] \right| \leq c|B| \|O_1\| \|O_2\| e^{-\operatorname{dist}(O_1, O_2)/\xi}. \tag{5.52}$$

We first prove a slightly weaker version of Condition 2 assuming Condition 1. We then improve our bound for commuting and 1D Hamiltonians.

Theorem 148 (Analyticity implies exponential decay of correlations)**.** *Suppose the free energy of a geometrically-local Hamiltonian is δ-analytic along $[0,\beta]$ as in Condition 1 for some $\delta = O(1)$. Then the correlations between any two operators O_1, O_2 with $\mathrm{dist}(O_1,O_2) = \Omega(\log n)$ decay exponentially in the range $[0,\beta]$. More precisely,*

$$\left| \operatorname{tr}\left[\rho_\beta(H) O_1 O_2\right] - \operatorname{tr}\left[\rho_\beta(H) O_1\right] \operatorname{tr}\left[\rho_\beta(H) O_2\right] \right| \leq cne^{-\mathrm{dist}(O_1,O_2)/\xi}. \tag{5.53}$$

Proof of Theorem 148. We can without loss of generality assume $\|O_1\|, \|O_2\| \leq 1$. Let $A_1 = \mathrm{supp}(O_1)$ and $A_2 = \mathrm{supp}(O_2)$. Each of the observables O_1 and O_2 can be decomposed into two positive semi-definite (PSD) matrices: $O_1 = O_1^+ - O_1^-$ and $O_2 = O_2^+ - O_2^-$, where O_1^+, O_2^+ include the positive eigenvalues of O_1, O_2 and $-O_1^-, -O_2^-$ include the negative ones. We can write the covariance in Eq. (5.52) as

$$\begin{aligned}
&\left| \operatorname{tr}\left[\rho_\beta(H) O_1 O_2\right] - \operatorname{tr}\left[\rho_\beta(H) O_1\right] \operatorname{tr}\left[\rho_\beta(H) O_2\right] \right| \\
&= \left| \sum_{\alpha,\gamma\in\{\pm\}} \alpha\gamma \left(\operatorname{tr}\left[\rho_\beta(H) O_1^\alpha O_2^\gamma\right] - \operatorname{tr}\left[\rho_\beta(H) O_1^\alpha\right] \operatorname{tr}\left[\rho_\beta(H) O_2^\gamma\right] \right) \right| \\
&\leq 4 \cdot \max_{\substack{N_2, N_1 \geq 0: \\ \|N_2\|, \|N_1\| \leq 1}} \left| \left(\operatorname{tr}\left[\rho_\beta(H) N_2 N_1\right] - \operatorname{tr}\left[\rho_\beta(H) N_2\right] \operatorname{tr}\left[\rho_\beta(H) N_1\right] \right) \right|,
\end{aligned} \tag{5.54}$$

where $\mathrm{supp}(N_2) = A_1$ and $\mathrm{supp}(N_1) = A_2$. Recall that the post-selected state $\rho_\beta(H|N)$ is defined by

$$\rho_\beta(H|N) = \frac{\sqrt{N} \exp(-\beta H) \sqrt{N}}{\operatorname{tr}[\exp(-\beta H) N]}. \tag{5.55}$$

The bound (5.54) can be rewritten as

$$\begin{aligned}
\left| \left(\operatorname{tr}\left[\rho_\beta(H) N_2 N_1\right] - \operatorname{tr}\left[\rho_\beta(H) N_2\right] \operatorname{tr}\left[\rho_\beta(H) N_1\right] \right) \right| &= \left| \operatorname{tr}\left[\rho_\beta(H) N_2\right] \right| \left| \operatorname{tr}[\rho_\beta(H|N_2) N_1] - \operatorname{tr}\left[\rho_\beta(H|\mathbb{1}) N_1\right] \right| \\
&\leq \left| \operatorname{tr}[\rho_\beta(H|N_2) N_1] - \operatorname{tr}\left[\rho_\beta(H|\mathbb{1}) N_1\right] \right|.
\end{aligned} \tag{5.56}$$

Hence, our goal is to show

$$\left| \operatorname{tr}[\rho_\beta(H|N_2) N_1] - \operatorname{tr}\left[\rho_\beta(H|\mathbb{1}) N_1\right] \right| \leq ce^{-\mathrm{dist}(O_1,O_2)/\xi}. \tag{5.57}$$

We instead show

$$\left| \log\left(\frac{\operatorname{tr}[\rho_\beta(H|N_2) N_1]}{\operatorname{tr}\left[\rho_\beta(H|\mathbb{1}) N_1\right]} \right) \right| \leq ce^{-\mathrm{dist}(O_1,O_2)/\xi}. \tag{5.58}$$

To see why this implies (5.57), we can further upper bound the right-hand side of (5.58) using the inequality $x \leq -\log(1-x)$ for $x < 1$ and choosing $x = c\exp(-\mathrm{dist}(O_1,O_2)/\xi)$. Then, by exponentiating both sides of (5.58) and using the fact that $\operatorname{tr}(\rho_\beta(H) N_1) \leq 1$, we arrive at (5.57). We can prove a similar bound even when instead of $\mathbb{1}$ there is any other PSD operator in the denominator. One way to interpret these bounds is that a local measurement on region A_2 is undetected from the perspective of local operators on region A_1.

The proof follows from Lemma 145. We first consider a perturbed version of (5.58) using

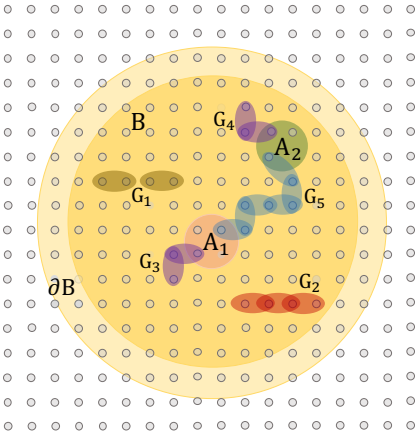


Figure 5-3: To study the correlations between regions A_1, A_2, we can restrict the Gibbs state to region B while adding an operator on the boundary ∂B to include the effect of the rest of the lattice. Regions $G_1, \ldots, G_5$ show up when studying the derivatives of the correlation function. See the proofs of Theorem 148 and Theorem 149.

Definition 147. We define $f(z_1, \ldots, z_m)$ as

$$\begin{aligned} f(z_1, \ldots, z_m) &= \log\left(\frac{\mathrm{tr}[\rho_{\vec{z}}(H|N_2)N_1]}{\mathrm{tr}[\rho_{\vec{z}}(H|\mathbb{1})N_1]}\right) \\ &= \log\left(\frac{\mathrm{tr}[e^{-\sum_{i=1}^m z_i H_i} N_2 N_1]}{\mathrm{tr}[e^{-\sum_{i=1}^m z_i H_i} N_2]} \frac{\mathrm{tr}[e^{-\sum_{i=1}^m z_i H_i}]}{\mathrm{tr}[e^{-\sum_{i=1}^m z_i H_i} N_1]}\right). \end{aligned} \tag{5.59}$$

This function is our choice for $f(z_1, \ldots, z_m)$ in Lemma 145. In particular, we prove that assuming Condition 1 is satisfied, $f(z_1, \ldots, z_m)$ is analytic in $\Omega_{\delta,\beta}$, has a bounded absolute value, and has vanishing derivatives at $z_1 = \cdots = z_m = 0$. Let us begin with the analyticity and boundedness.

Analyticity and boundedness: From (5.51) we see that for any positive operator N, the post-selected free energy is analytic and there exists some constant c such that

$$\left|\log\left(\mathrm{tr}\left[e^{-\sum_{i=1}^m z_i H_i} N\right]\right)\right| \leq cn, \tag{5.60}$$

By using a proper choice for N, we see that $f(z_1, \ldots, z_m)$ is a sum of analytic functions and therefore is itself analytic. We also get an upper bound on $|f(z_1, \ldots, z_m)|$, that is,

$$\begin{aligned} \forall (z_1, \ldots, z_m) \in \Omega_{\delta,\beta_c}, \quad |f(z_1, \ldots, z_m)| &\leq \left|\log\left(\mathrm{tr}[e^{-\sum_{i=1}^m z_i H_i} N_2 N_1]\right)\right| + \left|\log\left(\mathrm{tr}[e^{-\sum_{i=1}^m z_i H_i} N_2]\right)\right| \\ &\quad + \left|\log\left(\mathrm{tr}[e^{-\sum_{i=1}^m z_i H_i} N_1]\right)\right| + \left|\log\left(\mathrm{tr}[e^{-\sum_{i=1}^m z_i H_i}]\right)\right| \\ &\leq 4cn. \end{aligned} \tag{5.61}$$

Vanishing derivatives: It remains to show that certain derivatives of $f(z_1, \dots, z_m)$ are zero at the point $\beta = 0$, which is inside $\Omega_{\delta,\beta}$. The derivatives of $f(z_1, \dots z_m)$ are combinations of terms like

$$\frac{d^K}{d^{k_1} z_i \dots d^{k_m} z_m} \log \left(\operatorname{tr} \left[e^{-\sum_{i=1}^m z_i H_i} N_2 N_1 \right] \right) \Big|_{z=0}, \tag{5.62}$$

where $k_i \geq 0$ and $K = \sum_{i=1}^m k_i$. Notice that we are including the z_i that are not in the derivative by letting $k_i = 0$. We claim in certain instances that these terms are either zero or cancel each other. Consider all the local terms H_i that we are taking a derivative with respect to their z_i, i.e. $k_i \geq 1$. We denote the union of the support of these terms by G. Recall that A_1, A_2 are the support of O_1, O_2, respectively. Region G fits into (a union of) one of the following cases.

Case 1: G is not connected and does not intersect with $A_1 \cup A_2$ (see G_1 in Figure 5-3 for an example). In this case, the terms in the derivatives are

$$\begin{aligned}
\frac{d^K}{d^{k_1} z_i \dots d^{k_m} z_m} \log \left(\operatorname{tr} \left[e^{-\sum_{i=1}^m z_i H_i} N_2 N_1 \right] \right) \Big|_{z=0} &= \frac{d^K}{d^{k_1} z_i \dots d^{k_m} z_m} \log \left(\operatorname{tr}[N_2] \operatorname{tr}[N_1] \prod_{i:k_i \geq 1} \operatorname{tr} \left[e^{-z_i H_i} \right] \right) \Big|_{z=0} \\
&= \frac{d^K}{d^{k_1} z_i \dots d^{k_m} z_m} \log \left(\operatorname{tr}[N_2] \operatorname{tr}[N_1] \right) \\
&+ \sum_{i:k_i \geq 1} \frac{d^K}{d^{k_1} z_i \dots d^{k_m} z_m} \log \left(\operatorname{tr} \left[e^{-z_i H_i} \right] \right) \Big|_{z=0} \\
&= 0.
\end{aligned} \tag{5.63}$$

In the first line, we used the fact that sublattices A_1, A_2, and $\operatorname{supp}(H_i)$ with $k_i \geq 1$ do not intersect. The last line follows because $\operatorname{tr}[N_2] \operatorname{tr}[N_1]$ is a constant, and $\operatorname{tr} \left[e^{-z_i H_i} \right]$ only depends on z_i and its derivative with respect to other z_i is zero.

Case 2: G is connected but does not intersect with $A_1 \cup A_2$ (see G_2 in Figure 5-3 for an example). Similar to (5.63), we can still separate $\operatorname{tr}[N_2] \operatorname{tr}[N_1]$ from the remaining terms and their derivative is zero. Hence, we obtain

$$\frac{d^K}{d^{k_1} z_i \dots d^{k_m} z_m} \log \left(\operatorname{tr} \left[e^{-\sum_{i=1}^m z_i H_i} N_2 N_1 \right] \right) \Big|_{z=0} = \frac{d^K}{d^{k_1} z_i \dots d^{k_m} z_m} \log \left(\operatorname{tr} \left[e^{-\sum_{i:k_i \geq 1} z_i H_i} \right] \right) \Big|_{z=0}. \tag{5.64}$$

Although this term does not necessarily equal zero, the derivatives of $f(z_1, \dots, z_m)$ are combinations of terms like Eq. (5.64). These terms are all equal as we can separate traces involving N_2 and N_1 using the same argument as above, but they appear with opposite signs and thus cancel each other.

More precisely, we have

$$\begin{aligned}
&\frac{d^K}{d^{k_1}z_i \dots d^{k_m}z_m} f(z_1,\dots,z_m)\Big|_{z=0} \\
&= \frac{d^K}{d^{k_1}z_i \dots d^{k_m}z_m}\Big(\log\Big(\mathrm{tr}[N_2N_1]\,\mathrm{tr}\left[e^{-\sum_{i:k_i\geq 1} z_iH_i}\right]\Big) - \log\Big(\mathrm{tr}[N_2]\,\mathrm{tr}\left[e^{-\sum_{i:k_i\geq 1} z_iH_i}\right]\Big) \\
&- \log\Big(\mathrm{tr}[N_1]\,\mathrm{tr}\left[e^{-\sum_{i:k_i\geq 1} z_iH_i}\right]\Big) + \log\Big(\mathrm{tr}\left[e^{-\sum_{i:k_i\geq 1} z_iH_i}\right]\Big)\Big)\Big|_{z=0} \\
&= 0. \qquad (5.65)
\end{aligned}$$

Case 3: G is connected and intersects with A_1 or A_2, but not both (see G_3 or G_4 in Figure 5-3 for an example). Similar to Case 2, the derivatives of $f(z_1,\dots,z_m)$ consist of equal terms with opposite signs and therefore vanish. Here, we show the case where G only intersects A_2. The other cases similarly follow.

$$\begin{aligned}
&\frac{d^K}{d^{k_1}z_i \dots d^{k_m}z_m} f(z_1,\dots,z_m)\Big|_{z=0} \\
&= \frac{d^K}{d^{k_1}z_i \dots d^{k_m}z_m}\Big(\log\Big(\mathrm{tr}[N_1]\,\mathrm{tr}\left[e^{-\sum_{i:k_i\geq 1} z_iH_i}N_2\right]\Big) - \log\Big(\mathrm{tr}\left[e^{-\sum_{i:k_i\geq 1} z_iH_i}N_2\right]\Big) \\
&- \log\Big(\mathrm{tr}[N_1]\,\mathrm{tr}\left[e^{-\sum_{i:k_i\geq 1} z_iH_i}\right]\Big) + \log\Big(\mathrm{tr}\left[e^{-\sum_{i:k_i\geq 1} z_iH_i}\right]\Big)\Big)\Big|_{z=0} \\
&= 0, \qquad (5.66)
\end{aligned}$$

in which the first two and last two terms cancel each other.

A similar argument shows that if the set G is a union of any of the above cases, the derivatives vanish as well.

Case 4: G is connected and intersects with both A_1 and A_2 (see G_5 in Figure 5-3 for an example). Here, the cancellation that appeared in the other cases does not happen. Thus, this is the only situation in which the derivatives are non-zero.

The important observation is that for Case 4 to happen, G needs to be *long* enough to touch both A_1 and A_2. Hence, if the number of z_i with $k_i \geq 1$ is less than roughly $\mathrm{dist}(O_1, O_2)$, their corresponding derivative vanishes. Having all the criteria needed for applying Lemma 145, i.e analyticity, boundedness, and zero derivatives, we can get the following bound on $|f(z_1,\dots,z_m)|$ for some constant c and ξ:

$$|f(\beta,\dots,\beta)| \leq cne^{-\mathrm{dist}(O_1,O_2)/\xi}, \qquad (5.67)$$

which as explained before implies

$$\left|\,\mathrm{tr}\left[\rho_\beta(H)O_1O_2\right] - \mathrm{tr}\left[\rho_\beta(H)O_1\right]\mathrm{tr}\left[\rho_\beta(H)O_2\right]\,\right| \leq cne^{-\mathrm{dist}(O_1,O_2)/\xi}. \qquad (5.68)$$

Due to the extra factor of n in front of this bound, it implies the exponential decay of correlations only when $\mathrm{dist}(O_1, O_2) = \Omega(\log n)$. □

5.5.1 Tighter bounds for commuting Hamiltonians

Here we show how using the commutativity of H enables us to remove the extra factor of n in the bound (5.67) that we derived for general Hamiltonians. We state this in the following theorem.

Theorem 149. *Suppose H is a geometrically-local Hamiltonian with mutually commuting terms that satisfies Condition 1 for along $[0,\beta]$. Then the correlations between any two operators O_1, O_2 decay exponentially in the range $[0,\beta]$ as in Condition 2.*

Proof of Theorem 149. The proof of this theorem follows similar steps to that of Theorem 148. A crucial difference, which is the only part where we use the commutativity of local terms H_i, is the following. The Hamiltonian H in states $\rho_\beta(H|N_2)$ and $\rho_\beta(H|\mathbb{1})$ involves terms acting on all n sites in lattice Λ. In our analysis, we can essentially neglect the contribution of sites that are far from both region A_1 and A_2. In other words, as shown in Figure 5-3, let $B \subset \Lambda$ be any region that encloses the support of both A_1 and A_2. We restrict the Hamiltonian and states $\rho_\beta(H|N_2), \rho_\beta(H|\mathbb{1})$ to this region and include the effect of other sites by an operator acting on ∂B, the boundary of this enclosing region. We prove (5.58) for this smaller region. Without this step, we end up getting an upper bound like $cn\exp(-\text{dist}(O_1,O_2)/\xi)$, which has an extra factor of n, the number of sites in Λ, whereas with the restriction to the enclosing region, this factor is $|B|$, the number of sites in B. For the example in Figure 5-3, the size $|B| = O(\text{dist}(O_1,O_2)^D)$ which is negligible compared to the exponential decay factor $e^{-\text{dist}(O_1,O_2)/\xi}$. More formally, since H is a commuting Hamiltonian, we have $e^{-\beta H} = e^{-\beta H_B} e^{-\beta(H-H_B)}$. Hence, we get

$$\begin{aligned}
\text{tr}[\rho_\beta(H|N_2)N_1] &= \frac{\text{tr}[e^{-\beta H}N_2N_1]}{\text{tr}[e^{-\beta H}N_2]} \\
&= \frac{\text{tr}_B[e^{-\beta H_B}\,\text{tr}_{\bar{B}}[e^{-\beta(H-H_B)}]N_2N_1]}{\text{tr}_B[e^{-\beta H_B}\,\text{tr}_{\bar{B}}[e^{-\beta(H-H_B)}]N_2]} \\
&= \frac{\text{tr}_B[e^{-\beta H_B}\sigma N_2N_1]}{\text{tr}_B[e^{-\beta H_B}\sigma N_2]} \\
&= \text{tr}[\rho_\beta(H_B|\sigma N_2)N_1], \qquad (5.69)
\end{aligned}$$

where

$$\sigma = \frac{\text{tr}_{\bar{B}}[e^{-\beta(H-H_B)}]}{\text{tr}_{\bar{B}\cup\partial\bar{B}}[e^{-\beta(H-H_B)}]} \qquad (5.70)$$

is a state acting on the boundary $\partial\bar{B}$ [4]. Thus, we can replace the operator N_2 by $\sigma\otimes N_2$ acting on a larger region $\partial\bar{B}\cup A_2$ and restrict our attention to region B. We can now repeat the argument of Theorem 148. Let the perturbed Hamiltonian restricted to region B be $H_B(\vec{z}) = \sum_{H_i:\text{supp}(H_i)\subset B} z_i H_i$, where for simplicity, the number of local terms in H_B is denoted again by m. By plugging (5.69)

[4] Based on our definition of the boundary of a region, the boundary $\partial\bar{B}$ is inside B.

into (5.59), we see that the function $f(z_1, \ldots, z_m)$ is

$$\begin{aligned} f(z_1, \ldots, z_m) &= \log \left(\frac{\mathrm{tr}[\rho_{\vec{z}}(H_B|\sigma N_2)N_1]}{\mathrm{tr}[\rho_{\vec{z}}(H_B|\sigma)N_1]} \right) \\ &= \log \left(\frac{\mathrm{tr}[e^{-\beta H_B(\vec{z})}\sigma N_2 N_1]}{\mathrm{tr}[e^{-H_B(\vec{z})}\sigma N_2]} \frac{\mathrm{tr}[e^{-\beta H_B(\vec{z})}\sigma]}{\mathrm{tr}[e^{-\beta H_B(\vec{z})}\sigma N_1]} \right) \end{aligned} \tag{5.71}$$

The rest of the proof of Theorem 148 applies to this function. In particular, assuming Condition 1 holds, this function is bounded and analytic in $\Omega_{\delta,\beta}$, i.e. $|f(z_1, \ldots, z_m)| \leq c|B|$. Similarly, one can see that the low-order derivatives of $f(z_1, \ldots, z_m)$ are zero. Since the distance between ∂B and A_1 is still $O(\mathrm{dist}(O_1, O_2))$, Lemma 145 implies $|f(\beta, \ldots, \beta)| \leq c|B| \exp(-\mathrm{dist}(O_1, O_2)/\xi)$. Hence, we have

$$\left| \mathrm{tr}\left[\rho_\beta(H)O_1O_2\right] - \mathrm{tr}\left[\rho_\beta(H)O_1\right] \mathrm{tr}\left[\rho_\beta(H)O_2\right] \right| \leq c|B| e^{-\mathrm{dist}(O_1,O_2)/\xi} \tag{5.72}$$

□

5.5.2 Tighter bounds for 1D Hamiltonians

Theorem 150. *Let H be a geometrically-local Hamiltonian on a 1D chain that satisfies Condition 1 along $[0, \beta]$. Then, the exponential decay of correlations given in Condition 2 also holds for this Hamiltonian in the range $[0, \beta]$.*

Proof of Theorem 150. The proof is similar to that of Theorem 148 and Theorem 149. Recall that an important step is to introduce boundary states σ that include the effect of terms in the Hamiltonian H that are acting on the boundary or outside of some region B. Region B encloses the support of operators whose correlations we want to bound. There, we use the commutativity of H to find the boundary states σ which does not hold in general. Here, we show how, by using the quantum belief propagation operator η we introduced in Proposition 131, we can achieve the same boundary state in 1D.

We do not go through all steps of the proof of Theorem 148 again. Instead, we directly show that by restricting the Hamiltonian to region B and adding the boundary terms, the covariance in (5.52) changes negligibly. Then we apply bound (5.53) to this restricted covariance. Since the number of particles inside B is constant, instead of the extra factor of n, we get a constant prefactor as desired.

Recall that using the belief propagation equation (5.14) and the bound (5.15), we can remove the boundary terms $H_{\partial B}$ acting between $B, \bar{B}$ from the Gibbs state and get

$$\begin{aligned} \mathrm{tr}[\rho_\beta(H)O_1O_2] &= \mathrm{tr}\left[\frac{Z_\beta(H - H_{\partial B})}{Z_\beta(H)} \eta \rho_\beta(H - H_{\partial B}) \eta^\dagger O_1 O_2 \right] \\ &= \mathrm{tr}\left[\frac{Z_\beta(H - H_{\partial B})}{Z_\beta(H)} \eta_\ell \rho_\beta(H - H_{\partial B}) \eta_\ell^\dagger O_1 O_2 \right] \\ &+ \mathrm{tr}\left[\frac{Z_\beta(H - H_{\partial B})}{Z_\beta(H)} \eta_\ell \rho_\beta(H - H_{\partial B}) (\eta^\dagger - \eta_\ell^\dagger) O_1 O_2 \right] \\ &+ \mathrm{tr}\left[\frac{Z_\beta(H - H_{\partial B})}{Z_\beta(H)} (\eta - \eta_\ell) \rho_\beta(H - H_{\partial B}) \eta^\dagger O_1 O_2 \right], \end{aligned} \tag{5.73}$$

where in the second line, we replaced η with the truncated operator η_ℓ. To simplify this equation, we absorb the coefficient $Z_\beta(H - H_{\partial B})/Z_\beta(H)$ into the operators η, η_ℓ, and define

$$\tilde{\eta} = \left(\frac{Z_\beta(H - H_{\partial B})}{Z_\beta(H)}\right)^{1/2} \eta, \quad \tilde{\eta}_\ell = \left(\frac{Z_\beta(H - H_{\partial B})}{Z_\beta(H)}\right)^{1/2} \eta_\ell. \tag{5.74}$$

Hence, we have

$$\begin{aligned}\left|\mathrm{tr}\left[\rho_\beta(H)O_1O_2\right] - \mathrm{tr}\left[\tilde{\eta}_\ell\rho_\beta(H - H_{\partial B})\tilde{\eta}_\ell^\dagger O_1O_2\right]\right| \leq{}& \left|\mathrm{tr}\left[\tilde{\eta}\rho_\beta(H - H_{\partial B})(\tilde{\eta}^\dagger - \tilde{\eta}_\ell^\dagger)O_1O_2\right]\right| \\ &+ \left|\mathrm{tr}\left[(\tilde{\eta} - \tilde{\eta}_\ell)\rho_\beta(H - H_{\partial B})\tilde{\eta}^\dagger O_1O_2\right]\right|. \end{aligned} \tag{5.75}$$

According to (5.15), we have $\|\eta - \eta_\ell\| \leq e^{\alpha_1|\partial B| - \alpha_2\ell}$ and $\|\eta\| \leq e^{\beta/2\|H_{\partial B}\|}$. Also, Lemma 139 implies $Z_\beta(H - H_{\partial B})/Z_\beta(H) \leq e^{\alpha_3|\partial B|}$ for some constant α_3 that depends on the details of H. Using these bounds as well as the Cauchy-Schwarz and Hölder inequalities, we get the following bound for some constants c' and α_4:

$$\begin{aligned}\left|\mathrm{tr}\left[\rho_\beta(H)O_1O_2\right] - \mathrm{tr}\left[\tilde{\eta}_\ell\rho_\beta(H - H_{\partial B})\tilde{\eta}_\ell^\dagger O_1O_2\right]\right| &\leq 2\|O_1\|\|O_2\|\|\eta - \eta_\ell\|\|\eta\| \\ &\leq c'e^{-\alpha_4\ell}. \end{aligned} \tag{5.76}$$

To arrive at the last line, we used the fact that $|\partial B|$ in 1D is just a constant that depends on the range of H, and we assumed the truncation length ℓ is sufficiently larger than $|\partial B|$.

Note that since we removed the boundary terms $H_{\partial B}$, the Gibbs state decomposes into $\rho_\beta(H - H_{\partial B}) = \rho_\beta(H_{\bar{B}})\rho_\beta(H_B)$, which allows us to write

$$\mathrm{tr}\left[\tilde{\eta}_\ell\rho_\beta(H - H_{\partial B})\tilde{\eta}_\ell^\dagger O_1O_2\right] = \mathrm{tr}\left[\rho_\beta(H_B)\tilde{\sigma}_{\partial B}O_1O_2\right], \tag{5.77}$$

in which we assume region B is chosen to be wide enough so that both O_1, O_2 are sufficiently far from the boundary ∂B compared to length ℓ. This means η_ℓ does not overlap with O_1, O_2. We also define the *unnormalized* boundary state $\tilde{\sigma}_{\partial B}$ by

$$\tilde{\sigma}_{\partial B} = \tilde{\eta}_\ell^\dagger\tilde{\eta}_\ell \,\mathrm{tr}_{\bar{B}\setminus\mathrm{supp}(\eta_\ell)}[\rho_\beta(H_{\bar{B}})]. \tag{5.78}$$

Notice that $\tilde{\sigma}_{\partial B}$ is a PSD matrix. To see why, we use the fact that $\mathrm{tr}_{\bar{B}\setminus\mathrm{supp}(\eta_\ell)}[\rho_\beta(H_{\bar{B}})]$ is a PSD matrix and hence can be written as $WW^\dagger$ for some operator W supported on $\mathrm{supp}(\eta_\ell) \cap \bar{B}$. Then it is not hard to see that for any state $|\phi\rangle$, we have

$$\langle\phi|\tilde{\sigma}_{\partial B}|\phi\rangle = \sum_{i=1}^{\dim(\mathrm{supp}(W))} \langle i|W^\dagger\langle\phi|\tilde{\eta}_\ell^\dagger\tilde{\eta}_\ell W|i\rangle|\phi\rangle \geq 0. \tag{5.79}$$

Overall, we have

$$\left|\mathrm{tr}\left[\rho_\beta(H)O_1O_2\right] - \mathrm{tr}\left[\rho_\beta(H_B)\tilde{\sigma}_{\partial B}O_1O_2\right]\right| \leq c'e^{-\alpha_4\ell}. \tag{5.80}$$

Similarly, we can replace $\mathrm{tr}[\rho_\beta(H)O_i]$ with $\mathrm{tr}[\rho_\beta(H_B)\tilde{\sigma}_{\partial B}O_i]$ up to an exponentially small error in

ℓ,

$$|\mathrm{tr}\left[\rho_\beta(H)O_i\right] - \mathrm{tr}\left[\rho_\beta(H_B)\tilde{\sigma}_{\partial B}O_i\right]| \leq c'e^{-\alpha_4\ell}, \quad i \in \{1,2\}. \tag{5.81}$$

We can now plug these expressions into the covariance (5.52). Since $\|\mathrm{tr}[\rho_\beta(H_B)\tilde{\sigma}_{\partial B}O_i]\|$ is just a constant, we see that there exist constants c'' and α_5 such that

$$\begin{aligned}
&\left|\mathrm{tr}\left[\rho_\beta(H)O_1O_2\right] - \mathrm{tr}\left[\rho_\beta(H)O_1\right]\mathrm{tr}\left[\rho_\beta(H)O_2\right]\right| \\
&= \left|\left(\mathrm{tr}\left[\rho_\beta(H_B)\tilde{\sigma}_{\partial B}O_1O_2\right] + c'e^{-\alpha_4\ell}\right) - \left(\mathrm{tr}\left[\rho_\beta(H_B)\tilde{\sigma}_{\partial B}O_1\right] - c'e^{-\alpha_4\ell}\right)\left(\mathrm{tr}\left[\rho_\beta(H)\tilde{\sigma}_{\partial B}O_2\right] - c'e^{-\alpha_4\ell}\right)\right| \\
&\leq \left|\mathrm{tr}\left[\rho_\beta(H_B)\tilde{\sigma}_{\partial B}O_1O_2\right] - \mathrm{tr}\left[\rho_\beta(H_B)\tilde{\sigma}_{\partial B}O_1\right]\mathrm{tr}\left[\rho_\beta(H)\tilde{\sigma}_{\partial B}O_2\right]\right| + c''e^{-\alpha_5\ell}.
\end{aligned} \tag{5.82}$$

Using a similar approach that led to the bound (5.53) proved in Theorem 148, we get

$$\left|\mathrm{tr}\left[\rho_\beta(H_B)\tilde{\sigma}_{\partial B}O_1O_2\right] - \mathrm{tr}\left[\rho_\beta(H_B)\tilde{\sigma}_{\partial B}O_1\right]\mathrm{tr}\left[\rho_\beta(H)\tilde{\sigma}_{\partial B}O_2\right]\right| \leq c|B|e^{-\mathrm{dist}(O_1,O_2)/\xi}. \tag{5.83}$$

Combined with (5.82), we have

$$\left|\mathrm{tr}\left[\rho_\beta(H)O_1O_2\right] - \mathrm{tr}\left[\rho_\beta(H)O_1\right]\mathrm{tr}\left[\rho_\beta(H)O_2\right]\right| \leq c|B|e^{-\mathrm{dist}(O_1,O_2)/\xi} + c''e^{-\alpha_5\ell}. \tag{5.84}$$

Since all the coefficients in the bound on the right-hand side are constants, it suffices to choose ℓ large enough so that it is negligible compared to the $|B|e^{-\mathrm{dist}(O_1,O_2)/\xi}$ term. This is possible because we assumed ∂C is sufficiently (but still only constantly) far from O_1, O_2. This allows us to get a bound that does not depend on n as before, hence finishing the proof. □

Remark 151. *Recall that from (5.15) we know that the error of truncating the belief propagation operator η is*

$$\left\|\eta - \eta_\ell\right\| \leq e^{\alpha_1|\partial B| - \alpha_2\ell}. \tag{5.85}$$

In our setting, the dependence of the error bound on $e^{\alpha_1|\partial B|}$ makes this result only be applicable when Λ is a $1D$ lattice. Otherwise, since $|\partial B|$ is proportional to $\mathrm{diam}(B)^{D-1}$, we cannot choose length ℓ small enough compared to $\mathrm{diam}(B)$. Hence, we do not get a local operator as required.

5.6 Exponential decay of correlations implies analyticity

In this section, we focus on the converse of Theorem 148. In Section 5.5, we showed that the exponential decay of correlations is a *necessary* condition for the analyticity of the free energy. In this section, we ask if this condition is also *sufficient* for the analyticity. This was first established for classical systems by Dobrushin and Shlosman [DS87]. It appears that the quantum generalization of that proof requires the development of new tools. The goal in this section is to identify these tools. Our contribution is to extend the result of [DS87] to classical systems that are not translationally invariant and express the proof in a language that is suitable for the quantum case.

Here, for clarity, we consider a simpler version of Condition 1 that is stated below:

Condition 1' (Analyticity of the free energy)**.** *The free energy of a geometrically-local Hamiltonian H is δ-analytic at inverse temperature $\beta \in \mathbb{R}^+$ if for all $\beta' \in \mathbb{C}$ such that $|\beta' - \beta| \leq \delta$, the free*

energy is analytic and there exists a constants c such that

$$\left|\log\left(\operatorname{tr}\left[e^{-\beta' H}\right]\right)\right| \leq cn. \tag{5.86}$$

Recall that in Condition 1, we assumed that the free energy of a *post-selected* state is analytic and bounded. In comparison, Condition 1' only includes partition functions with an *open boundary condition*. For algorithmic purposes, like the one in Section 5.3, this version is sufficient. However, with small modifications, the same proof can be adapted to show Condition 1 with arbitrary boundary conditions.

Our goal is to derive Condition 1' assuming that the correlations in the system decay exponentially. We restate this condition for convenience.

Restatement of Condition 2. *The correlations in the Gibbs state $\rho_\beta(H)$ of a geometrically-local Hamiltonian decay exponentially if for any local Hermitian operators O_1 and O_2 and any region B such that $\operatorname{supp}(O_1), \operatorname{supp}(O_2) \subset B$, there exist constants ξ and c such that*

$$\left|\operatorname{tr}\left[\rho_\beta(H) O_1 O_2\right] - \operatorname{tr}\left[\rho_\beta(H) O_1\right] \operatorname{tr}\left[\rho_\beta(H) O_2\right]\right| \leq c|B| \|O_1\| \|O_2\| e^{-\operatorname{dist}(O_1, O_2)/\xi}. \tag{5.87}$$

Remark 152. *Technically, we use a stronger version of Condition 2 in our proof which is often called strong spatial mixing. Under this condition, when we set two boundary conditions, their effect on a distant site decays exponentially with the distance from the part of the boundary where the two different boundary conditions differ rather than with the distance from the whole boundary.*

Although we consider classical systems, we find it more convenient to continue using quantum notation. This also makes it easier to point out where the proof breaks for quantum systems. The reader, however, should note that the terms in the Hamiltonian are all diagonal in a product basis and the projector operators we use basically *fix the value* of classical spins.

More formally, we prove the following theorem in this section.

Theorem 153 (The decay of correlations implies analyticity for classical systems)**.** *Let $H = \sum_{i=1}^{m} H_i$ be a geometrically-local Hamiltonian of a classical spin system, i.e. the local terms H_i are all diagonal in the same product basis. For such a system, the exponential decay of correlations given in Condition 2 implies analyticity of the free energy in Condition 1'.*

We prove this theorem in multiple steps that are formulated in Propositions 154, 155, and 157. An outline of the proof is given in Figure 5-4. It turns out that Proposition 154 and Proposition 155 continue to hold for commuting Hamiltonians, so we give their statements and proofs for these Hamiltonians. However, for reasons to be highlighted in its proof, Proposition 157 only holds for classical systems.

Proof of Theorem 153. The proof is immediate from the combination of Proposition 155, Proposition 154, and Proposition 157. □

5.6.1 Step 1: Condition 1' from the complex site removal bound

Our first step, stated in Proposition 154, is to show how a variant of the complex site removal bound that we discussed in Section 5.4 allows us to find an upper bound on the absolute value of the free energy as in Condition 1'. Compared to the bound (5.25) in Section 5.4, this variant

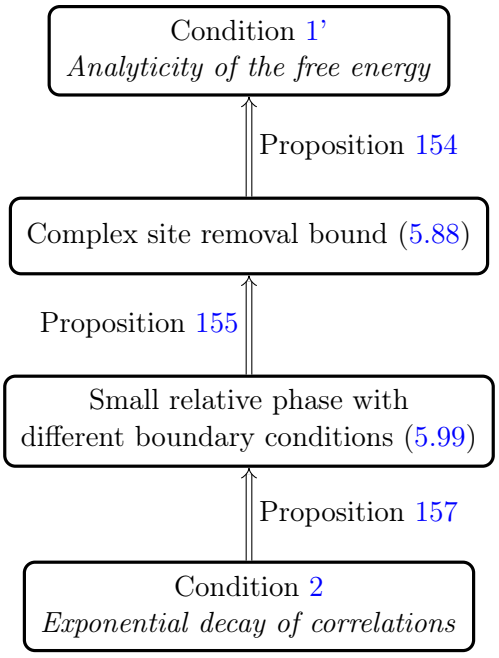


Figure 5-4: The structure of the proof of Theorem 153. We follow a series of reductions to show Condition 1'.

includes setting a non-trivial boundary condition after removing a subset of lattice sites. To avoid subtleties arising from entangled boundary conditions and projectors, we need to give a slightly different proof compared to what we did before (5.24).

Proposition 154 (Condition 1' from the complex site removal bound)**.** *Let $H = \sum_{k=1}^{m} H_k$ be a geometrically-local Hamiltonian with mutually commuting terms on lattice Λ. Let P be a projector acting on $\partial\bar{A}$ where $A \subset \Lambda$ is a region of constant size*[5]*. We denote the terms in H acting on $\bar{A}$ or $\partial\bar{A}$ by H' and the real and imaginary parts of $\beta \in \mathbb{C}$ by β_r and β_i. Suppose when $|\beta_i| \leq \delta$ for some sufficiently small δ, there exists a constant c such that*

$$\left|\log\left(\frac{\mathrm{tr}_{\bar{A}\cup A}[e^{-\beta H}]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[e^{-\beta H'}P]}\right)\right| \leq c. \tag{5.88}$$

Then,

i. The observables supported on A like H_A have bounded expectations with respect to the complex perturbed Gibbs state $\rho_\beta(H)$. That is, there exists a constant c' such that $|\,\mathrm{tr}\,[H_A\rho_\beta(H)]\,| \leq c'\|H_A\|$.

ii. Condition 1' holds for this system.

Proof of Proposition 154. By using Lemma 139, we have $|\log(\mathrm{tr}[e^{-\beta_r H}])| \leq O(n)$. Hence to

[5]Recall $\partial\bar{A}$ is the boundary of $\bar{A}$ and is inside A. For a (κ, R)-local H, $\partial A = \{v \in \Lambda\backslash A : \exists v' \in A,\ \mathrm{dist}(v-v') \leq R\}$.

show (5.86), it is sufficient to show that

$$\left|\log\left(\frac{\mathrm{tr}[e^{-\beta H}]}{\mathrm{tr}[e^{-\beta_r H}]}\right)\right| \leq cn. \tag{5.89}$$

The difference between the numerator and denominator of (5.89) is the addition of the complex perturbations $\beta_i H = \sum_{k=1}^{m} \beta_i H_k$ to the exponent of the numerator. Instead of adding these terms all together, we can add local terms $\beta_i H_k$ step by step. We do this by setting up a telescoping series of products such that in each fraction, a new term $\beta_i H_k$ is added. We have

$$\frac{\mathrm{tr}[e^{-\beta H}]}{\mathrm{tr}[e^{-\beta_r H}]} = \frac{\mathrm{tr}[e^{-\beta_r H - i\beta_i \sum_{k=1}^{m} H_k}]}{\mathrm{tr}[e^{-\beta_r H - i\beta_i \sum_{k=1}^{m-1} H_k}]} \frac{\mathrm{tr}[e^{-\beta_r H - i\beta_i \sum_{k=1}^{m-1} H_k}]}{\mathrm{tr}[e^{-\beta_r H - i\beta_i \sum_{k=1}^{m-2} H_k}]} \cdots \frac{\mathrm{tr}[e^{-\beta_r H - i\beta_i H_1}]}{\mathrm{tr}[e^{-\beta_r H}]}. \tag{5.90}$$

Hence,

$$\left|\log\left(\frac{\mathrm{tr}[e^{-\beta H}]}{\mathrm{tr}[e^{-\beta_r H}]}\right)\right| \leq \sum_{j=0}^{m-1} \left|\log\left(\frac{\mathrm{tr}[e^{-\beta_r H - i\beta_i \sum_{k=0}^{j+1} H_k}]}{\mathrm{tr}[e^{-\beta_r H - i\beta_i \sum_{k=0}^{j} H_k}]}\right)\right|, \tag{5.91}$$

in which we set $H_0 = 0$. Since for interactions considered in this paper $m = O(n)$, we can derive the bound in (5.99) by showing for any j,

$$\left|\log\left(\frac{\mathrm{tr}[e^{-\beta_r H - i\beta_i \sum_{k=0}^{j+1} H_k}]}{\mathrm{tr}[e^{-\beta_r H - i\beta_i \sum_{k=0}^{j} H_k}]}\right)\right| \leq O(1). \tag{5.92}$$

To do so, we define $\gamma_j(t)$ for $t \in [0,1]$ to be

$$\gamma_j(t) = \log\left(\mathrm{tr}[e^{-\beta_r H - i\beta_i \sum_{k=0}^{j} H_k - i\beta_i t H_{j+1})}]\right). \tag{5.93}$$

Then, the left hand side of (5.92) can be written as

$$\begin{aligned}
\left|\log\left(\frac{\mathrm{tr}[e^{-\beta_r H - i\beta_i \sum_{k=0}^{j+1} H_k}]}{\mathrm{tr}[e^{-\beta_r H - i\beta_i \sum_{k=0}^{j} H_k}]}\right)\right| &= |\gamma_j(1) - \gamma_j(0)| \\
&\leq \max_{t\in[0,1]} \left|\frac{d\gamma_j(t)}{dt}\right| \\
&= |\beta_i| \max_{t\in[0,1]} \left|\frac{\mathrm{tr}[H_{j+1} e^{-it\beta_i H_{j+1}} e^{-\beta_r H - i\beta_i \sum_{k=0}^{j} H_k}]}{\mathrm{tr}[e^{-it\beta_i H_{j+1}} e^{-\beta_r H - i\beta_i \sum_{k=0}^{j} H_k}]}\right|.
\end{aligned} \tag{5.94}$$

For a region $A \subset \Lambda$, let H_A and H' be parts of the Hamiltonian acting on A and $\bar{A} \cup \partial\bar{A}$, respectively. One can see that for any choice of j and t, finding an upper bound like the one in (5.94) is equivalent to bounding a local expectation term like

$$\mathrm{tr}\left[H_A e^{-(\beta_r + it\beta_i)H_A} \frac{e^{-\beta H'}}{Z_\beta(H)}\right] = \mathrm{tr}[H_A \rho_\beta(H)] \tag{5.95}$$

for some suitable choice of A. We also assume, without loss of generality, that *all* local terms in H'

are complex perturbed. Using the Hölder inequality, we get

$$\left|\mathrm{tr}\left[H_A e^{-(\beta_r+it\beta_i)H_A}\frac{e^{-\beta H'}}{Z_\beta(H)}\right]\right| = \left|\mathrm{tr}_A\left[H_A e^{-(\beta_r+it\beta_i)H_A}\frac{\mathrm{tr}_{\bar{A}}[e^{-\beta H'}]}{Z_\beta(H)}\right]\right| \leq \|H_A\| e^{|\beta|\|H_A\|} d^{|\partial\bar{A}|}\left\|\frac{\mathrm{tr}_{\bar{A}}[e^{-\beta H'}]}{Z_\beta(H)}\right\|. \tag{5.96}$$

Since $|A| = O(1)$, we only need to upper bound the largest singular value of $\mathrm{tr}_{\bar{A}}[e^{-\beta H'}]/Z_\beta(H)$, whose support is only on $\partial\bar{A}$, by a constant. Let $|u\rangle$ and $|v\rangle$ be the left and right singular vectors associated with the largest singular value. We claim that there exists a rank 1 projector P supported on $\partial\bar{A}$ such that

$$\left\|\frac{\mathrm{tr}_{\bar{A}}[e^{-\beta H'}]}{Z_\beta(H)}\right\| = \mathrm{tr}_{\partial\bar{A}}\left[\frac{\mathrm{tr}_{\bar{A}}[e^{-\beta H'}]}{Z_\beta(H)}|u\rangle\langle v|\right] \leq (2+\sqrt{2})\left|\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[\frac{e^{-\beta H'}}{Z_\beta(H)}P\right]\right|. \tag{5.97}$$

This can be derived by noting that $|u\rangle\langle v|$ can be decomposed as sum of rank 1 projectors as follows

$$|u\rangle\langle v| = -\frac{1+i}{2}\left(|u\rangle\langle u| + |v\rangle\langle v|\right) + i|w^-\rangle\langle w^-| + |w^+\rangle\langle w^+|, \tag{5.98}$$

where $|w^+\rangle = \frac{1}{\sqrt{2}}(|u\rangle + |v\rangle)$ and $|w^-\rangle = \frac{1}{\sqrt{2}}(|u\rangle + i|v\rangle)$.

Finally, using the premise of this proposition given in (5.88), we get both (154) and Condition 1', which concludes the proof. □

5.6.2 Step 2: The complex site removal bound from the small relative phase condition

Proposition 155. *Consider the same setup as that of Proposition 154. Let P and Q be projectors acting on $\partial\bar{A}$. Let $\theta(\delta)$ be a complex function depending on H, P, and Q, but constant in n such that for any positive constant c, $c|\theta(\delta)| \geq \delta$ for sufficiently small δ. We can, for instance, assume $|\theta(\delta)| = \sqrt{\delta}$. Suppose when $|\beta_i| \leq \delta$ for some sufficiently small δ, we have*

$$\frac{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[e^{-i\beta_i H'}\rho_{\beta_r}(H'|P)]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[e^{-i\beta_i H'}\rho_{\beta_r}(H'|Q)]} = 1 + |\partial\bar{A}|\theta(\delta). \tag{5.99}$$

Then, the complex site removal bound (5.88) given in Proposition 154 holds, i.e.

$$\left|\log\left(\frac{\mathrm{tr}_{\bar{A}\cup A}\left[e^{-\beta H}\right]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-\beta H'}P\right]}\right)\right| \leq c. \tag{5.100}$$

Before getting to the proof of this proposition, we first state and prove a relevant lemma.

Lemma 156. *Consider the same definitions as in Proposition 155. The ratio of the* unperturbed

partition functions (with real β) with different boundary conditions can be bounded as

$$\left|\log\left(\frac{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-\beta_r H'}Q\right]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-\beta_r H'}P\right]}\right)\right| \leq c' \tag{5.101}$$

for some constant c' depending on $|\partial\bar{A}|$.

Proof of Lemma 156. Let $H_{\bar{A}}$ be terms in H' that are acting solely on $\bar{A}$. That is, the support of $H_{\bar{A}}$ does not overlap that of P and Q. This means for instance, $\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-\beta_r H_{\bar{A}}}Q\right] = \mathrm{tr}_{\bar{A}}\left[e^{-\beta_r H_{\bar{A}}}\right]\mathrm{tr}[Q]$. We have

$$\frac{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-\beta_r H'}Q\right]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-\beta_r H'}P\right]} = \frac{\mathrm{tr}_{\partial\bar{A}}[Q]}{\mathrm{tr}_{\partial\bar{A}}[P]}\frac{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-\beta_r H'}Q\right]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-\beta_r H_{\bar{A}}}Q\right]}\frac{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-\beta_r H_{\bar{A}}}P\right]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-\beta_r H'}P\right]}. \tag{5.102}$$

The first ration on the RHS of (5.102) can be bounded as $|\log(\mathrm{tr}[Q]/\mathrm{tr}[P])| \leq O(|\partial\bar{A}|)$. For the remaining two ratios we have

$$\begin{aligned}
\left|\log\left(\frac{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-\beta_r H'}P\right]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-\beta_r H_{\bar{A}}}P\right]}\right)\right| &= \left|\int_0^1 dt\frac{d}{dt}\log\left(\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-\beta_r(H_{\bar{A}}+t(H'-H_{\bar{A}}))}P\right]\right)\right| \\
&\leq \max_{t\in[0,1]}\left|\frac{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[-\beta_r(H'-H_{\bar{A}})e^{-\beta_r(H_{\bar{A}}+t(H'-H_{\bar{A}}))}P\right]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-\beta_r(H_{\bar{A}}+t(H'-H_{\bar{A}}))}P\right]}\right| \\
&\leq \max_{t\in[0,1]}\left|\frac{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[-\beta_r(H'-H_{\bar{A}})e^{-\frac{\beta_r}{2}(H_{\bar{A}}+t(H'-H_{\bar{A}}))}Pe^{-\frac{\beta_r}{2}(H_{\bar{A}}+t(H'-H_{\bar{A}}))}\right]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-\frac{\beta_r}{2}(H_{\bar{A}}+t(H'-H_{\bar{A}}))}Pe^{-\frac{\beta_r}{2}(H_{\bar{A}}+t(H'-H_{\bar{A}}))}\right]}\right| \\
&\leq \left\|\beta_r(H'-H_{\bar{A}})\right\| \leq \beta_r O(|\partial\bar{A}|)h
\end{aligned} \tag{5.103}$$

Similarly, we can exchange the role of P and Q to get a similar bound for Q. All these bounds together imply the bound (5.101). □

Proof of Proposition 155. We show how assuming equation (5.99), we can derive a *lower* and an *upper* bound for

$$\left|\frac{\mathrm{tr}_{\bar{A}\cup A}\left[e^{-\beta H}\right]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-\beta H'}P\right]}\right|. \tag{5.104}$$

We decompose the expression (5.104) into two parts denoted by L_1 and L_2 as follows

$$\begin{aligned}
\frac{\mathrm{tr}_{\bar{A}\cup A}\left[e^{-\beta H}\right]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-\beta H'}P\right]} &= \mathrm{tr}_A\left[e^{-i\beta_i H_A}e^{-\beta_r H_A}\frac{\mathrm{tr}_{\bar{A}}\left[e^{-\beta H'}\right]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-\beta H'}P\right]}\right] \\
&= L_1 + L_2,
\end{aligned} \tag{5.105}$$

where

$$L_1 = \mathrm{tr}_A \left[e^{-\beta_r H_A} \frac{\mathrm{tr}_{\bar{A}} \left[e^{-\beta H'} \right]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}} \left[e^{-\beta H'} P \right]} \right] \tag{5.106}$$

$$L_2 = \mathrm{tr}_A \left[(e^{-i\beta_i H_A} - \mathbb{1}) e^{-\beta_r H_A} \frac{\mathrm{tr}_{\bar{A}} \left[e^{-\beta H'} \right]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}} \left[e^{-\beta H'} P \right]} \right]. \tag{5.107}$$

All the complex perturbations acting on A are moved to the second part L_2 which is analyzed later and shown to have only a small contribution.

Let $\{|\psi_k\rangle\}$ be the set of eigenstates of the operator H_A that span the Hilbert space of A. The term L_1 can be written as

$$\begin{aligned} L_1 &= \sum_k \langle\psi_k| e^{-\beta_r H_A} |\psi_k\rangle \left[\frac{\mathrm{tr}_{\bar{A}\cup A} \left[e^{-\beta H'} |\psi_k\rangle\langle\psi_k| \right]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}} \left[e^{-\beta H'} P \right]} \right] \\ &= \sum_k e_k \left[\frac{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}} \left[e^{-\beta H'} \, \mathrm{tr}_{A\setminus\partial\bar{A}} |\psi_k\rangle\langle\psi_k| \right]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}} \left[e^{-\beta H'} P \right]} \right] \\ &= \sum_{j,k} e_k r_{j,k} \left[\frac{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}} \left[e^{-\beta H'} Q_{j,k} \right]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}} \left[e^{-\beta H'} P \right]} \right], \end{aligned} \tag{5.108}$$

where the first line follows from $\{|\psi_k\rangle\}$ spanning the Hilbert space of A. In the second line, we denoted $\langle\psi_k| e^{-\beta_r H_A} |\psi_k\rangle$ by positive coefficients e_k. In the last line, we used the fact that $\mathrm{tr}_{A\setminus\partial\bar{A}} |\psi_k\rangle\langle\psi_k|$ is a density operator on $\partial\bar{A}$ and can be decomposed into a convex combination of projectors $Q_{j,k}$ supported on $\partial\bar{A}$ with positive coefficients $r_{j,k}$. In other words,

$$\mathrm{tr}_{A\setminus\partial\bar{A}} |\psi_k\rangle\langle\psi_k| = \sum_j r_{j,k} Q_{j,k}. \tag{5.109}$$

From the assumption of the theorem given in (5.99) we get

$$\left[\frac{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}} \left[e^{-\beta H'} Q_{j,k} \right]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}} \left[e^{-\beta H'} P \right]} \right] = \alpha_{j,k} \left(1 + |\partial\bar{A}| \theta_{j,k}(\delta) \right), \tag{5.110}$$

where

$$\alpha_{j,k} = \frac{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}} \left[e^{-\beta_r H'} Q_{j,k} \right]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}} \left[e^{-\beta_r H'} P \right]} \tag{5.111}$$

is the ratio of the *real* partition functions, and according to Lemma 156,

$$|\log \alpha_{j,k}| \leq O(|\partial\bar{A}|). \tag{5.112}$$

Hence, we get the following expression for L_1:

$$L_1 = \sum_{j,k} \alpha_{j,k} r_{j,k} e_k \left(1 + |\partial\bar{A}|\theta_{j,k}(\delta)\right). \tag{5.113}$$

This allows us to find a lower bound on this term. Since all coefficients $\alpha_{j,k}$, $r_{j,k}$, and e_k are positive constants, Eq. (5.113) is sum of complex numbers with various magnitudes that have small complex phases at most proportional to $|\partial\bar{A}|\theta_{j,k}(\delta)$. The absolute value of the sum of these complex numbers is at least the sum of their real parts. In particular, since A is a region of *constant* size, by choosing a sufficiently small δ such that $\delta|\partial\bar{A}| \ll 1$, we can ensure that the real parts are all positive and add up to some non-zero value. More precisely,

$$|L_1| \geq \left(\sum_{j,k} \alpha_{j,k} r_{j,k} e_k\right) \cos\left(c''|\partial\bar{A}|\theta(\delta)\right) \geq \Omega(1) \quad \text{for } \delta \ll |\partial\bar{A}|. \tag{5.114}$$

We can also get an *upper* bound on $|L_1|$ using the expression (5.113). We have

$$|L_1| \leq \left(1 + |\partial\bar{A}|\right) \left(\sum_{j,k} \alpha_{j,k} r_{j,k} e_k\right) \leq O(1) \tag{5.115}$$

Now, we look at the second term L_2. Similar to the previous bound, we can find a projector Q and a constant c' such that

$$\begin{aligned} |L_2| = \left|\mathrm{tr}_A \left[(e^{-i\beta_i H_A} - \mathbb{1})e^{-\beta_r H_A} \frac{\mathrm{tr}_{\bar{A}}\left[e^{-\beta H'}\right]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-\beta H'}P\right]}\right]\right| &\leq \|e^{-i\beta_i H_A} - \mathbb{1}\|\|e^{-\beta_r H_A}\| \frac{\left\|\mathrm{tr}_{\bar{A}}\left[e^{-\beta H'}\right]\right\|_1}{\left|\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-\beta H'}P\right]\right|} \\ &\leq c'\delta\|H_A\| d^{|\partial\bar{A}|} e^{|\beta|\|H_A\|}. \end{aligned} \tag{5.116}$$

We used a bound similar to (5.110) to get to the last line.

All bounds (5.114), (5.115), and (5.116) depend on $|A|$ which is a constant. Also, as δ is made smaller, (5.116) becomes negligible compared to (5.114) or (5.115). Hence, if δ is chosen to be sufficiently small yet still a constant, we get the desired bounds:

$$O(1) \geq |L_1| + |L_2| \geq \left|\frac{\mathrm{tr}_{\bar{A}}\left[e^{-\beta H'}\right]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-\beta H'}P\right]}\right| \geq |L_1| - |L_2| \geq \Omega(1). \tag{5.117}$$

□

5.6.3 Step 3: The small relative phase condition from Condition 2

Proposition 157. *Let $H = \sum_{i=1}^m H_i$ be a geometrically-local Hamiltonian of a* classical *spin system. Suppose the correlations in this system decay exponentially as in Condition 2. Then, the bound given in (5.99) holds for this system.*

Proof of Proposition 157. The proof is by induction. The lattice Λ is already divided into regions A and $\bar{A}$ according to Propositions 154 and 155. We further split the region $\bar{A} \cup \partial\bar{A}$ into

a constant region B and its complement $\bar{B}$. For reasons that will become clear shortly, it suffices to fix an arbitrary site x on $\partial\bar{A}$ and choose region B such that $\mathrm{dist}(\partial\bar{B}, x) \gg \xi$, where ξ is the correlation length in Condition 2. We assume inductively that (5.99) holds for $\bar{B}$. Then, using the decay of correlations, we show that even after adding the contribution of region B, Equation (5.99) still holds for the region $\bar{A} \cup \partial\bar{A} = B \cup \bar{B}$.

Since we are considering classical systems, the projectors P and Q set the value of the boundary spins, each of which attains d distinct states, to some fixed values denoted by strings s_p and s_q, where $s_{p \text{ or } q} \in [d]^{|\partial\bar{A}|}$. Hence, $P = |s_p\rangle\langle s_p|$ and $Q = |s_q\rangle\langle s_q|$. Assume s_p and s_q differ on t sites. Consider a series of strings $s_1, \ldots, s_t$ such that $s_1 = s_p$, $s_t = s_q$, and s_i and s_{i+1} differ only on one site. Denote the corresponding projectors by $P_1, P_2, \ldots, P_t$. We can set up a telescoping product for (5.99) as follows:

$$\begin{aligned}
&\frac{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[e^{-i\beta_i H'}\rho_{\beta_r}(H'|P)]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[e^{-i\beta_i H'}\rho_{\beta_r}(H'|Q)]} \\
&= \frac{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[e^{-i\beta_i H'}\rho_{\beta_r}(H'|P_1)]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[e^{-i\beta_i H'}\rho_{\beta_r}(H'|P_2)]} \frac{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[e^{-i\beta_i H'}\rho_{\beta_r}(H'|P_2)]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[e^{-i\beta_i H'}\rho_{\beta_r}(H'|P_3)]} \cdots \frac{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[e^{-i\beta_i H'}\rho_{\beta_r}(H'|P_{t-1})]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[e^{-i\beta_i H'}\rho_{\beta_r}(H'|P_t)]}.
\end{aligned} \tag{5.118}$$

One can see that to get the desired bound in (5.99), it is enough to show the following bound on these ratios:

$$\frac{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[e^{-i\beta_i H'}\rho_{\beta_r}(H'|P_i)]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[e^{-i\beta_i H'}\rho_{\beta_r}(H'|P_{i+1})]} = 1 + \theta(\delta) \tag{5.119}$$

for $\theta(\delta)$ satisfying the conditions given in Proposition 154. This is why we define region B around a *single* site on $\partial\bar{A}$.

To simplify the notation, we keep using P, Q instead of P_i, P_{i+1} for the rest of the proof bearing in mind that they differ on one site. In order to show (5.119), we change the left-hand side to a slightly different expression that makes it easier to see the connection to the decay of correlations. We have

$$\frac{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[e^{-i\beta_i H'}\rho_{\beta_r}(H'|P)]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[e^{-i\beta_i H'}\rho_{\beta_r}(H'|Q)]} = 1 + \frac{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-i\beta_i H'}\left(\rho_{\beta_r}(H'|P) - \rho_{\beta_r}(H'|Q)\right)\right]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-i\beta_i H'}\rho_{\beta_r}(H'|Q)\right]}. \tag{5.120}$$

To derive (5.99), we can alternatively show

$$\frac{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-i\beta_i H'}\left(\rho_{\beta_r}(H'|P) - \rho_{\beta_r}(H'|Q)\right)\right]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-i\beta_i H'}\rho_{\beta_r}(H'|Q)\right]} = \theta(\delta). \tag{5.121}$$

The steps that we take to prove this equation are very similar to the ones in the proof of Proposition 154. Recall that H' consists of the terms in H that act on $\bar{A} \cup \partial\bar{A}$. Similarly, let H'' be part of H' that acts on $\bar{B} \cup \partial\bar{B}$. We also define T to be a projector (which again assigns a value from $[d]$ to the boundary spins) supported on $\partial\bar{B}$.

We divide both the numerator and the denominator of (5.121) by $\mathrm{tr}_{\bar{B}\cup\partial\bar{B}}[e^{-i\beta_i H''}\rho_{\beta_r}(H''|T)]$. This does not change the fraction but allows us to use the induction hypothesis. Similar to what we did in (5.105), we split the numerator into two parts, denoted by M_1 and M_2, such that the

complex perturbations acting on B are all moved to M_2. We get

$$\frac{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-i\beta_i H'}\left(\rho_{\beta_r}(H'|P)-\rho_{\beta_r}(H'|Q)\right)\right]}{\mathrm{tr}_{\bar{B}\cup\partial\bar{B}}\left[e^{-i\beta_i H''}\rho_{\beta_r}(H''|T)\right]} = M_1 + M_2, \tag{5.122}$$

where

$$\begin{aligned} M_1 &= \frac{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-i\beta_i H''}\left(\rho_{\beta_r}(H'|P)-\rho_{\beta_r}(H'|Q)\right)\right]}{\mathrm{tr}_{\bar{B}\cup\partial\bar{B}}\left[e^{-i\beta_i H''}\rho_{\beta_r}(H''|T)\right]} \\ M_2 &= \frac{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-i\beta_i H''}(e^{-i\beta_i(H'-H'')}-\mathbb{1})\left(\rho_{\beta_r}(H'|P)-\rho_{\beta_r}(H'|Q)\right)\right]}{\mathrm{tr}_{\bar{B}\cup\partial\bar{B}}\left[e^{-i\beta_i H''}\rho_{\beta_r}(H''|T)\right]}. \end{aligned} \tag{5.123}$$

Now we use the crucial fact that $\rho_{\beta_r}(H'|P \text{ or } Q)$ is a classical probability distribution that has the *Markov property*. In other words,

Lemma 158. *For any* diagonal *operator O supported on $\bar{B}\cup\partial\bar{B}$, we have*

$$\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[O\ \rho_{\beta_r}(H'|P)] = \sum_{s\in[d]^{|\partial\bar{B}|}} \mathrm{tr}_{\bar{B}\cup\partial\bar{B}}\left[O\ \rho_{\beta_r}(H''|P_s)\right]\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[P_s\ \rho_{\beta_r}(H'|P)], \tag{5.124}$$

where s denotes the state of the spins on $\partial\bar{B}$ and P_s is the corresponding projector.

This equality follows from the law of total probability. The term $\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[P_s\ \rho_{\beta_r}(H'|P)]$ is the probability of the boundary spins being in state s while $\mathrm{tr}_{\bar{B}\cup\partial\bar{B}}\left[O\ \rho_{\beta_r}(H''|P_s)\right]$ is the expectation value of O *conditioned* on the state of the boundary spins. The fact that we only need to condition on the value of the *boundary* spins follows from the Markov property of the Gibbs distribution. We postpone a more detailed proof of Eq. (5.124) until after the end of this proof.

Using (5.124), the term M_1 can be written as

$$\begin{aligned} M_1 &= \sum_{s\in[d]^{|\partial\bar{B}|}}\left(\frac{\mathrm{tr}_{\bar{B}\cup\partial\bar{B}}\left[e^{-i\beta_i H''}\ \rho_{\beta_r}(H''|P_s)\right]}{\mathrm{tr}_{\bar{B}\cup\partial\bar{B}}\left[e^{-i\beta_i H''}\rho_{\beta_r}(H''|T)\right]}-1\right)\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[P_s\ \rho_{\beta_r}(H'|P)] \\ &\quad - \sum_{s\in[d]^{|\partial\bar{B}|}}\left(\frac{\mathrm{tr}_{\bar{B}\cup\partial\bar{B}}\left[e^{-i\beta_i H''}\ \rho_{\beta_r}(H''|P_s)\right]}{\mathrm{tr}_{\bar{B}\cup\partial\bar{B}}\left[e^{-i\beta_i H''}\rho_{\beta_r}(H''|T)\right]}-1\right)\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[P_s\ \rho_{\beta_r}(H'|Q)] \\ &= \sum_{s\in[d]^{|\partial\bar{B}|}}\left(\frac{\mathrm{tr}_{\bar{B}\cup\partial\bar{B}}\left[e^{-i\beta_i H''}\ \rho_{\beta_r}(H''|P_s)\right]}{\mathrm{tr}_{\bar{B}\cup\partial\bar{B}}\left[e^{-i\beta_i H''}\rho_{\beta_r}(H''|T)\right]}-1\right)\left(\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[P_s\ \rho_{\beta_r}(H'|P)]-\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[P_s\ \rho_{\beta_r}(H'|Q)]\right). \end{aligned} \tag{5.125}$$

For later convenience, we added and subtracted 1 in the first equality. We can now use the induction hypothesis *and* the exponential decay of correlations property to bound M_1. From the induction

assumption (5.99) we see that for sufficiently small δ

$$\frac{\mathrm{tr}_{\bar{B}\cup\partial\bar{B}}\left[e^{-i\beta_i H''}\ \rho_{\beta_r}(H''|P_s)\right]}{\mathrm{tr}_{\bar{B}\cup\partial\bar{B}}\left[e^{-i\beta_i H''}\rho_{\beta_r}(H''|T)\right]} = 1 + |\partial\bar{B}|\theta(\delta). \tag{5.126}$$

Moreover, we let $x \in B$ be the site on which P and Q differ. Then, the assumption of the exponential decay of correlations (5.52) implies

$$\left|\mathrm{tr}_{\Lambda\setminus\partial\bar{B}}[\rho_{\beta_r}(H'|P)] - \mathrm{tr}_{\Lambda\setminus\partial\bar{B}}[\rho_{\beta_r}(H'|Q)]\right| \leq ce^{-\mathrm{dist}(x,\partial\bar{B})/\xi}. \tag{5.127}$$

Overall, (5.126) and (5.127) show that $|M_1|$ can be bounded as follows:

$$|M_1| \leq c|\theta(\delta)||\partial\bar{B}|e^{-\mathrm{dist}(x,\partial\bar{B})/\xi}. \tag{5.128}$$

Similarly, one can show that

$$M_2 = \sum_{s\in[d]^{|B|}} \left(\frac{\mathrm{tr}_{\bar{B}\cup\partial\bar{B}}\left[e^{-i\beta_i H''}\ \rho_{\beta_r}(H''|P_s)\right]}{\mathrm{tr}_{\bar{B}\cup\partial\bar{B}}\left[e^{-i\beta_i H''}\rho_{\beta_r}(H''|T)\right]}\right) \mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[(e^{-i\beta_i(H'-H'')} - \mathbb{1})P_s\ (\rho_{\beta_r}(H'|P) - \rho_{\beta_r}(H'|Q))], \tag{5.129}$$

which again by using (5.99) and $\sum_{s\in[d]^{|B|}} \mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[P_s\ \rho_{\beta_r}(H'|P)] = 1$ can be bounded as

$$|M_2| \leq c''\delta\|H_B\|(1 + |\partial\bar{B}||\theta(\delta)|). \tag{5.130}$$

We next analyze the denominator of (5.121) that similar to the numerator is first divided by $\mathrm{tr}_{\bar{B}\cup\partial\bar{B}}\left[e^{-i\beta_i H''}\rho_{\beta_r}(H''|T)\right]$. We can follow similar arguments to Section 5.6.2 to show that for sufficiently small δ, we can lower bound this term by a constant:

$$\frac{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-i\beta_i H'}\rho_{\beta_r}(H'|Q)\right]}{\mathrm{tr}_{\bar{B}\cup\partial\bar{B}}\left[e^{-i\beta_i H''}\rho_{\beta_r}(H''|T)\right]} \geq \Omega(1). \tag{5.131}$$

Finally, we can insert these bounds in (5.121) to get the following upper bound for some constants c_1 and c_2:

$$\left|\frac{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-i\beta_i H'}\ (\rho_{\beta_r}(H'|P) - \rho_{\beta_r}(H'|Q))\right]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-i\beta_i H'}\rho_{\beta_r}(H'|Q)\right]}\right| \leq c_1|\theta(\delta)||\partial\bar{B}|e^{-\mathrm{dist}(x,\partial\bar{B})/\xi} + c_2\delta\|H_B\|(1 + |\partial\bar{B}||\theta(\delta)|). \tag{5.132}$$

Since $\theta(\delta)$ can be chosen as $\sqrt{\delta}$ for instance, for a fixed $\mathrm{dist}(x,\partial\bar{B})$, we can always choose δ small

enough such that

$$\frac{c_2\delta\|H_B\|(1+|\partial\bar{B}||\theta(\delta)|)}{c_1|\theta(\delta)||\partial\bar{B}|e^{-\mathrm{dist}(x,\partial\bar{B})/\xi}} \leq c_3 \tag{5.133}$$

for some constant $c_3 \leq 1$. We can also choose $\mathrm{dist}(x, \partial\bar{B})$ sufficiently large enough so that

$$c_1|\theta(\delta)||\partial\bar{B}|e^{-\mathrm{dist}(x,\partial\bar{B})/\xi} \leq |\theta(\delta)|. \tag{5.134}$$

Without the term $e^{-\mathrm{dist}(x,\partial\bar{B})/\xi}$ that originates from the decay of correlations property, we could not ensure that the bound $|\theta(\delta)|$ is recovered after the induction step. □

Here, we prove Lemma 158 that we mentioned during the proof of Proposition 157. We restate the lemma for convenience.

Restatement of Lemma 158. *Consider the same setup as in Proposition 157 in which we restrict ourselves to classical Hamiltonians. For any diagonal operator O supported on $\bar{B} \cup \partial\bar{B}$, we have*

$$\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[O\ \rho_{\beta_r}(H'|P)] = \sum_{s\in[d]^{|\partial\bar{B}|}} \mathrm{tr}_{\bar{B}\cup\partial\bar{B}}\left[O\ \rho_{\beta_r}(H''|P_s)\right] \mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[P_s\ \rho_{\beta_r}(H'|P)], \tag{5.135}$$

where s denotes the state of the spins on $\partial\bar{B}$ and P_s is the corresponding projector.

Proof of Lemma 158. We have

$$\begin{aligned}
\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[O\ \rho_{\beta_r}(H'|P)\right] &= \mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[O_{\bar{B}\cup\partial\bar{B}}\ \frac{e^{-\beta_r H'}P}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-\beta_r H'}P\right]}\right] \\
&= \sum_{s\in[d]^{|\partial\bar{B}|}} \mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[O_{\bar{B}\cup\partial\bar{B}}\ \frac{e^{-\beta_r H''}P_s}{\mathrm{tr}_{\bar{B}\cup\partial\bar{B}}[e^{-\beta H''}P_s]} \frac{P_s e^{-\beta(H'-H'')}P\ \mathrm{tr}_{\bar{B}\cup\partial\bar{B}}[e^{-\beta H''}P_s]}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[e^{-\beta_r H'}P\right]}\right] \\
&= \sum_{s\in[d]^{|\partial\bar{B}|}} \mathrm{tr}_{\bar{B}\cup\partial\bar{B}}\left[O_{\bar{B}\cup\partial\bar{B}}\ \frac{e^{-\beta_r H''}P_s}{\mathrm{tr}_{\bar{B}\cup\partial\bar{B}}[e^{-\beta H''}P_s]}\right] \mathrm{tr}_{\bar{A}\cup\partial\bar{A}}\left[P_s\ \frac{e^{-\beta H'}P}{\mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[e^{-\beta H'}P]}\right] \\
&= \sum_{s\in[d]^{|\partial\bar{B}|}} \mathrm{tr}_{\bar{B}\cup\partial\bar{B}}\left[O_{\bar{B}\cup\partial\bar{B}}\ \rho_{\beta_r}(H''|P_s)\right] \mathrm{tr}_{\bar{A}\cup\partial\bar{A}}[P_s\ \rho_{\beta_r}(H'|P)],
\end{aligned} \tag{5.136}$$

□

Remark 159. *A first step in generalizing the proof of Theorem 153 to the quantum case would be to consider commuting Hamiltonians. While some parts of the proof already apply to these systems, the one in Proposition 157 does not immediately go through. One issue is that the decomposition (5.118) does not have a quantum counterpart. In particular, when comparing the effect of two entangled boundary projectors, we cannot write a telescoping product that reduces this to comparing local projectors. Perhaps by using the commutativity of the terms in the Hamiltonian, we can find a structure in the projectors that allows us to overcome this problem. We leave this for future work.*

5.7 Extrapolating from high external fields and Lee-Yang zeros

In this section, we study spin systems whose interactions are described by two- or one-body terms. For qubits, such systems are generally described by Hamiltonians of the form

$$H(\mu) = -\sum_{\substack{(i,j)\in E \\ a,b\in\{x,y,z\}}} J_{ij}^{ab}\sigma_a \otimes \sigma_b - \sum_{i\in V}(h_i^x X_i + h_i^y Y_i + \mu h_i^z Z_i), \tag{5.137}$$

where $J_{ij}^{ab}, h_i^a, \mu \in \mathbb{R}$ and $\sigma_a \in \{X, Y, Z, \mathbb{1}\}$ are Pauli matrices. The interaction graph, as usual, is denoted by $G = (V, E)$ with $|V| = n$ and $|E| = m$. Physically, the two-body interactions J_{ij}^{ab} are due to the *coupling* between the spins of the particles on adjacent sites, whereas the one-body terms h_i^a characterize the interaction of spins with some *external magnetic field*.

Remark 160. *For later convenience, we introduce an extra factor μ before the Z_i terms in (5.137). One can think of μ as the maximum strength of the external field in the z-direction. As explained below, this parameter plays the same role as β in the extrapolation algorithm of Section 5.3.*

In Section 5.3, we developed approximation algorithms for the partition function of a quantum many-body systems by extrapolating from high to low temperatures. In this section, we again use the idea of extrapolation, but this time our parameter of interest is μ, the magnitude of the one-body terms in the z-direction. The physical motivation for this approach is that when the system is subject to a large enough external field in a specific direction (the z-direction in our case), all the spins align themselves in that direction, and estimating the properties of the system becomes trivial. However, as we move to smaller fields, the other interaction terms between the particles gain significance, making the problem non-trivial.

In order to apply the extrapolation algorithm in Section 5.3, we need to know the locus of the complex zeros of the partition function as a function of the external field μ. As mentioned in Section 5.1, these are called Lee-Yang zeros. We can exactly determine the locus of these zeros when the Hamiltonian (5.137) describes a *ferromagnetic* system, i.e. when the neighboring spins tend to align along the same direction. This is a result of Suzuki and Fisher [SF71]. There, by generalizing the result of Lee and Yang [LY52], they show that all the complex zeros lie on the imaginary axis in the μ-plane. Theorem 163 covers this result.

The key step is to map the quantum system to a classical spin system using the *quantum-to-classical mapping* (see for example [SF71, Bra15]). Then, by the result of Theorem 133, instead of studying the zeros of the quantum system, we can focus on the zeros of a classical system.

The classical spin system that we obtain involves 1-, 2-, and 4-body terms in its Hamiltonian. We represent the terms in the Hamiltonian with functions $V_{1,i}$, $V_{2,i}$, and $V_{4,i,j}$ that assign possibly complex numbers to their input spins. The indices of these functions refer to the number of particles that they act on and the coefficients of the original quantum Hamiltonian that they depend on.

Proposition 161 (Quantum-to-classical mapping, cf. [SF71])**.** *Consider a 2-local Hamiltonian H as in Eq. (5.137). Let $z_i = e^{\beta\mu h_i^z/\eta}$ and $\varepsilon = \beta/\eta$. This Hamiltonian can be mapped to a 4-local classical spin model involving $n' = n\eta$ spins $s \in \{\pm 1\}$ with the interactions of the form $V_{1,i} : \{\pm 1\} \to \mathbb{C}$,*

$V_{2,i} : \{\pm 1\}^2 \to \mathbb{C}$, *and* $V_{4,i,j} : \{\pm 1\}^4 \to \mathbb{C}$ *such that* $\exp(V_{1,i}(s_a)) = z_i^{s_a}$ *and*

$$\sum_{s_a, s_b \in \{\pm 1\}} \exp\left(V_{2,i}(s_a, s_b)\right) |s_a\rangle\langle s_b| = \begin{pmatrix} 1 & \varepsilon(h_i^x + i h_i^y) \\ \varepsilon(h_i^x - i h_i^y) & 1 \end{pmatrix},$$

$$\sum_{s_a, s_b, s_{a'}, s_{b'} \in \{\pm 1\}} \exp\left(V_{4,i,j}(s_a, s_b, s_{a'}, s_{b'})\right) |s_a, s_{a'}\rangle\langle s_b, s_{b'}| =$$

$$\begin{pmatrix} 1+\varepsilon J_{ij}^{zz} & \varepsilon(-iJ_{ij}^{zy} + J_{ij}^{zx}) & \varepsilon(-iJ_{ij}^{yz} + J_{ij}^{xz}) & \varepsilon(J_{ij}^{xx} - J_{ij}^{yy} - iJ_{ij}^{xy} - iJ_{ij}^{yx}) \\ \varepsilon(iJ_{ij}^{zy} + J_{ij}^{zx}) & 1-\varepsilon J_{ij}^{zz} & \varepsilon(J_{ij}^{xx} + J_{ij}^{yy} + iJ_{ij}^{xy} - iJ_{ij}^{yx}) & \varepsilon(iJ_{ij}^{yz} - J_{ij}^{xz}) \\ \varepsilon(iJ_{ij}^{yz} + J_{ij}^{xz}) & \varepsilon(J_{ij}^{xx} + J_{ij}^{yy} - iJ_{ij}^{xy} + iJ_{ij}^{yx}) & 1-\varepsilon J_{ij}^{zz} & \varepsilon(iJ_{ij}^{zy} - J_{ij}^{zx}) \\ \varepsilon(J_{ij}^{xx} - J_{ij}^{yy} + iJ_{ij}^{xy} + iJ_{ij}^{yx}) & \varepsilon(-iJ_{ij}^{yz} - J_{ij}^{xz}) & \varepsilon(-iJ_{ij}^{zy} - J_{ij}^{zx}) & 1+\varepsilon J_{ij}^{zz} \end{pmatrix}. \quad (5.138)$$

The partition function of this classical system is of the form

$$Z_{c\ell}(\mu) = \sum_{s_1, \ldots, s_\eta \in \{\pm 1\}} \exp\left(\sum_{\substack{i \in V \\ a \in E_{1,i}}} V_{1,i}(s_a) + \sum_{\substack{i \in V \\ (a,b) \in E_{2,i}}} V_{2,i}(s_a, s_b) + \sum_{\substack{(i,j) \in E \\ (a,b,a',b') \in E_{4,i,j}}} V_{4,i,j}(s_a, s_b, s_{a'}, s_{b'}) \right), \quad (5.139)$$

where $E_{1,i}$, $E_{2,i}$, *and* $E_{4,i,j}$ *are certain unordered subsets of vertices that depend on the choice of* i, j *(see the remark below), and we included the effective temperature of the classical system in the coefficients* $V_{1,i}$, $V_{2,i}$, *and* $V_{4,i,j}$. *Moreover, in the limit* $\eta \to \infty$, *the partition function of the classical model uniformly converges to that of the quantum system.*

Remark 162. *The details of the interaction (hyper)graph of the classical system in Proposition 161 is not important for our purposes. We can think of this graph as* η *copies of the original interaction graph* $G = (V, E)$ *stacked on top of each other. These copies are coupled together by the application of* $V_{1,i}, V_{2,i}$, *and* $V_{4,i,j}$. *While the interaction terms like* $V_{1,i}$ *apply to all vertices, the terms* $V_{2,i}$ *act on a vertex in one of the copies of* G *and its clones in the neighboring graphs. The set* $E_{2,i}$ *denotes the set of all such two vertices that* $V_{2,i}$ *acts on. Similarly, the set* $E_{4,i,j}$ *corresponds to all four vertices that interact through* $V_{4,i,j}$.

In Proposition 161, the dependency on μ only appears in the 1-body terms $V_{1,i}$ and parameters z_i. Also, since we do not rely on sampling algorithms, we do not restrict ourselves to *stoquastic* Hamiltonians as in [BDOT08] or [BG17], but we later put constraints on the coefficients J_{ij}^{ab} to make the Hamiltonian ferromagnetic.

5.7.1 Complex zeros of ferromagnetic systems

We now state a *generalized* Lee-Yang theorem that characterizes the locus of the complex zeros of certain classical spin systems.

Theorem 163 (Generalized Lee-Yang theorem, cf. [SF71])**.** *Consider the classical spin system described in Proposition 161 or more generally one that satisfies the following conditions:*

(i)

$$V_{2,i}(-s_a,-s_b) = V^*_{2,i}(s_a,s_b)$$
$$V_{4,i,j}(-s_a,-s_b,-s_{a'},-s_{b'}) = V^*_{4,i,j}(s_a,s_b,s_{a'},s_{b'}) \tag{5.140}$$

(ii)

$$|\exp(V_{2,i}(+1,+1))| \geq \frac{1}{4}\sum_{s_a,s_b\in\{\pm1\}} |\exp(V_{2,i}(s_i,s_j))|$$
$$|\exp(V_{4,i,j}(+1,+1,+1,+1))| \geq \frac{1}{4}\sum_{s_a,s_b,s_{a'},s_{b'}\in\{\pm1\}} |\exp(V_{4,i,j}(s_a,s_b,s_{a'},s_{b'}))|. \tag{5.141}$$

Let $Z_{c\ell}(\mu)$ be the partition function of this system as a function of μ for a fixed β. Then, the zeros of this partition function, i.e. the solutions of $Z_{c\ell}(\mu)=0$, are all on the imaginary axis in the complex μ-plane, that is, $\mathrm{Re}(\mu)=0$.

Proof of Theorem 163. Refer to [SF71] for the detailed proof of this proposition. Here we only sketch one of the main ideas in their proof.

For simplicity and in order to roughly see why conditions (i) and (ii) are sufficient for the zeros of the partition function to lie on the imaginary axis, we neglect the $V_{4,i,j}$ terms and focus on the $V_{1,i}$ and $V_{2,i}$ interactions. Recall that $z_i = e^{\beta\mu h_i^z/\eta}$ and $\exp(V_{1,i}(s_a)) = z_i^{s_a}$. Then, $Z_{ij}(z_i,z_j) = \sum_{s_a,s_b\in\{\pm1\}} z_i^{s_a} z_j^{s_b} \exp(V_{2,i}(s_i,s_j))$ is proportional to the partition function of the system when all spins except s_a and s_b are fixed to some certain values $\{s_k\}_{k\neq a,b}$.

Consider the solutions of $Z_{ij}(z_i,z_j)=0$. It is shown in [SF71] that if such a solution satisfies $|z_j|>1$ and $|z_i|>1$, then we can find another solution such that $|z_j|=1$ and $|z_i|>1$ (a similar result holds for $|z_i|,|z_j|<1$).

Here, we show that when $|z_j|=1$, we also necessarily have $|z_i|=1$. Since z_i and z_j depend on μ through $z_i = e^{\beta\mu h_i^z/\eta}$, we see that the partition function can only vanish when $\mathrm{Re}(\mu)=0$. Although we do not show it here, it turns out that this condition is actually sufficient to show that the whole partition function, without any fixed spins, also has complex zeros only on the imaginary axis.

We have

$$Z_{ij}(z_i,z_j) = \left(\sum_{s_b\in\{\pm1\}} z_j^{s_b}\exp(V_{2,i}(+1,s_b))\right) z_i + \left(\sum_{s_b\in\{\pm1\}} z_j^{s_b}\exp(V_{2,i}(-1,s_b))\right) z_i^{-1}. \tag{5.142}$$

Using the condition (i) in (ii) we see that

$$|\exp(V_{2,i}(+1,+1))| \geq |\exp(V_{2,i}(+1,-1))|. \tag{5.143}$$

If we consider $|z_j|=1$, this implies $\sum_{s_b\in\{\pm1\}} z_j^{s_b}\exp(V_{2,i}(+1,s_b)) \neq 0$. We use this in Eq. (5.142) to find the solutions of $Z_{ij}(z_i,z_j)=0$ for some $|z_j|=1$. We get

$$|z_i|^2 = \frac{|\sum_{s_b\in\{\pm1\}} z_j^{s_b}\exp(V_{2,i}(-1,s_b))|}{|\sum_{s_b\in\{\pm1\}} z_j^{s_b}\exp(V_{2,i}(+1,s_b))|}, \tag{5.144}$$

but another application of condition (i) implies $|z_i| = 1$ as desired. The rest of the proof for the whole partition function involves a recursive use of this conclusion and shows that the location of the zeros remains on the imaginary axis when different interactions are summed over in the partition function. □

Remark 164. *Instead of μ, it is common to consider the partition function as a function of e^{μ}. In this case, the complex zeros are located on the unit circle in the e^{μ}-plane. Hence, the Lee-Yang theorem is often called the* circle theorem.

The connection between Theorem 163 and quantum ferromagnetic systems is established through the following theorem.

Theorem 165 (Zeros of ferromagnetic systems, cf. [SF71])**.** *Let $H(\mu)$ be a 2-local Hamiltonian as in Eq. (5.137) with $J_{ij}^{xz}, J_{ij}^{zx}, J_{ij}^{yz}, J_{ij}^{zy} = 0$ defined over an arbitrary interaction graph that is not necessarily geometrically local. Suppose $h_i^z \geq 0$, and additionally, the following constraint is satisfied by the coefficients:*

$$J_{ij}^{zz} \geq \frac{1}{2}\left[\left(J_{ij}^{xx} - J_{ij}^{yy}\right)^2 + \left(J_{ij}^{xy} + J_{ij}^{yx}\right)^2\right]^{\frac{1}{2}} + \frac{1}{2}\left[\left(J_{ij}^{xx} + J_{ij}^{yy}\right)^2 + \left(J_{ij}^{xy} - J_{ij}^{yx}\right)^2\right]^{\frac{1}{2}}. \tag{5.145}$$

Then, the partition function of this system only vanishes when $\mathrm{Re}(\mu) = 0$.

When $J_{ij}^{xy} = J_{ij}^{yx} = 0$, this condition simplifies to

$$J_{ij}^{zz} \geq |J_{ij}^{yy}|, \quad J_{ij}^{zz} \geq |J_{ij}^{xx}|. \tag{5.146}$$

This characterizes the ferromagnetic Heisenberg model given by

$$H = -\sum_{(i,j)\in E}\left(J_{ij}^{xx}X_iX_j + J_{ij}^{yy}Y_iY_j + J_{ij}^{zz}Z_iZ_j\right) - \sum_{i\in V}\left(h_i^xX_i + h_i^yY_i + \mu h_i^zZ_i\right). \tag{5.147}$$

Proof of Theorem 165. The proof follows by applying Proposition 161 to map the quantum system (5.137) to the classical system in (5.138). One can see that if the quantum system satisfies (5.145), then the corresponding classical system satisfies the conditions (5.140) and (5.141). Hence, the generalized Lee-Yang theorem in Theorem 163 shows that the zeros of the classical system are located on the imaginary axis. As the error ε in the mapping goes to zero, we get a family of classical partition functions that approach the quantum partition function. Theorem 133 implies that the complex zeros of the quantum and classical systems coincide in the limit of $\varepsilon \to 0$. Thus, the complex zeros of the quantum system are also located on the imaginary axis. □

Remark 166. *One can extend the result of Theorem 163 to include interactions between spins greater than spin 1/2. It is shown in [Suz69] that the partition function of the Heisenberg model with spin s particles can be mapped to that of a spin 1/2 Heisenberg model as in (5.147). Therefore, the Lee-Yang theorem holds for these systems too.*

5.7.2 An algorithm for the anisotropic XXZ model

In Section 5.7.1, we studied the location of the complex zeros of a 2-local Hamiltonian when the external magnetic field μ is varied. Here, we focus on a specific subclass of those Hamiltonians

for which we can find an approximation algorithm. Particularly, we consider the *anisotropic* XXZ model which has the following Hamiltonian:

$$H(\mu) = -\sum_{(i,j)\in E} \left(J_{ij}(X_iX_j + Y_iY_j) + J_{ij}^{zz} Z_iZ_j\right) - \frac{\mu}{2}\sum_{i\in V}(Z_i + \mathbb{1}). \tag{5.148}$$

Compared to the Heisenberg model, the XXZ model assigns equal coefficients to the X_iX_j and Y_iY_j terms and does not include X_i and Y_i terms. An important property of this model that we use in our algorithm is that

$$\left[H(\mu), \frac{\mu}{2}\sum_{i\in V}(Z_i + \mathbb{1})\right] = 0. \tag{5.149}$$

To see this, notice that $[X_iX_j + Y_iY_j, Z_i + Z_j] = 0$.

Let $|s_1, s_2, \ldots, s_n\rangle$ be an assignment of spins ± 1 to all the vertices. Any such vector is an eigenstate of $1/2\sum_{i=1}^{n}(Z_i + \mathbb{1})$, that is,

$$\frac{1}{2}\sum_{i=1}^{n}(Z_i + \mathbb{1})|s_1, s_2, \ldots, s_n\rangle = \frac{1}{2}(\sum_{i=1}^{n} s_i + n)|s_1, s_2, \ldots, s_n\rangle. \tag{5.150}$$

Let $\mathcal{H}_k$ denote the eigenspace of $1/2\sum_{i=1}^{n}(Z_i + \mathbb{1})$ that corresponds to the kth eigenvalue. This subspace is spanned by the binary strings of length n with Hamming weight k. We have:

$$\forall |v\rangle \in \mathcal{H}_k, \quad \frac{1}{2}\sum_{i=1}^{n}(Z_i + \mathbb{1})|v\rangle = k|v\rangle. \tag{5.151}$$

We can partition the Hilbert space of the n vertices $\mathcal{H}$ into $\mathcal{H} = \oplus_k \mathcal{H}_k$. The dimension of each of these subspaces $\dim(\mathcal{H}_k)$ is $\binom{n}{k}$. Since (5.149) holds, the partition function of this model can be written as a polynomial in $z = \exp(\beta\mu)$.

Lemma 167. *The partition function of the anisotropic* XXZ *model with $\mu < 0$ given in (5.148) can be written as*

$$Z_\beta(H(\mu)) = \sum_{k=0}^{n} q_k z^k, \tag{5.152}$$

where $z = e^{\beta\mu}$ and the coefficients q_k are defined by

$$q_k = \mathrm{tr}_{\mathcal{H}_k}[e^{\beta\sum_{(i,j)\in E}\left(J_{ij}(X_iX_j + Y_iY_j) + J_{ij}^{zz}Z_iZ_j\right)}]. \tag{5.153}$$

Proof of Lemma 167. We have

$$\begin{aligned} Z_\beta(H(\mu)) &= \mathrm{tr}_{\mathcal{H}}[e^{-\beta H(\mu)}] \\ &= \mathrm{tr}_{\mathcal{H}}[e^{\beta \sum_{(i,j)\in E}\left(J_{ij}(X_iX_j+Y_iY_j)+J_{ij}^{zz}Z_iZ_j\right)}e^{\beta\mu/2\sum_{i\in V}(Z_i+\mathbb{1})}.] \\ &= \sum_{k=0}^{n} e^{\beta\mu k}\,\mathrm{tr}_{\mathcal{H}_k}[e^{\beta \sum_{(i,j)\in E}\left(J_{ij}(X_iX_j+Y_iY_j)+J_{ij}^{zz}Z_iZ_j\right)}] \\ &= \sum_{k=0}^{n} q_k z^k. \end{aligned}$$

□

Remark 168. *In Lemma 167 and what follows we assume for simplicity that $\mu < 0$ and Taylor expand in z. If $\mu > 0$ we can pull out a factor of z^n and expand in $1/z$ instead.*

Now, we are ready to state an algorithm for this model.

Theorem 169 (Approximation algorithm for the partition function of the XXZ model)**.** *There is an algorithm that runs in $n^{O(\log(n/\varepsilon))}$ time and outputs an ε-multiplicative approximation to the partition function of the anisotropic* XXZ *model in the ferromagnetic regime, i.e. when $J_{ij}^{zz} \geq |J_{ij}|$ and μ is an arbitrary nonzero constant.*

Proof of Theorem 169. By Lemma 167, the partition function is a polynomial of degree n given in (5.152). The location of its zeros is given by Theorem 165. Hence, we can apply the truncated Taylor series of Proposition 135 to obtain an approximation algorithm for $Z_\beta(H(\mu))$.

According to Lemma 167, the partition function of this system is

$$Z_\beta(H(\mu)) = \sum_{k=0}^{n} q_k z^k.$$

The running time of the extrapolation algorithm is dominated by the calculation of the coefficients q_k of the Taylor expansion, where q_k is

$$q_k = \mathrm{tr}_{\mathcal{H}_k}[e^{\beta \sum_{(i,j)\in E}\left(J_{ij}(X_iX_j+Y_iY_j)+J_{ij}^{zz}Z_iZ_j\right)}] \tag{5.154}$$

and $\dim(\mathcal{H}_k) = \binom{n}{k}$. In general, we can decompose the Hilbert space of the system as $\mathcal{H} = \oplus_k \mathcal{H}_k$. The local Hamiltonian H is block diagonal in this basis. Since H is sum of local terms, it takes time $n^{O(k)}$ to compute the entries of H and diagonalize it in the block corresponding to the subspace $\mathcal{H}_k$. Then we can find the trace of the exponential of this block also in time $n^{O(k)}$. Since we only need $k = O(\log(n))$ in the truncated Taylor expansion, we achieve an overall running time of $n^{O(\log(n/\varepsilon))}$.

□

Even though Theorem 165 applies to a broader class of 2-local Hamiltonians such as the Heisenberg model, our method does not immediately give an algorithm for those Hamiltonians. The reason is a technical difficulty in representing the partition function of these models as a polynomial in $\exp(\beta\mu)$ (or $\exp(-\beta\mu)$) . This turns out not to be an issue for the XXZ model since the 1-body terms $\sum_i Z_i$ commute with the rest of the Hamiltonian.

One might wonder why we could not use the extrapolation algorithm directly for the classical system that we get after the mapping in Proposition 161. After all, the partition function of this system is also a polynomial of degree $\mathrm{poly}(n)$ and the location of its zeros is the same as that of the quantum system. It seems that at least naively applying this idea does not work. This is because the point that we want to extrapolate to in the classical system is μ/η instead of μ. For the error of the mapping to be $1/\mathrm{poly}(n)$, we need η to be $\mathrm{poly}(n)$. Thus, the ending point of the extrapolation is vanishingly close to the imaginary axis where the zeros are located. This makes the running time blow up and become exponential instead of quasi-polynomial.

Note that sampling algorithms like the ones used in [Bra15, BG17] do not encounter this problem. The running time of these algorithms remains efficient even if the parameters of the classical Hamiltonian scale with the number of particles n. There are unfortunately no randomized algorithms based on sampling known for the 4-local classical Hamiltonian obtained in the mapping of Proposition 161. We leave extending our result to cover all the Hamiltonians considered in Theorem 165 as a challenge for future work.

Chapter 6

Improved approximation algorithms for bounded-degree local Hamiltonians

Chapter summary: The low-temperature properties of interacting quantum systems are believed to require exponential resources to compute in the general case. Quantifying the extent to which such properties can be approximated using efficient algorithms remains a significant open challenge. In this chapter, we consider the task of approximating the ground state energy of two-local quantum Hamiltonians with bounded-degree interaction graphs. Most existing algorithms optimize the energy over the set of product states. We propose and analyze a family of shallow quantum circuits that can be used to improve the approximation ratio achieved by a given product state. The algorithm takes as input an n-qubit product state $|v\rangle$ with variance $\mathrm{Var}_v(H)$ and improves its energy by an amount proportional to $\mathrm{Var}_v(H)^2/n$. In a typical case, this results in an extensive improvement in the estimated energy. We extend our results to k-local Hamiltonians and entangled initial states. This chapter is based on:

[AGMKS21] Anurag Anshu, David Gosset, Karen J. Morenz Korol, and Mehdi Soleimanifar. Improved approximation algorithms for bounded-degree local Hamiltonians. *Phys. Rev. Lett., 127:250502, Dec 2021*

6.1 Introduction

Quantum computers are capable of efficiently computing the dynamics of quantum many-body systems [Llo96], and it is anticipated that they can be useful for scientific applications in physics, materials science and quantum chemistry. The extent of the quantum advantage for other important simulation tasks, such as computing low temperature properties of quantum systems, is still unknown. In this chapter we consider the task of approximating the ground state energy of local Hamiltonians. Here it is natural to expect some improvement over classical machines which cannot even store the state of such systems efficiently. Indeed, classical methods such as the mean-field or Hartree-Fock approximations do not capture the entanglement structure present in the true ground state.

Motivated by small quantum computers that may be available in the near future, there has been increased interest in devising algorithms that consume few quantum resources and can be implemented across a wide range of hardware platforms. In this vein, heuristic algorithms for ground state preparation have been proposed based on variationally minimizing the energy over the output

states of shallow (low-depth) quantum circuits [PMS+14, FGG14, KMT+17]. Although variational algorithms have been rigorously analyzed for specific problems and some limitations are known [MBS+18, FGG20, BKKT20, BGM21], no general treatment of their efficacy exists. Characterizing the advantage offered by shallow quantum circuits and variational quantum algorithms stands as a pressing challenge.

In this chapter, we derive rigorous bounds on the performance of shallow quantum circuits in estimating the ground state energy of local Hamiltonians. For simplicity, we state our results for a system of qubits with two-local interactions. Later in Section 6.3.5, we discuss extensions of our results to k-local Hamiltonians.

To begin, let $G = (V, E)$ be a graph, and consider a Hamiltonian

$$H = \sum_{\{i,j\} \in E} h_{ij} \tag{6.1}$$

with $n = |V|$ qubits and nearest-neighbor interactions h_{ij} that act nontrivially only on qubits $\{i, j\}$ at vertices connected by an edge. We assume without loss of generality that $\|h_{ij}\| \leq 1$. We are interested in the problem of approximating the ground energy or smallest eigenvalue $\lambda_{\min}(H)$ of the Hamiltonian. It will be convenient to instead approximate the largest eigenvalue $\lambda_{\max}(H)$; this convention matches the one used in classical optimization and is without loss of generality, since $\lambda_{\min}(H) = -\lambda_{\max}(-H)$. In the worst case, the problem of estimating the largest eigenvalue $\lambda_{\max}(H)$ of Eq. (6.1) to within an additive error scaling inverse polynomially with n is believed to be intractable for quantum or classical computers [1]. Here we consider the approximation task where the goal is to compute an estimate e for $\max(H)$ such that the approximation ratio $r \equiv e/\lambda_{max}(H)$ is as close to 1 as possible. We will also be interested in efficient quantum algorithms that prepare states $|\psi\rangle$ with good approximation ratios.

Besides describing local interactions encountered in physics, Hamiltonians of the form Eq. (6.1) can encode notable cost-functions considered in computer science and thus provide a physically motivated extension of the classical *approximation algorithm* setting [Vaz13]. For example, one may consider an Ising Hamiltonian for which $h_{ij} = (I - Z_i Z_j)/2$, where Z is the Pauli operator. This Hamiltonian is classical—that is, diagonal in the computational basis—and computing its maximum eigenvalue is equivalent to finding the Max-Cut of the graph G, a well-studied classical optimization problem. More generally, two-local quantum Hamiltonians may involve noncommuting terms such as Heisenberg interactions $h_{ij} = 1/4(I - X_i X_j - Y_i Y_j - Z_i Z_j)$ (with Pauli X, Y and Z operators); the resulting optimization problem can be viewed as a quantum analogue of Max-Cut [GP19]. Quantum approximation algorithms aim to estimate the largest eigenvalue of such Hamiltonians and have been studied in several previous works. This includes the Heisenberg interactions mentioned above [GP19, AGM20] and more general settings in which the interaction terms h_{ij} are restricted to be positive semidefinite [GK12, HLP20, PT20], or traceless [HM17, BGKT19].

Despite considerable interest, the ultimate limits of efficient algorithms for quantum approximation algorithms are poorly understood. Approximation ratios approaching 1 are only known to be achievable for certain special families of graphs, including lattices or bounded-degree planar graphs using tensor product of $O(1)$-qubit states [BBT07] or high degree graphs using tensor products of single-qubit states [BBT07, GK12, BH16]. In certain cases, one may ascertain limitations on

[1] In particular, a decision version of this problem is complete for the complexity class QMA which is a quantum generalization of NP

efficient achievable approximation ratios from the classical Probabilistically Checkable Proof (PCP) theorem [ALM+98, AS98, Din07], though stronger and more general limitations may follow from the quantum PCP conjecture if some version of it can be proven [AAV13].

A quantum approximation algorithm typically outputs an estimate of the form $\langle v|H|v\rangle$ where $|v\rangle$ is a quantum state computed by the algorithm. A central challenge is to understand the structure of quantum states $|v\rangle$ that achieve good approximation ratios in the general case. Most existing algorithms are based on tensor products of one- or few-qubit states, while Ref. [AGM20] also considers states prepared by shallow quantum circuits. In this chapter, we describe conditions under which the performance of such algorithms can be *improved*. We restrict our attention to local Hamiltonians on bounded-degree graphs and consider an improvement strategy based on shallow quantum circuits.

6.2 Summary of results

6.2.1 Improvement of product states

To this end, suppose we are given an n-qubit state $|v\rangle$ and a Hamiltonian Eq. (6.1) defined on a graph $G = (V, E)$ with maximum degree $d \geq 2$. It will be convenient to assume (without loss of generality) that G is d-regular—we can ensure this by possibly adding some local terms h_{ij} which are equal to zero. We imagine that $|v\rangle$ may be the output of some approximation algorithm such as the ones described above. Our aim is to efficiently compute a state with energy larger than $\langle v|H|v\rangle$. Moreover, we would like to increase this energy by an amount proportional to $|E|$ in order to guarantee that the *approximation ratio* is improved by some additive constant. We show that this is possible if the following two conditions hold:

(i) The variance of the energy, defined by

$$\mathrm{Var}_v(H) = \langle v|H^2|v\rangle - \langle v|H|v\rangle^2,$$

satisfies $\mathrm{Var}_v(H) = \Omega(|E|)$ [2].

(ii) The state $|v\rangle$ is a product state. That is, $|v\rangle = |v_1\rangle \otimes |v_2\rangle \otimes \ldots \otimes |v_n\rangle$ where each $|v_i\rangle$ is a single-qubit state.

Theorem 170. *Given a product state $|v\rangle$ and a Hamiltonian H as in Eq. (6.1) defined on a d-regular graph, we can efficiently compute a depth-$(d+1)$ quantum circuit U such that the state $|\psi\rangle = U|v\rangle$ satisfies*

$$\langle\psi|H|\psi\rangle \geq \langle v|H|v\rangle + \Omega\left(\frac{\mathrm{Var}_v(H)^2}{d^2|E|}\right). \tag{6.2}$$

This result applies broadly to quantum optimization problems, but does not provide any improvement when specialized to the classical setting. To see this, note that condition (i) is not satisfied in the purely classical case where $|v\rangle$ is a computational basis state and H is diagonal in the computational basis. Indeed, we have $\mathrm{Var}_v(H) = 0$ whenever $|v\rangle$ is an eigenstate of H. On the other hand, condition (i) is fairly mild in the quantum setting and we later show it holds for

[2] i.e. there is a universal constant c, such that asymptotically $\mathrm{Var}_v(H) \geq c \cdot |E|$

a generic product state $|v\rangle$ whenever the Hamiltonian contains nontrivial interactions on each edge of the graph. Since G is d-regular, the number of terms in the sum is $O(d|E|)$. So condition (i) is satisfied if the sum is proportional to the number of terms appearing in it.

Simple examples demonstrate that neither of the two conditions alone is enough to even guarantee the existence of a state with approximation ratio better than $|v\rangle$ for large regular graphs. Condition (ii) alone is not sufficient because it is possible for a product state to have maximal energy $\lambda_{\max}(H)$ (i.e., this occurs for all classical Hamiltonians). To see that condition (i) is not sufficient, one can consider the Max-Cut Hamiltonian on (say) an even cycle graph, and let $|v\rangle$ be an equal superposition of two eigenstates of H, one with maximal energy $|E|$ and one with energy $|E| - \Theta(\sqrt{|E|})$. The resulting state has approximation ratio $1 - O(|E|^{-1/2})$ and variance $\mathrm{Var}_v(H) = \Omega(|E|)$. Thus condition (i) is satisfied, but the approximation ratio cannot be improved by an additive constant.

In the special case where $|v\rangle$ achieves the largest energy of any product state, we are able to strengthen the bound Eq. (6.2). We say that the product state $|v\rangle$ is *locally optimal* for H if for any single-qubit Pauli Q, we have

$$\frac{d}{d\phi}\langle v|e^{-i\phi Q}He^{i\phi Q}|v\rangle\big|_{\phi=0} = 0,$$

or equivalently $\langle v|[Q,H]|v\rangle = 0$. As we later show in Claim 177 in Section 6.3.2, the bound in Eq. (6.2) can be improved to $\langle v|H|v\rangle + \Omega(\frac{\mathrm{Var}_v(H)^2}{d|E|})$ for locally optimal states.

Generally, however, the improvement stated in Eq. (6.2) is optimal in the sense that there exists a Hamiltonian H and a product state $|v\rangle$ with $\mathrm{Var}_v(H) = \Theta(|E|)$ for which

$$\lambda_{\max}(H) - \langle v|H|v\rangle \leq O\left(\frac{\mathrm{Var}_v(H)^2}{d^2|E|}\right). \tag{6.3}$$

For example, Eq. (6.3) is satisfied by the Hamiltonian with $h_{ij} = Z_i + Z_j$ on any d-regular graph and the product state $|v\rangle = (\cos(\theta)|0\rangle + \sin(\theta)|1\rangle)^{\otimes n}$, for any $\theta \in (0, \pi/2)$. In this simple case, the left-hand side can be computed exactly and is equal to $\frac{\mathrm{Var}_v(H)^2}{d^2|E|} \cdot \frac{\sin^2(\theta)}{\sin^4(2\theta)}$.

To establish Theorem 170, we consider a variational family of states obtained from $|v\rangle = \otimes_{i\in V}|v_i\rangle$ by applying a quantum circuit composed of nearest neighbor commuting gates on the interaction graph G. In particular, let $P_1, P_2, \ldots, P_n$ be any collection of single-qubit operators such that $\|P_i\| \leq 1$ and

$$\langle v_i|P_i|v_i\rangle = 0 \quad \text{for all} \quad i \in V.$$

Following [AGM20], we define the circuit

$$V(\vec{\theta}) = \prod_{\{i,j\}\in E} e^{i\theta_{ij}P_iP_j} = e^{i\sum_{\{i,j\}\in E}\theta_{ij}P_iP_j}. \tag{6.4}$$

Here, $\vec{\theta}$ is an array of real parameters $\{\theta_{ij}\}_{\{i,j\}\in E}$. Since by assumption, the interaction graph G is d-regular, the quantum circuit $V(\vec{\theta})$ can be implemented with circuit depth $d+1$. It is not hard to see that this variational family includes as a special case the level-1 Quantum Approximate Optimization Algorithm (QAOA) for 2-local classical Hamiltonians [FGG14]. For a given choice of operators $\{P_i\}_{i\in V}$, the following theorem lower bounds the improvement in the energy after applying

the the quantum circuit $V(\vec{\theta})$ to $|v\rangle$.

Theorem 171. *Let $|v\rangle$ be a product state and $|\psi\rangle = V(\vec{\theta})|v\rangle$ be the state prepared by the quantum circuit Eq. (6.4). Define the positive real parameter α by*

$$\alpha = \mathbb{E}_{\{i,j\}\in E}|\langle v_i, v_j|[P_iP_j, h_{ij}]|v_i, v_j\rangle|, \tag{6.5}$$

where the expectation is with respect to the uniform distribution over the edges. There is an efficient classical algorithm to select parameters $\vec{\theta}$ satisfying

$$\langle\psi|H|\psi\rangle \geq \langle v|H|v\rangle + \Omega\big(|E|\alpha^2/d\big). \tag{6.6}$$

Proof. Write N_{ij} for the set of edges $\{k,\ell\} \in E$ incident to a given edge $\{i,j\} \in E$. The latter edge is included as well, i.e., $\{i,j\} \in N_{ij}$. Consider the energy of a term

$$\langle\psi|h_{ij}|\psi\rangle = \langle v|V(\vec{\theta})^\dagger h_{ij} V(\vec{\theta})|v\rangle.$$

The gates in $V(\vec{\theta})$ which are associated with edges that are not incident with $\{i,j\}$ can be cancelled, leaving $\langle v|V_{ij}^\dagger h_{ij} V_{ij}|v\rangle$ where $V_{ij} = \prod_{\{k,\ell\}\in N_{ij}} e^{i\theta_{k\ell}P_kP_\ell}$. Thus

$$\begin{aligned}\langle\psi|h_{ij}|\psi\rangle = \langle v|h_{ij}|v\rangle + \\ \sum_{m=1}^{\infty}\frac{i^m}{m!}\langle v|\Bigg[\sum_{\{k,\ell\}\in N_{ij}} -\theta_{k\ell}P_kP_\ell, h_{ij}\Bigg]_m|v\rangle.\end{aligned} \tag{6.7}$$

Here, $[A,B]_m$ is the m-nested commutator $[A,[A,\ldots[A,B]]]$. Using the fact that $\langle v_k|P_k|v_k\rangle = 0$ for all k, the $m=1$ term simplifies to

$$\sum_{\{k,\ell\}\in N_{ij}} -i\theta_{k\ell}\langle v|\left[P_kP_\ell, h_{ij}\right]|v\rangle = -i\theta_{ij}\langle v|[P_iP_j, h_{ij}]|v\rangle. \tag{6.8}$$

At this stage, we make the choice

$$\theta_{ij} = \theta\cdot\operatorname{sign}\left(-i\langle v|[P_iP_j, h_{ij}]|v\rangle\right), \tag{6.9}$$

where the parameter θ will be determined later. Substituting in Eq. (6.8) gives

$$\sum_{\{k,\ell\}\in N_{ij}} -i\theta_{k\ell}\langle v|\left[P_kP_\ell, h_{ij}\right]|v\rangle = \theta|\langle v_i, v_j|[P_iP_j, h_{ij}]|v_i, v_j\rangle|. \tag{6.10}$$

For $m > 1$, we have

$$\left|\langle v|\left[\sum_{\{k,\ell\}\in N_{ij}} -\theta_{k\ell}P_kP_\ell, h_{ij}\right]_m|v\rangle\right| \leq \sum_{\substack{\{k_1,\ell_1\},\{k_2,\ell_2\},\ldots \\ \{k_m,\ell_m\}\in N_{ij}}} \theta^m \left|\langle v|\left[P_{k_1}P_{\ell_1},[\ldots,[P_{k_m}P_{\ell_m},h_{ij}]]\right]|v\rangle\right|.$$

The only nonzero terms are those in which the expression $\langle v_s|P_s|v_s\rangle$ does not appear. To upper bound the number of nonzero terms, we count the number of tuples $(\{k_1,\ell_1\},\{k_2,\ell_2\},\ldots\{k_m,\ell_m\})$ such that no vertex in $V\setminus\{i,j\}$ appears exactly once. An upper bound is as follows (a proof is provided in Section 6.3.1).

Claim 172. *Let $m \geq 2$. The number of ordered tuples of edges $(\{k_1,\ell_1\},\{k_2,\ell_2),\ldots,\{k_m,\ell_m\}) \in N_{ij}^{\times m}$ in which no vertex in $V\setminus\{i,j\}$ appears exactly once is at most $(2m\sqrt{d})^m$.*

Finally, using Eq. (6.9) and the fact that $\|h_{ij}\|, \|P_i\| \leq 1$, we can upper bound

$$\theta^m \left|\langle v|\left[P_{k_1}P_{\ell_1},[\ldots,[P_{k_m}P_{\ell_m},h_{ij}]]\right]|v\rangle\right| \leq (2\theta)^m.$$

Thus, the sum of all $m > 1$ terms in Eq. (6.7) has magnitude at most

$$\sum_{m=2}^{\infty}\frac{1}{m!}\left(4m\sqrt{d}\right)^m\theta^m \leq \sum_{m=0}^{\infty}\left(4e\sqrt{d}\theta\right)^{m+2} \leq 32e^2d\theta^2$$

assuming $\theta \leq \frac{1}{8e\sqrt{d}}$ (where we used the bound $m^m/m! \leq e^m$). Combining with Eqs. (6.7,6.10) and summing over all $\{i,j\}\in E$, we get

$$\langle\psi|H|\psi\rangle \geq \langle v|H|v\rangle + |E|\left(\theta\alpha - 32e^2d\theta^2\right).$$

We may then choose $\theta = O(\alpha/d)$ to get the desired lower bound. □

Let us now see how Theorem 170 is obtained as a consequence of Theorem 171 (the detailed proof is provided in Section 6.3.2). The lower bound (6.6) applies to any choice of operators $\{P_i\}_{i\in V}$. We will choose these operators in a way that gives the variance bound Eq. (6.2). In the following, for convenience and without loss of generality, we shall work in a local basis in which our initial product state is $|v\rangle \equiv |0^n\rangle$. Our starting point is the observation that the variance of a 2-local Hamiltonian can be expressed in this basis as

$$\mathrm{Var}_v(H) = \langle 0^n|HQ_1H|0^n\rangle + \langle 0^n|HQ_2H|0^n\rangle,$$

where Q_t is the projector onto computational basis states with Hamming weight $t\in\{1,2\}$. This implies that

$$\langle 0^n|HQ_tH|0^n\rangle \geq \mathrm{Var}_v(H)/2 \tag{6.11}$$

for some $t\in\{1,2\}$. Suppose $t=2$ and let X_i,Y_i, and Z_i be the Pauli operators. We define α_1 to be the RHS of Eq. (6.5) with $P_i = X_i$ for all i, and similarly α_2 with $P_i = (X_i+Y_i)/\sqrt{2}$ for all i.

By a direct calculation we see that

$$\alpha_1 = \frac{2}{|E|} \sum_{\{i,j\}\in E} |\mathrm{Im}\left(\langle 11|h_{ij}|00\rangle\right)|$$
$$\alpha_2 = \frac{2}{|E|} \sum_{\{i,j\}\in E} |\mathrm{Re}\left(\langle 11|h_{ij}|00\rangle\right)| \tag{6.12}$$

and therefore

$$\langle 0^n|HQ_2H|0^n\rangle = \sum_{\{i,j\}\in E} |\langle 11|h_{ij}|00\rangle|^2 \leq |E|\left(\frac{\alpha_1+\alpha_2}{2}\right).$$

This means $\max\{\alpha_1,\alpha_2\} \geq |E|^{-1}\langle 0^n|HQ_2H|0^n\rangle$ which together with Eq. (6.11) implies that when $t=2$, we can efficiently find a series of operators P_i such that the parameter α satisfies $\alpha \geq (2|E|)^{-1}\mathrm{Var}_v(H)$. By plugging this in Eq. (6.6), we obtain $\langle\psi|H|\psi\rangle \geq \langle v|H|v\rangle + \Omega(\frac{\mathrm{Var}_v(H)^2}{d|E|})$. Thus if $t=2$ we obtain a better lower bound than the one claimed in Theorem 170. Otherwise, if $t=1$, then a simple calculation (the detailed proof is given in Section 6.3.2) shows that one can efficiently compute a product state with energy at least $\langle v|H|v\rangle + \Omega(\frac{\mathrm{Var}(H)^2}{d^2|E|})$. In general, the choice between $t=1$ and $t=2$ can be efficiently determined. Thus we obtain Theorem 170. In Section 6.3.2, we show that if $|v\rangle$ is locally optimal for H, then $\langle 0^n|HQ_1H|0^n\rangle = 0$ and $t=2$, so we obtain the better bound described above.

Let us briefly illustrate how these results can be applied to the quantum Max-Cut Hamiltonian considered in Refs. [GP19, AGM20]. The Hamiltonian is built from local terms $h_{ij} = w_{ij}\Pi_{ij}$, where $0 \leq w_{ij} \leq 1$ and $\Pi_{ij} = (\mathbb{1} - X_iX_j - Y_iY_j - Z_iZ_j)/4$ is the projector onto the antisymmetric state of two qubits. This Hamiltonian has the special feature that any product state $|v\rangle$ is locally optimal, and moreover, we have $|\langle v_i^\perp, v_j^\perp|h_{ij}|v_i,v_j\rangle| = \langle v_i,v_j|h_{ij}|v_i,v_j\rangle$. Therefore

$$\mathrm{Var}_v(H) = \sum_{\{i,j\}\in E} \langle v|h_{ij}|v\rangle^2 \geq |E|^{-1}\langle v|H|v\rangle^2$$

using Cauchy-Schwarz. We may then efficiently compute a state $|\psi\rangle$ such that

$$\langle\psi|H|\psi\rangle \geq \langle v|H|v\rangle + \Omega\left(\frac{\langle v|H|v\rangle^4}{d|E|^3}\right). \tag{6.13}$$

We see that if the initial state has approximation ratio $\langle v|H|v\rangle/|E| = r$ then the state $|\psi\rangle$ improves this to $r + \Omega(r^4/d)$ [3].

This example demonstrates the power of Theorem 170 and shows that for the quantum Max-Cut problem, the approximation ratio of *any* product state can be improved by applying a shallow quantum circuit. For more general two-local Hamiltonians, we can guarantee an improvement in the approximation ratio whenever the condition $\mathrm{Var}_v(H) = \Omega(|E|)$ holds, which we expect for typical product states and Hamiltonians. Below we discuss two natural extensions of our results. First, we ask whether one can improve approximation ratios attained by more general families of quantum

[3] A better bound can be obtained by directly computing the parameter α for a randomized choice of operators $\{P_i\}$. In that case, $\mathbb{E}\alpha \geq \Omega(\langle v|H|v\rangle/|E|)$ which results in an improvement of $\Omega\left(\langle v|H|v\rangle^2/(d|E|)\right)$

states. Along these lines, we provide an extension of Theorem 170 to the more general case where $|v\rangle$ is any state prepared by a quantum circuit of depth $D = O(1)$. Next, we show how one can improve the approximation ratio achieved by a random product state $|v\rangle$. Using Theorem 171, we show that the approximation ratio can be improved by $\Omega(1/d)$ for any Hamiltonian with nontrivial two-local interactions, and by $\Omega(1/\sqrt{d})$ if the interaction graph is triangle-free.

6.2.2 Improvement of bounded-depth entangled states

Recall that for any n-qubit quantum circuit and any qubit $j \in [n]$, we may define the lightcone $\mathcal{L}(j) \subseteq [n]$ which consists of all output qubits that are causally connected to j. Define the maximum lightcone size $\ell = \max_{j\in[n]} \mathcal{L}(j)$. We have $\ell \leq 2^D$ for any depth D circuit composed of two-qubit gates.

Theorem 173. *Let $|v\rangle = W|0^n\rangle$ where W is a quantum circuit with maximum lightcone size ℓ. There is an efficient classical algorithm that computes a quantum circuit U such that $|\psi\rangle = U|v\rangle$ satisfies*

$$\langle\psi|H|\psi\rangle = \langle v|H|v\rangle + \Omega\left(\frac{\mathrm{Var}(H)^2}{\ell^{10}d^2|E|}\right).$$

For constant-depth circuits we have $\ell = O(1)$ and we get the same asymptotic energy improvement as we established previously in Theorem 170 for product states. However, in this case the circuit U that we construct is not constant-depth. In Section 6.3.4, we show that the improvement stated above can also be obtained for states $|v\rangle$ that are the unique ground states of a gapped ℓ-local Hamiltonian F. In that case, ℓ is replaced by the locality of the Hamiltonian F. Thus, Theorem 173 extends to a broad class of tensor network states (such as PEPS of low bond dimension) that have a gapped parent Hamiltonian.

The theorem provides limitations on the energy that can be achieved by any state $|v\rangle$ produced by a bounded-depth circuit. In particular, since $\langle\psi|H|\psi\rangle \leq \lambda_{\max}(H)$, we find that

$$\langle v|H|v\rangle \leq \lambda_{\max}(H) - \Omega\left(\frac{\mathrm{Var}(H)^2}{\ell^{10}d^2|E|}\right).$$

This shows that the approximation ratio achievable by constant-depth states $|v\rangle$ with $\mathrm{Var}(H) = \Omega(|E|)$ is bounded away from 1. An interesting direction for future work is to explore whether one can use this fact to exhibit new local Hamiltonian systems with the almost-linear NLTS (No Low-energy Trivial States) property [AN22, FH14].

6.2.3 Improvement of random assignments

Given an instance of a (classical) constraint satisfaction problem, one may consider the trivial algorithm in which each variable is chosen independently and uniformly at random. Remarkably, efficient algorithms which improve over the approximation ratio achieved by this simple strategy are not likely to exist in the general case [Has01]. On the other hand, for structured cases such as bounded-degree graphs, improvement is possible. In particular, on degree-d graphs, one can efficiently find an assignment satisfying a $\mu + \Omega(\frac{1}{d})$ fraction of constraints [Has00]. Here μ is the expected fraction of constraints satisfied by a uniformly random assignment. It has been shown that when a degree-d graph is triangle-free, there are efficient "local" algorithms that find a binary

string satisfying a $\mu+\Omega(\frac{1}{\sqrt{d}})$ fraction of constraints by starting with a uniformly random assignment [BMO+15, Has19] or quantum superposition [FGG15] and then locally updating each bit/qubit as a function of the state of its neighbors.

Below we show that this optimal dependence on d can be recovered and generalized to the local Hamiltonian setting by applying our algorithm in Theorem 171 to a randomly chosen product state. For randomly chosen $|v\rangle$, the parameter α in Theorem 171 can be related to the 2-norm of the quadratic terms in the Pauli expansion of the Hamiltonian. More precisely, for an n-qubit operator $O = \sum_{i<j} \sum_{x,y} f_{xy}^{ij} \sigma_x^i \otimes \sigma_y^j$ where $\sigma_0 = I$ and $\{\sigma_1, \sigma_2, \sigma_3\}$ are the Pauli matrices, we define

$$\mathrm{quad}(O) = \sum_{i<j} \sum_{x>0, y>0} (f_{xy}^{ij})^2.$$

Theorem 174. *There is an efficient randomized algorithm that given a Hamiltonian H Eq. (6.1) defined on a d-regular graph, computes a depth-$d+1$ circuit U such that $|\psi\rangle = U|v\rangle$ satisfies*

$$\mathbb{E}_v \langle\psi|H|\psi\rangle \geq \mathbb{E}_v \langle v|H|v\rangle + \Omega\left(\frac{\mathrm{quad}(H)^2}{d|E|}\right).$$

If the graph is triangle-free then the right-hand side can be replaced with $\mathbb{E}_v \langle v|H|v\rangle + \Omega\left(\frac{\mathrm{quad}(H)}{\sqrt{d}}\right)$.

The proof of Theorem 174 is provided in Section 6.3.3. We also show in Section 6.3.6 that for triangle-free graphs one can efficiently compute *product* states matching the approximation ratios quoted above using a local classical algorithm similar to the ones described in Refs. [Has19, BMO+15, HM17]. Thus, low-depth quantum circuits are not necessary to achieve the asymptotic $\Omega(1/\sqrt{d})$ scaling in this case. Similar results have been shown before in [HM17] for general 2-local Hamiltonians. Nevertheless, one may take the output product state of such algorithms and improve it *further* using the shallow quantum circuit from Theorem 170.

6.3 Deferred proofs

6.3.1 Proof of Claim 172

Proof. First, we count all such tuples $(\{k_1, \ell_1\}, \{k_2, \ell_2\}, \ldots \{k_m, \ell_m\})$ in which the edge $\{i, j\}$ does not appear. Each one can be generated by choosing a tuple of vertices $(v_1, v_2, \ldots v_m)$ incident to $\{i, j\}$ and then specifying a neighbor, either i or j, for each of them. An upper bound is obtained by counting the number of tuples $(v_1, v_2, \ldots v_m)$ such that each v_p occurs at least twice and then multiplying by 2^m. Any tuple $(v_1, v_2, \ldots v_m)$ of this form can be generated as follows. First, for each $i = 1, 2, \ldots, m$ we choose a color $c(i) \in \{1, 2 \ldots, m/2\}$. We set $v_k = v_{k'}$ whenever $c(k) = c(k')$. We then assign a neighbor of i or j to each color $\{1, 2, \ldots, m/2\}$. Since vertices i and j each have at most d neighbors, we see that the number of tuples $(v_1, v_2, \ldots v_m)$ such that each v_p occurs at least twice is at most $(m/2)^m \cdot (2d)^{m/2}$. The number of tuples of edges $(\{k_1, \ell_1\}, \{k_2, \ell_2\}, \ldots \{k_m, \ell_m\}\} \in N_{ij}^{\times m}$ in which no vertex in $V \setminus \{i, j\}$ appears exactly once, and the edge $\{i, j\}$ does not occur, is then at most $2^m \cdot (m/2)^m \cdot (2d)^{m/2}$.

In order to account for the appearance of the edge $\{i, j\}$, we fix the number of places u where

the edge appears and then count as before for the $m-u$ places. This number is

$$\sum_{u=0}^{m} \binom{m}{u} ((m-u)\sqrt{2d})^{m-u} \leq (2m\sqrt{d})^m.$$

□

6.3.2 Improvement of product states: proof of Theorem 170

In this section we provide the full details of the proof of Theorem 170. It will be convenient to work in a local basis defined by $|v\rangle$, such that $|v\rangle = |0^n\rangle$ and

$$\mathrm{Var}_v(H) = \langle 0^n|H^2|0^n\rangle - (\langle 0^n|H|0^n\rangle)^2.$$

For ease of notation we write $\mathrm{Var}(H) = \mathrm{Var}_v(H)$. Recall the quantity α defined in Eq. (6.5):

$$\alpha = \mathbb{E}_{\{i,j\}\in E} |\langle v_i, v_j|[P_i P_j, h_{ij}]|v_i, v_j\rangle|, \tag{6.14}$$

We will use the following proposition.

Proposition 175. *Let Q_2 be the projector onto computational basis states with Hamming weight 2. We can efficiently choose operators $\{P_i\}_{i\in V}$ such that*

$$\alpha \geq \frac{1}{|E|} \cdot \langle 0^n|HQ_2H|0^n\rangle. \tag{6.15}$$

Proof. Let α_1 be Eq. (6.5) with $P_i = X_i$ for all i, and let α_2 be Eq. (6.5) with $P_i = (X_i + Y_i)/\sqrt{2}$ for all i. Direct calculation shows that

$$\begin{aligned} \alpha_1 &= \frac{2}{|E|} \sum_{\{i,j\}\in E} \left|\mathrm{Im}\left(\langle 11|_{ij} \otimes \langle 0^{n-2}|h_{ij}|0^n\rangle\right)\right| \\ \alpha_2 &= \frac{2}{|E|} \sum_{\{i,j\}\in E} \left|\mathrm{Re}\left(\langle 11|_{ij} \otimes \langle 0^{n-2}|h_{ij}|0^n\rangle\right)\right| \end{aligned} \tag{6.16}$$

We can express $\langle 0^n|HQ_2H|0^n\rangle$ as

$$\begin{aligned} \langle 0^n|HQ_2H|0^n\rangle. &= \sum_{\{i,j\}\in E} |\langle 11|_{ij} \otimes \langle 0^{n-2}|h_{ij}|0^n\rangle|^2 \\ &= \sum_{\{i,j\}\in E} \left(\mathrm{Im}\left(\langle 11|_{ij} \otimes \langle 0^{n-2}|h_{ij}|0^n\rangle\right)\right)^2 + \left(\mathrm{Re}\left(\langle 11|_{ij} \otimes \langle 0^{n-2}|h_{ij}|0^n\rangle\right)\right)^2 \\ &\leq \sum_{\{i,j\}\in E} \left|\mathrm{Im}\left(\langle 11|_{ij} \otimes \langle 0^{n-2}|h_{ij}|0^n\rangle\right)\right| + \left|\mathrm{Re}\left(\langle 11|_{ij} \otimes \langle 0^{n-2}|h_{ij}|0^n\rangle\right)\right| \\ &= |E| \cdot \frac{\alpha_1 + \alpha_2}{2}, \end{aligned}$$

where we used the fact that $\|h_{ij}\| \leq 1$ in going from the second to the third line above. Now the

last line implies that either α_1 or α_2 achieves the bound from Eq. (6.15). Moreover, the choice of α_1 or α_2 can be efficiently determined. □

Proof of Theorem 170. Let Q_t be the projector onto computational basis states with Hamming weight $t \in \{1, 2\}$. Since H is two-local we have

$$\mathrm{Var}(H) = \langle 0^n|HQ_1H|0^n\rangle + \langle 0^n|HQ_2H|0^n\rangle.$$

Therefore $\langle 0^n|HQ_tH|0^n\rangle \geq \mathrm{Var}(H)/2$ for some $t \in \{1, 2\}$. If $t = 2$ then we may use Proposition 175 which gives

$$\max\{\alpha_1, \alpha_2\} \geq \frac{1}{2|E|}\mathrm{Var}(H).$$

Combining this with Theorem 171, we arrive at

$$\langle\psi|H|\psi\rangle \geq \langle 0^n|H|0^n\rangle + \Omega\left(\frac{\mathrm{Var}(H)^2}{d|E|}\right)$$

which is better than the desired lower bound.

Next suppose $\langle 0^n|HQ_1H|0^n\rangle \geq \mathrm{Var}(H)/2$. Define

$$L = \sum_{j=1}^{n}(-1)^{a_j}P_j$$

where each P_j is a single-qubit Pauli operator acting nontrivially only on qubit j, and $a_j \in \{0, 1\}$ is chosen so that

$$i\langle 0^n|(-1)^{a_j}[P_j, H]|0^n\rangle = |\langle 0^n|[P_j, H]|0^n\rangle|.$$

Define $|\theta\rangle = e^{-i\theta L}|0^n\rangle$ where θ is a real parameter that we will fix later. Then

$$\langle\theta|H|\theta\rangle = \langle 0^n|H|0^n\rangle + \theta\sum_{j=1}^{n}|\langle 0^n|[P_j, H]|0^n\rangle| + \mathrm{Err},$$

where

$$\begin{aligned}|\mathrm{Err}| &= \left|\sum_{m\geq 2}\frac{i^m\theta^m}{m!}\langle 0^n|[L, H]_m|0^n\rangle\right| \\ &\leq |E|\sum_{m\geq 2}\frac{\theta^m 4^m}{m!} \\ &\leq 16\theta^2|E|e^{4\theta}.\end{aligned}$$

In the second line we used the fact that

$$[L, h_{ij}]_m = [(-1)^{a_i}P_i + (-1)^{a_j}P_j, h_{ij}]_m$$

can be expanded as a sum of 2^m terms each of norm at most 2^m. Now define

$$\beta = \frac{1}{|E|}\sum_{j=1}^{n} |\langle 0^n|[P_j, H]|0^n\rangle|. \tag{6.17}$$

and note that since $\|[P_j, H]\| \leq 2d$ for all j we have

$$\beta \leq \frac{1}{|E|}\sum_{j=1}^{n} 2d \leq 4.$$

Then

$$\langle\theta|H|\theta\rangle \geq \langle 0^n|H|0^n\rangle + |E|\left(\theta\beta - 16\theta^2 e^{4\theta}\right).$$

Choosing $\theta = \beta/32$ gives

$$\begin{aligned}
\langle\theta|H|\theta\rangle &\geq \langle 0^n|H|0^n\rangle + |E|\left(\frac{\beta^2}{32} - \frac{\beta^2}{64}e^{\beta/8}\right) \\
&\geq \langle 0^n|H|0^n\rangle + |E|\left(\frac{\beta^2}{32} - \frac{\beta^2}{64}e^{1/2}\right) \\
&\geq \langle 0^n|H|0^n\rangle + 0.001\cdot|E|\beta^2.
\end{aligned} \tag{6.18}$$

Now let β_1 be given by Eq. (6.17) with $P_i = X_i$ for all i, and let β_2 be given by Eq. (6.17) with $P_i = Y_i$ for all i. Then

$$\begin{aligned}
\frac{\beta_1+\beta_2}{2} &= \frac{1}{2|E|}\sum_{j=1}^{n} |\langle 0^n|[X_j, H]|0^n\rangle| + |\langle 0^n|[Y_j, H]|0^n\rangle| \\
&\geq \frac{1}{4d|E|}\sum_{j=1}^{n} |\langle 0^n|[X_j, H]|0^n\rangle|^2 + |\langle 0^n|[Y_j, H]|0^n\rangle|^2 \\
&= \frac{1}{4d|E|}\sum_{j=1}^{n} |2\mathrm{Im}\left(\langle\hat{e}_j|H|0^n\rangle\right)|^2 + |2\mathrm{Re}\left(\langle\hat{e}_j|H|0^n\rangle\right)|^2 \\
&= \frac{1}{d|E|}\langle 0^n|HQ_1H|0^n\rangle \\
&\geq \frac{1}{2d|E|}\mathrm{Var}(H).
\end{aligned}$$

Therefore either β_1 or β_2 is larger than the RHS above. Plugging this into Eq. (6.18) we arrive at

$$\langle\theta|H|\theta\rangle \geq \langle 0^n|H|0^n\rangle + \Omega\left(\frac{\mathrm{Var}(H)^2}{d^2|E|}\right).$$

□

The improvement in energy $\langle v|H|v\rangle$ in Theorem 170 is extensive, proportional to $|E|$, when $\mathrm{Var}_v(H) = \Omega(|E|)$. Here we argue that this condition on the variance is mild in the quantum

setting. One way to see this is using the following expression for the variance:

$$\mathrm{Var}_v(H) = \sum_{\{i,j\}\cap\{k,l\}\neq\emptyset} \left(\langle v|h_{ij}h_{kl}|v\rangle - \langle v|h_{ij}|v\rangle\cdot\langle v|h_{kl}|v\rangle\right). \tag{6.19}$$

Since G is d-regular, the number of terms in the sum is $O(d|E|)$. So condition $\mathrm{Var}_v(H) = \Omega(|E|)$ is satisfied if the sum is proportional to the number of terms appearing in it.

A more precise argument can be given by considering the expectation of $\mathrm{Var}_v(H)$ for a fixed Hamiltonian and random product states. Suppose the Hamiltonian H in the Pauli basis is given by $H = \sum_{\{i,j\}\in E}\sum_{a,b} f^{ij}_{a,b}\sigma^i_a\otimes\sigma^j_b$ where $\sigma_0 = I$ and $\{\sigma_1,\sigma_2,\sigma_3\}$ are the Pauli matrices. We define as before $\mathrm{quad}(H) = \sum_{\{i,j\}\in E}\sum_{a>0,b>0}(f^{ij}_{a,b})^2$. By choosing vectors $|v_1\rangle,\ldots,|v_n\rangle$ uniformly at random, we get $\mathbb{E}_v\langle v|\sigma^i_a\sigma^i_b|v\rangle = \delta_{a,b}$ and $\mathbb{E}_v\langle v|\sigma^i_a\otimes\sigma^j_b|v\rangle = \delta_{a,0}\delta_{b,0}$. We also have $\mathbb{E}_v\langle v|\sigma^i_a|v\rangle\langle v|\sigma^i_b|v\rangle = 1/3$ if $a = b\neq 0$ and $\mathbb{E}_v\langle v|\sigma^i_a|v\rangle\langle v|\sigma^i_b|v\rangle = 0$ if $a\neq b$. To derive the last two equations, we can write

$$\mathbb{E}_v\langle v|\sigma^i_a|v\rangle\langle v|\sigma^i_b|v\rangle = \langle 00|\cdot\mathbb{E}_{U\sim\mathrm{Haar}}\left[(U^{\otimes 2})^\dagger\cdot\sigma_a\otimes\sigma_b\cdot U^{\otimes 2}\right]\cdot|00\rangle.$$

It follows from the properties of the Haar measure (see a similar derivation in Eq. (6.26)) that

$$\mathbb{E}_{U\sim\mathrm{Haar}}\left[(U^{\otimes 2})^\dagger\cdot\sigma_a\otimes\sigma_b\cdot U^{\otimes 2}\right] = \frac{1}{3}\left(\mathrm{tr}(\sigma_a\otimes\sigma_b)(I-\frac{1}{2}S) + \mathrm{tr}(\sigma_a\otimes\sigma_b\cdot S)(S-\frac{1}{2}I)\right),$$

where S is the swap operator. From here, a direct calculation leads to the claimed equations. Using these equations, we see that when $i = k$ and $j\neq l$ in Eq. (6.19), the only terms that have a non-zero contribution in the sum are of the form

$$\langle v|(\sigma^i_a\otimes\sigma^j_0)(\sigma^i_a\otimes\sigma^l_0)|v\rangle - \langle v|\sigma^i_a\otimes\sigma^j_0|v\rangle\cdot\langle v|\sigma^i_a\otimes\sigma^l_0|v\rangle,$$

with $a>0$. When $i = k$ and $j = l$, the relevant terms involve

$$\langle v|(\sigma^i_a\otimes\sigma^j_b)(\sigma^i_a\otimes\sigma^j_b)|v\rangle - \langle v|\sigma^i_a\otimes\sigma^j_b|v\rangle\cdot\langle v|\sigma^i_a\otimes\sigma^j_b|v\rangle..$$

Substituting the expectation of Eq. (6.19), we arrive at the following expression for $\mathbb{E}_v\mathrm{Var}_v(H)$:

$$\mathbb{E}_v\mathrm{Var}_v(H) = \frac{8}{9}\sum_{\{i,j\}\in E}\sum_{a>0,b>0}(f^{ij}_{a,b})^2 + \frac{2}{3}\sum_{i\in V,a>0}\left(\sum_{j:\{i,j\}\in E} f^{ij}_{a,0}\right)^2 \geq \frac{8}{9}\mathrm{quad}(H) \tag{6.20}$$

For typical Hamiltonians where the coefficients $f^{ij}_{a,b}$ with $a>0, b>0$ are non-zero constants, we see that $\mathbb{E}_v\mathrm{Var}_v(H) = \Omega(|E|)$.

Finally, let us discuss a special case in which the bound from Theorem 170 can be improved. We say that a product state $|v\rangle$ is locally optimal for H if, for any single-qubit Pauli Q we have

$$\frac{d}{d\phi}\langle v|e^{-i\phi Q}He^{i\phi Q}|v\rangle\big|_{\phi=0} = 0,$$

or equivalently

$$\langle v|[Q,H]|v\rangle = 0. \tag{6.21}$$

As in the above, for simplicity we shall work in a local basis defined by v, so that $|v\rangle = |0^n\rangle$.

Claim 176. *Suppose $|0^n\rangle$ is locally optimal for H. Then for any string $z \in \{0,1\}^n$ with Hamming weight $|z| = 1$ we have*

$$\langle z|H|0^n\rangle = 0. \tag{6.22}$$

Proof. Without loss of generality consider the case where $z = 10^{n-1}$. Then

$$|2\mathrm{Im}(\langle z|H|0^n\rangle)| = |\langle 0^n|[X_1, H]|0^n\rangle| = 0 \quad \text{and} \quad |2\mathrm{Re}(\langle z|H|0^n\rangle)| = |\langle 0^n|[Y_1, H]|0^n\rangle| = 0,$$

where we used Eq. (6.21). □

Claim 177. *Suppose $|0^n\rangle$ is locally optimal for H. We may efficiently choose $\{P_i\}$ and $\{\theta_{ij}\}$ such that*

$$\langle\psi|H|\psi\rangle \geq \langle 0^n|H|0^n\rangle + \Omega\left(\frac{\mathrm{Var}(H)^2}{d|E|}\right). \tag{6.23}$$

Proof. Since H is two-local we have

$$\mathrm{Var}(H) = \langle 0^n|HQ_1H|0^n\rangle + \langle 0^n|HQ_2H|0^n\rangle = \langle 0^n|HQ_2H|0^n\rangle,$$

where in the last equality we used claim 176. The claim then follows directly by combining Proposition 175 and Theorem 171. □

6.3.3 Improvement of random states: proof of Theorem 174

General degree-d graphs

We prove the first part of Theorem 174 regarding general degree-d graphs, which is implied by the following lemma.

Lemma 178. *Let $|v\rangle = |v_1\rangle \otimes |v_2\rangle \ldots \otimes |v_n\rangle$ where each v_i is a Haar random single-qubit state. Then there is an efficient randomized process with random coins r, that constructs the matrices P_i (depending on both r and $|v\rangle$) such that the resulting state $|\psi_{r,v}\rangle$ satisfies*

$$\mathbb{E}_{r,v}\langle\psi_{r,v}|H|\psi_{r,v}\rangle \geq \mathbb{E}_v\langle v|H|v\rangle + \Omega\left(\frac{\mathrm{quad}(H)^2}{d|E|}\right).$$

Proof. Pick $|v\rangle = \otimes_i|v_i\rangle$, where each $|v_i\rangle$ is chosen uniformly at random from Haar measure on qubits. Also choose n uniformly random real numbers μ_i i.i.d in the interval $[0, \frac{\pi}{2}]$. The latter choice is made using the coins r. Given $|v\rangle, r$ define $P_i = e^{i\mu_i}|v_i\rangle\langle v_i^\perp| + e^{-i\mu_i}|v_i^\perp\rangle\langle v_i|$ (we drop the labels $|v\rangle, r$ from P_i for convenience). Observe that $\langle v_i|P_i|v_i\rangle = 0$ and $\|P_i\| \leq 1$, as required. Then $\alpha_{v,r}$ (as given in Eq (6.5)) can be evaluated to be

$$\begin{aligned}\alpha_{v,r} &= \mathbb{E}_{\{i,j\}\in E}\left|\left(e^{i(\mu_i+\mu_j)}\langle v_i^\perp, v_j^\perp|h_{ij}|v_i, v_j\rangle - e^{-i(\mu_i+\mu_j)}\langle v_i, v_j|h_{ij}|v_i^\perp, v_j^\perp\rangle\right)\right| \\ &= 2\mathbb{E}_{\{i,j\}\in E}\left|\mathrm{Im}\left(e^{i(\mu_i+\mu_j)}\langle v_i^\perp, v_j^\perp|h_{ij}|v_i, v_j\rangle\right)\right|. \end{aligned} \tag{6.24}$$

Let $\langle v_i^\perp, v_j^\perp|h_{ij}|v_i, v_j\rangle = e^{i\kappa_{i,j}}|\langle v_i^\perp, v_j^\perp|h_{ij}|v_i, v_j\rangle|$ be the polar decomposition. Then

$$|\mathrm{Im}\left(e^{i(\mu_i+\mu_j)}\langle v_i^\perp, v_j^\perp|h_{ij}|v_i, v_j\rangle\right)| = |\sin(\mu_i+\mu_j+\kappa_{i,j})| \cdot |\langle v_i^\perp, v_j^\perp|h_{ij}|v_i, v_j\rangle|.$$

Note that

$$\mathbb{E}_r|\sin(\mu_i+\mu_j+\kappa_{i,j})| = \frac{4}{\pi^2}\int_0^{\frac{\pi}{2}}\int_0^{\frac{\pi}{2}}|\sin(\mu_i+\mu_j+\kappa_{i,j})|d\mu_i d\mu_j \geq \frac{2}{5},$$

for all $\kappa_{i,j}$. Then Eq 6.24 ensures that

$$\mathbb{E}_r\alpha_{v,r} = 2\mathbb{E}_{\{i,j\}\in E}\mathbb{E}_r|\mathrm{Im}\left(e^{i(\mu_i+\mu_j)}\langle v_i^\perp, v_j^\perp|h_{ij}|v_i, v_j\rangle\right)| \geq \frac{4}{5}\cdot\mathbb{E}_{\{i,j\}\in E}|\langle v_i^\perp, v_j^\perp|h_{ij}|v_i, v_j\rangle|.$$

Then we can evaluate

$$\begin{aligned}
&\mathbb{E}_{v,r}\alpha_{v,r} \geq \frac{4}{5}\cdot\mathbb{E}_{\{i,j\}\in E}\int|\left(\langle v_i^\perp, v_j^\perp|h_{ij}|v_i, v_j\rangle\right)|dv_i dv_j\\
&\geq \frac{4}{5}\cdot\mathbb{E}_{\{i,j\}\in E}\int|\langle v_i^\perp, v_j^\perp|h_{ij}|v_i, v_j\rangle|^2 dv_i dv_j\\
&= \frac{4}{5}\cdot\mathbb{E}_{\{i,j\}\in E}\int\mathrm{tr}\left(|v_i^\perp, v_j^\perp\rangle\langle v_i^\perp, v_j^\perp|h_{ij}|v_i, v_j\rangle\langle v_i, v_j|h_{ij}\right)dv_i dv_j\\
&= \frac{4}{5}\cdot\mathbb{E}_{\{i,j\}\in E}\int\langle 11|\left(U_i^\dagger\otimes V_j^\dagger\right)h_{ij}\left(U_i\otimes V_j\right)|00\rangle\langle 00|\left(U_i^\dagger\otimes V_j^\dagger\right)h_{ij}\left(U_i\otimes V_j\right)|11\rangle dU_i dV_j\\
&= \frac{4}{5}\cdot\mathbb{E}_{\{i,j\}\in E}\int\langle 1100|\left(U_{i1}^\dagger\otimes V_{j1}^\dagger\otimes U_{i2}^\dagger\otimes V_{j2}^\dagger\right)h_{i1,j1}\otimes h_{i2,j2}\left(U_{i1}\otimes V_{j1}\otimes U_{i2}\otimes V_{j2}\right)|0011\rangle dU_i dV_j,
\end{aligned} \tag{6.25}$$

where in the second last equality we fixed a basis $\{|0\rangle, |1\rangle\}$ for each qubit and introduced random unitaries U_i, V_j that specify $|v_i\rangle = U_i|0\rangle, |v_j\rangle = V_j|0\rangle$. Using the well known properties of Haar integral, we have

$$\begin{aligned}
&\int\left(U_{i1}^\dagger\otimes V_{j1}^\dagger\otimes U_{i2}^\dagger\otimes V_{j2}^\dagger\right)h_{i1,j1}\otimes h_{i2,j2}\left(U_{i1}\otimes V_{j1}\otimes U_{i2}\otimes V_{j2}\right)dU_i dV_j\\
&= a\mathrm{id}_{i1,i2}\otimes\mathrm{id}_{j1,j2} + bS_{i1,i2}\otimes\mathrm{id}_{j1,j2} + c\mathrm{id}_{i1,i2}\otimes S_{j1,j2} + dS_{i1,i2}\otimes S_{j1,j2}.
\end{aligned} \tag{6.26}$$

Above, id is the identity operator, S is the swap operator and the subscripts represent the qubits on which the operator acts. Coefficients a, b, c, d can be evaluated using the following system of equations, obtained from Eq. 6.26 by tracing each of the four operators.

$$\begin{aligned}
(\mathrm{tr}_{i,j}h_{ij})^2 &= 16a + 8b + 8c + 4d\\
\mathrm{tr}_j\left(\mathrm{tr}_i h_{ij}\mathrm{tr}_i h_{ij}\right) &= 8a + 4b + 16c + 8d\\
\mathrm{tr}_i\left(\mathrm{tr}_j h_{ij}\mathrm{tr}_j h_{ij}\right) &= 8a + 16b + 4c + 8d\\
\mathrm{tr}_{i,j}\left(h_{ij}^2\right) &= 4a + 8b + 8c + 16d.
\end{aligned}$$

One can solve for d to obtain

$$d = \frac{(\mathrm{tr}_{i,j} h_{ij})^2}{36} + \frac{\mathrm{tr}_{i,j}\left(h_{ij}^2\right)}{9} - \frac{\mathrm{tr}_j\left(\mathrm{tr}_i h_{ij} \mathrm{tr}_i h_{ij}\right) + \mathrm{tr}_i\left(\mathrm{tr}_j h_{ij} \mathrm{tr}_j h_{ij}\right)}{18}.$$

In order to obtain a simpler lower bound and see that d is positive, we expand $h_{ij} = \sum_{x,y} f_{x,y}^{i,j} \sigma_x^i \otimes \sigma_y^j$ in the two qubit Pauli basis. Then

$$(\mathrm{tr}_{i,j} h_{ij})^2 = 16\left(f_{0,0}^{i,j}\right)^2, \quad \mathrm{tr}_{i,j}\left(h_{ij}^2\right) = 4\sum_{x,y}\left(f_{x,y}^{i,j}\right)^2,$$

$$\mathrm{tr}_j\left(\mathrm{tr}_i h_{ij} \mathrm{tr}_i h_{ij}\right) = 8\sum_{y}\left(f_{0,y}^{i,j}\right)^2, \quad \mathrm{tr}_i\left(\mathrm{tr}_j h_{ij} \mathrm{tr}_j h_{ij}\right) = 8\sum_{y}\left(f_{y,0}^{i,j}\right)^2.$$

Hence,

$$\begin{aligned}
&\frac{(\mathrm{tr}_{i,j} h_{ij})^2}{36} + \frac{\mathrm{tr}_{i,j}\left(h_{ij}^2\right)}{9} - \frac{\mathrm{tr}_j\left(\mathrm{tr}_i h_{ij} \mathrm{tr}_i h_{ij}\right) + \mathrm{tr}_i\left(\mathrm{tr}_j h_{ij} \mathrm{tr}_j h_{ij}\right)}{18} \\
&= \frac{4}{9}\left(\left(f_{0,0}^{i,j}\right)^2 + \sum_{x,y}\left(f_{x,y}^{i,j}\right)^2 - \sum_{y}\left(f_{0,y}^{i,j}\right)^2 - \sum_{y}\left(f_{y,0}^{i,j}\right)^2\right) \\
&= \frac{4}{9}\left(\sum_{x>0,y>0}\left(f_{x,y}^{i,j}\right)^2\right).
\end{aligned}$$

Conjugating Eq. 6.26 with $\langle 1100| (\cdot) |0011\rangle$, it can be seen that only the term corresponding to d survives and evaluates to 1. Thus, Eq. 6.25 gives

$$\begin{aligned}
&\mathbb{E}_{v,r}\alpha_{v,r} \geq \\
&\frac{4}{5}\mathbb{E}_{\{i,j\}\in E}\left(\frac{(\mathrm{tr}_{i,j} h_{ij})^2}{36} + \frac{\mathrm{tr}_{i,j}\left(h_{ij}^2\right)}{9} - \frac{\mathrm{tr}_j\left(\mathrm{tr}_i h_{ij} \mathrm{tr}_i h_{ij}\right) + \mathrm{tr}_i\left(\mathrm{tr}_j h_{ij} \mathrm{tr}_j h_{ij}\right)}{18}\right) \\
&= \frac{16}{45}\mathbb{E}_{\{i,j\}\in E}\left(\sum_{x>0,y>0}\left(f_{x,y}^{i,j}\right)^2\right) = \frac{16}{45}\frac{\mathrm{quad}(H)}{|E|}.
\end{aligned}$$

Thus, using the convexity of square function,

$$\mathbb{E}_{v,r}\alpha_{v,r}^2 \geq \frac{16^2}{45^2}\left(\frac{\mathrm{quad}(H)}{|E|}\right)^2 \geq \frac{1}{8}\left(\frac{\mathrm{quad}(H)}{|E|}\right)^2.$$

This completes the proof by employing Theorem 171. □

Triangle-free graphs

In this section we establish the second part of Theorem 174, which concerns triangle-free graphs. The proof is based on the following exact expression. It will be convenient in what follows to work

in a local basis in which the product state of interest is $|v\rangle = |0^n\rangle$.

Lemma 179 (Improvement for triangle-free Hamiltonians). *Suppose G is a triangle-free, degree-d graph. Suppose we are given single-qubit Hermitian operators $\{P_i\}_{i\in[n]}$ satisfying $P_i^2 = I$ and $\langle 0|P_i|0\rangle = 0$ for all $i \in [n]$, and consider the state $|\psi\rangle = e^{i\sum_{\{r,s\}\in E}\theta_{rs}P_rP_s}|0^n\rangle$ as a function of the real parameters $\{\theta_{rs}\}$. Define*

$$\alpha_{kl} = |\langle 00|[h_{kl}, P_kP_l]|00\rangle|$$

We can efficiently choose $\theta_{ij} \in \{\pm\theta\}$ for each edge $\{i,j\} \in E$ so that

$$\langle\psi|h_{kl}|\psi\rangle = \frac{1}{4}\mathrm{Tr}(h_{kl}) + \frac{1}{4}\mathrm{Tr}(h_{kl}Z_kZ_l)\cos^{2d-2}(2\theta) + \frac{1}{4}\mathrm{Tr}(h_{kl}(Z_k+Z_l))\cos^d(2\theta). + \frac{\alpha_{kl}}{2}\sin(2\theta)\cos^{d-1}(2\theta) \tag{6.27}$$

for all edges $\{k,l\} \in E$

Proof. We have

$$\langle\psi|h_{kl}|\psi\rangle = \langle 0^n|V_{kl}^\dagger h_{kl}(\theta)V_{kl}|0^n\rangle \tag{6.28}$$

where $h_{kl}(\theta_{kl}) = e^{-i\theta_{kl}P_kP_l}h_{kl}e^{i\theta_{kl}P_kP_l}$ and

$$V_{kl} = \prod_{\{k,s\}\in E\setminus\{k,l\}} e^{i\theta_{ks}P_kP_s} \prod_{\{r,l\}\in E\setminus\{k,l\}} e^{i\theta_{rl}P_rP_l} \tag{6.29}$$

$$= \prod_{\{k,s\}\in E\setminus\{k,l\}} (\cos(\theta) + i\sin(\theta_{ks})P_kP_s) \prod_{\{r,l\}\in E\setminus\{k,l\}} (\cos(\theta) + i\sin(\theta_{rl})P_rP_l)\,. \tag{6.30}$$

Plugging Eq. (6.30) into Eq. (6.28) and using $\langle 0|P_i|0\rangle = 0$ and the fact that G is triangle-free gives

$$\langle\psi|h_{kl}|\psi\rangle = \sum_{A\subseteq N(k)\setminus\{l\}} \sum_{B\subseteq N(l)\setminus\{k\}} \left(\cos^2(\theta)\right)^{2d-2-|A|-|B|} \left(\sin^2(\theta)\right)^{|A|+|B|} \cdot \langle 0^n| \left(\prod_{s\in A} P_kP_s \prod_{r\in B} P_rP_l\right) h_{kl}(\theta_{kl}) \left(\prod_{s\in A} P_kP_s \prod_{r\in B} P_rP_l\right) |0^n\rangle$$

In the above we also used our choice $|\theta_{ij}| = \theta$ for all edges $\{i,j\} \in E$. Observe that the matrix element appearing in the above depends only on the parity (even/odd) of $|A|$ and $|B|$. In particular,

$$\langle\psi|h_{kl}|\psi\rangle = F_{EE} + F_{EO} + F_{OE} + F_{OO}$$

where the even/even term is

$$F_{EE} = \langle 00|h_{kl}(\theta_{kl})|00\rangle \left(\sum_{j=0,2,\ldots} \binom{d-1}{j} \left(\cos^2(\theta)\right)^{d-1-j} \left(\sin^2(\theta)\right)^j\right)^2 \tag{6.31}$$

$$= \langle 00|h_{kl}(\theta_{kl})|00\rangle \frac{1}{4}\left(1 + \cos^{d-1}(2\theta)\right)^2, \tag{6.32}$$

and by similar calculations one arrives at

$$F_{EO} = \langle 10|h_{kl}(\theta_{kl})|10\rangle \frac{1}{4}\left(1 - \cos^{2d-2}(2\theta)\right) \tag{6.33}$$

$$F_{OE} = \langle 01|h_{kl}(\theta_{kl})|01\rangle \frac{1}{4}\left(1 - \cos^{2d-2}(2\theta)\right) \tag{6.34}$$

$$F_{OO} = \langle 11|h_{kl}(\theta_{kl})|11\rangle \frac{1}{4}\left(1 - \cos^{d-1}(2\theta)\right)^2 \tag{6.35}$$

Now for ease of presentation in the following we write $c = \cos^{d-1}(2\theta)$ and $a_{xy} = \langle xy|h_{kl}(\theta_{kl})|xy\rangle$, for $x, y \in \{0, 1\}$. Then expanding the above expression gives

$$\langle\psi|h_{kl}|\psi\rangle = F_{EE} + F_{EO} + F_{OE} + F_{OO} \tag{6.36}$$

$$= a_{00} + (1-c)\left(\frac{a_{01} + a_{10}}{2} - a_{00}\right) + \frac{1}{4}(1-c)^2\left(a_{11} + a_{00} - a_{01} - a_{10}\right). \tag{6.37}$$

Now let

$$b_{xy} = \langle xy|h_{kl}|xy\rangle. \tag{6.38}$$

So that

$$a_{00} = \cos^2(\theta)b_{00} + \sin^2(\theta)b_{11} + i\cos(\theta)\sin(\theta_{kl})\langle 00|[h_{kl}, P_kP_l]|00\rangle.$$

We now fix the sign of θ_{kl} so that

$$a_{00} = \cos^2(\theta)b_{00} + \sin^2(\theta)b_{11} + \frac{1}{2}\sin(2\theta)\alpha_{kl}. \tag{6.39}$$

With this choice we have

$$a_{11} = \cos^2(\theta)b_{11} + \sin^2(\theta)b_{00} - \frac{1}{2}\sin(2\theta)\alpha_{kl}. \tag{6.40}$$

To compute the third term in the above equation we used the fact that $P_k^2 = P_l^2 = I$ and $\langle 0|P_k|0\rangle = \langle 0|P_l|0\rangle = 0$ which implies

$$\langle 11|[P_kP_l, h_{kl}]|11\rangle = \langle 00|P_kP_l[P_kP_l, h_{kl}]P_kP_l|00\rangle = -\langle 00|[P_kP_l, h_{kl}]|00\rangle.$$

Similarly, by a direct calculation we get

$$a_{01} + a_{10} = b_{10} + b_{01}. \tag{6.41}$$

Plugging Eqs. (6.39, 6.40, 6.41) into Eq. (6.37) we get

$$\begin{aligned}\langle\psi|h_{kl}|\psi\rangle &= b_{00} + (a_{00} - b_{00})c + (1-c)\left(\frac{b_{01}+b_{10}}{2} - b_{00}\right) + \frac{1}{4}(1-c)^2\left(b_{00}+b_{11}-b_{10}-b_{01}\right) \\ &= b_{00} + \frac{\alpha_{kl}}{2}\sin(2\theta)\cos^{d-1}(2\theta) + \sin^2(\theta)\cos^{d-1}(2\theta)\left(b_{11}-b_{00}\right) \\ &+ (1-\cos^{d-1}(2\theta))\left(\frac{b_{01}+b_{10}}{2} - b_{00}\right) + \frac{1}{4}(1-\cos^{d-1}(2\theta))^2\left(b_{00}+b_{11}-b_{10}-b_{01}\right)\end{aligned} \tag{6.42}$$

Rearranging the above expression we arrive at

$$\begin{aligned}\langle\psi|h_{kl}|\psi\rangle &= \frac{\alpha_{kl}}{2}\sin(2\theta)\cos^{d-1}(2\theta) + b_{00}\left(\frac{1}{4} + \frac{1}{4}\cos^{2d-2}(2\theta) + \frac{1}{2}\cos^{d}(2\theta)\right) \\ &+ b_{11}\left(\frac{1}{4} + \frac{1}{4}\cos^{2d-2}(2\theta) - \frac{1}{2}\cos^{d}(2\theta)\right) + (b_{01}+b_{10})\left(\frac{1}{4} - \frac{1}{4}\cos^{2d-2}(2\theta)\right)\end{aligned} \tag{6.43}$$

By noting that $\sum_{x,y} b_{xy} = \mathrm{Tr}(h_{kl})$, $\sum_{x,y}(-1)^{x+y}b_{xy} = \mathrm{Tr}(h_{kl}Z_kZ_l)$, and $b_{00} - b_{11} = \frac{1}{2}\mathrm{Tr}(h_{kl}(Z_k+Z_l))$, we arrive at Eq. (6.27).

□

Using the expression in (6.27), we prove the bound for triangle-free graphs from Theorem 174:

Proof. As shown above, the exact formula for the energy of $|\psi\rangle = V(\vec{\theta})|0^n\rangle$ on a triangle-free graph is

$$\begin{aligned}\langle\psi|h_{kl}|\psi\rangle &= \frac{1}{4}\mathrm{Tr}(h_{kl}) + \frac{1}{4}\mathrm{Tr}(h_{kl}Z_kZ_l)\cos^{2d-2}(2\theta) \\ &+ \frac{1}{4}\mathrm{Tr}(h_{kl}(Z_k+Z_l))\cos^{d}(2\theta). + \frac{\alpha_{kl}}{2}\sin(2\theta)\cos^{d-1}(2\theta).\end{aligned} \tag{6.44}$$

Here α_{kl} depends on the choices of P_k, P_l. We either choose $P_i = X_i$ for all i, or $P_i = (X+Y)_i/\sqrt{2}$ for all i, each with probability $1/2$. Then

$$\mathbb{E}(\alpha_{kl}) = 2|\mathrm{Re}(\langle 00|h_{kl}|11\rangle)| + 2|\mathrm{Im}(\langle 00|h_{kl}|11\rangle)| \geq 2|\langle 00|h_{kl}|11\rangle|.$$

Substituting in Eq. (6.44) gives

$$\begin{aligned}\mathbb{E}\left(\langle\psi|h_{kl}|\psi\rangle\right) &\geq \frac{1}{4}\mathrm{Tr}(h_{kl}) + \frac{1}{4}\mathrm{Tr}(h_{kl}Z_kZ_l)\cos^{2d-2}(2\theta) \\ &+ \frac{1}{4}\mathrm{Tr}(h_{kl}(Z_k+Z_l))\cos^{d}(2\theta). + |\langle 00|h_{kl}|11\rangle|\sin(2\theta)\cos^{d-1}(2\theta).\end{aligned} \tag{6.45}$$

Now instead of using the starting state $|0^n\rangle$, suppose we start from a random computational basis state $|s\rangle = X(s)|0^n\rangle$. Running through the above argument in the rotated basis defined by $|s\rangle$ we

see that for a suitable random choice of $\{P_i\}$ we have

$$\begin{aligned}\mathbb{E}\left(\langle\psi|h_{kl}|\psi\rangle\right) &\geq \frac{1}{4}\left(\mathrm{Tr}(h_{kl})\right) + \frac{|\langle 00|h_{kl}|11\rangle| + |\langle 01|h_{kl}|10\rangle|}{2}\sin(2\theta)\cos^{d-1}(2\theta) \\ &\geq \frac{1}{4}\left(\mathrm{Tr}(h_{kl})\right) + \frac{1}{4}|\mathrm{Tr}(h_{kl}X_kX_l)|\sin(2\theta)\cos^{d-1}(2\theta). \end{aligned} \tag{6.46}$$

Here we used the fact that

$$\mathbb{E}_s\left(\mathrm{Tr}(X(s)h_{kl}X(s)Z_kZ_l)\right) = \mathbb{E}_s\left(\mathrm{Tr}(X(s)h_{kl}X(s)(Z_k+Z_l))\right) = 0.$$

Summing Eq. (6.46) over all edges $\{k,l\} \in E$ gives

$$\mathbb{E}\left(\langle\psi|H|\psi\rangle\right) \geq \frac{1}{4}\left(\mathrm{Tr}(H)\right) + \frac{1}{4}\sin(2\theta)\cos^{d-1}(2\theta)\sum_{\{k,l\}\in E}\left|\mathrm{Tr}\left(h_{kl}X_kX_l\right)\right|$$

Since there is nothing special about the X-basis we can again use our freedom to randomize the local basis of each qubit to get

$$\begin{aligned}\mathbb{E}\left(\langle\psi|H|\psi\rangle\right) &\geq \frac{1}{4}\mathrm{Tr}(H) + \frac{1}{36}\sin(2\theta)\cos^{d-1}(2\theta)\sum_{\{k,l\}\in E}\sum_{Q,R\in\{X,Y,Z\}}\left|\mathrm{Tr}\left(h_{kl}Q_kR_l\right)\right| \\ &\geq \frac{1}{4}\mathrm{Tr}(H) + \frac{1}{36}\sin(2\theta)\cos^{d-1}(2\theta)\sum_{\{k,l\}\in E}\sum_{Q,R\in\{X,Y,Z\}}\left|\mathrm{Tr}\left(h_{kl}Q_kR_l\right)\right|^2/4 \\ &= \frac{1}{4}\mathrm{Tr}(H) + \sin(2\theta)\cos^{d-1}(2\theta)\frac{\mathrm{quad}(H)}{36}, \end{aligned} \tag{6.47}$$

where in the second-to-last line we used the fact that $|\mathrm{Tr}\left(h_{kl}Q_kR_l\right)| \leq 4$ which follows from $\|h_{kl}\| \leq 1$. Finally, we can find the maximum value of the second term with respect to θ by noting that $\sin(2\theta)\cos^{d-1}(2\theta)$ reaches a maximum when $\theta = \arcsin(\frac{1}{\sqrt{d}})$. Using this fact we get

$$\mathbb{E}(\langle\psi|H|\psi\rangle) \geq \frac{1}{4}\mathrm{Tr}(H) + \Omega\left(\frac{\mathrm{quad}(H)}{\sqrt{d}}\right). \tag{6.48}$$

□

6.3.4 Improvement of bounded-depth entangled states: proof of Theorem 173

We prove Theorem 173. Given the d-regular graph $G = (V, E)$, we consider the state $|v\rangle = W|0\rangle^n$, where W has a maximum lightcone of size ℓ. The aim is to increase the energy of $|v\rangle$ with respect to H. The light cones of the edges have sizes at most 2ℓ. Define

$$F = \sum_{j=1}^{n} W|1\rangle\langle 1|_j W^\dagger. \tag{6.49}$$

The locality of F is ℓ. Let $A = i[H, F]$ and define $|\psi\rangle = e^{iA\theta}|v\rangle$ (thus $U = e^{iA\theta}$ in the statement of Theorem 173). We can write

$$A = \sum_{e\in E} i[h_e, F] := \sum_{e\in E} A_e, \tag{6.50}$$

where

$$A_e = i\sum_{j=1}^{n}[h_e, W|1\rangle\langle 1|_j W^\dagger] = i\sum_{j:\mathrm{supp}(h_e)\cap\mathrm{supp}(W|1\rangle\langle 1|_j W^\dagger)\neq\phi}[h_e, W|1\rangle\langle 1|_j W^\dagger]. \tag{6.51}$$

Any j satisfying $\mathrm{supp}(h_e)\cap\mathrm{supp}(W|1\rangle\langle 1|_j W^\dagger)\neq\phi$ is in the light cone of h_e. Thus there are $\leq 2\ell$ such j's. For any such j, $W|1\rangle\langle 1|_j W^\dagger$ has locality ℓ. Thus, A_e is supported on $\leq 2\ell^2$ qubits. Further,

$$\|A_e\| \leq 2\ell \cdot \max_{j:\mathrm{supp}(h_e)\subset\mathrm{supp}(W|1\rangle\langle 1|_j W^\dagger)} \|[h_e, W|1\rangle\langle 1|_j W^\dagger]\| \leq 2\ell,$$

where we used

$$\|[h_e, W|1\rangle\langle 1|_j W^\dagger]\| = \|[h_e, W|1\rangle\langle 1|_j W^\dagger - \mathrm{id}/2]\| \leq 2\|h_e\|\|W|1\rangle\langle 1|_j W^\dagger - I/2\| \leq 1.$$

We have

$$\langle\psi|H|\psi\rangle = \langle v|H|v\rangle - i\langle v|[A,H]|v\rangle\theta + \sum_{m=2}^{\infty}\frac{(-i\theta)^m}{m!}\langle v|[A,H]_m|v\rangle, \tag{6.52}$$

Now, using the identities $F|v\rangle = 0$ and $F \geq \mathrm{I} - |v\rangle\langle v|$, we find

$$-i\langle v|[A,H]|v\rangle\theta = \langle v|[[H,F],H]|v\rangle\theta = 2\langle v|HFH|v\rangle\theta \geq 2\theta\langle v|H(\mathrm{I}-|v\rangle\langle v|)H|v\rangle = 2\theta\mathrm{Var}(H). \tag{6.53}$$

Thus, let us focus on the terms with $m \geq 2$. We upper bound

$$\langle v|[A,H]_m|v\rangle \leq \sum_{e\in E}\|[A,h_e]_m\| \leq |E|\max_e\|[A,h_e]_m\|. \tag{6.54}$$

Now, consider for each e,

$$[A,h_e]_m = \sum_{e_1,\ldots e_m}[A_{e_m},[A_{e_{m-1}}\ldots[A_{e_1},h_e]]], \tag{6.55}$$

where we used Eq. 6.50. Most terms are zero and we will bound the number of non-zero terms. We will use the following simple fact.

Fact 180. *Let $S\subset V$. The number of e such that the support of A_e overlaps with S is at most $|S|\ell^2 d$. For each such e and any operator O_S on S , the support of $[O_S, A_e]$ is at most $|S|+2\ell^2$.*

Proof. Since e is such that the support of A_e overlaps with S, there exist a j satisfying $\mathrm{supp}(h_e)\cap\mathrm{supp}(W|1\rangle\langle 1|_j W^\dagger)\neq\phi$ for which the support of $[h_e, W|1\rangle\langle 1|_j W^\dagger]$ overlaps with S. Thus, either $\mathrm{supp}(h_e)\cap S\neq\phi$ or j belongs to the light cone of S. Since the support of h_e overlaps with the light cone of j in the latter case, we have that the support of h_e overlaps with the light cone of the light cone of S (in both the cases). To upper bound the number of possible e, we hence we count the size of the light cone of the light cone of S ($\leq |S|\ell^2$) and then count the number of edges intersecting with this light cone ($\leq d|S|\ell^2$). For any such e, the support of $[O_S, A_e]$ is contained in the union of S and the support of A_e. This completes the proof. □

Using Fact 180, let us estimate the number of $(e_1, e_2, \ldots e_m)$ that contribute to Eq. 6.55. Setting S to be the set of two vertices of e, we find that the number of e_1 is at most $2d\ell^2$. Arguing

inductively, suppose we have fixed $e_1, e_2, \dots e_{k-1}$. The support size of $[A_{e_{k-1}}, [A_{e_{m-1}} \dots [A_{e_1}, h_e]]]$ is at most $2 + 2(k-1)\ell^2$ (by Fact 180). Thus, the number of e_k contributing to Eq. 6.55 is at most $(2 + 2(k-1)\ell^2)\ell^2 d$. Hence, the total number of $(e_1, \dots e_m)$ is at most

$$(2d\ell^2)\cdot(2+2\ell^2)\ell^2 d\cdot(2+4\ell^2)\ell^2 d \dots (2+2(m-1)\ell^2)\ell^2 d \leq 2^m\cdot(m-1)!\cdot\ell^{2m-2}\cdot(2d\ell^2)^m \leq (m-1)!(4d\ell^4)^m.$$

Thus,

$$\begin{aligned}
\|[A, h_e]_m\|_\infty &\leq (m-1)!(4d\ell^4)^m \max_{e_1,\dots e_m} \|[A_{e_m}, [A_{e_{m-1}} \dots [A_{e_1}, h_e]]]\| \\
&\leq (m-1)!(4d\ell^4)^m \cdot 2^m \|h_e\| \cdot \max_{e_1,\dots e_m} \|A_{e_1}\| \cdot \|A_{e_2}\| \dots \cdot \|A_{e_m}\| \\
&\overset{(a)}{\leq} (m-1)!(4d\ell^4)^m \cdot 2^m \cdot (2\ell)^m = (m-1)! \cdot (16d\ell^5)^m, \qquad (6.56)
\end{aligned}$$

where (a) uses $\|A_e\| \leq 2\ell$. Combining with Eq. 6.54, this ensures that

$$\begin{aligned}
\sum_{m=2}^{\infty} \left| \frac{(i\theta)^m}{m!} \langle v|[[H, A]]_m|v\rangle \right| &\leq |E| \sum_{m=2}^{\infty} \frac{\theta^m}{m!} (m-1)! \cdot (16d\ell^5)^m \\
&\leq |E| \cdot \sum_{m=2}^{\infty} (16d\ell^5\theta)^m \\
&\leq 2|E| \cdot (16d\ell^5\theta)^2, \qquad (6.57)
\end{aligned}$$

where the last inequality assumes $\theta \leq \frac{1}{32d\ell^5}$ (our choice below will satisfy this). Thus, using Eq. 6.52 and Eq. 6.53,

$$\begin{aligned}
\langle\theta|H|\theta\rangle &\geq \langle v|H|v\rangle + 2\theta \mathrm{Var}(H) - 2|E| \cdot (16d\ell^5\theta)^2 \\
&= \langle v|H|v\rangle + 2\theta|E| \left(\frac{\mathrm{Var}(H)}{|E|} - 2(16d\ell^5)^2\theta \right). \qquad (6.58)
\end{aligned}$$

Setting $\theta = \frac{\mathrm{Var}(H)}{2^{10}d^2\ell^{10}|E|} \leq \frac{1}{32d\ell^5}$, we conclude that

$$\langle\psi|H|\psi\rangle \geq \langle v|H|v\rangle + \frac{\mathrm{Var}(H)^2}{2^{10}d^2\ell^{10}|E|}. \qquad (6.59)$$

We highlight that the above proof can be applied with minor modifications to the more general case in which F is a ℓ-local Hamiltonian with the unique ground state $|v\rangle$ and constant spectral gap $\gamma = \Omega(1)$. In this case, we set the ground energy of F at 0, leading to the relations $F|v\rangle = 0$ and $F \succeq \gamma(\mathrm{id} - |v\rangle\langle v|)$. Thus, the first order contribution in (6.53) is replaced by

$$-i\langle v|[A, H]|v\rangle\theta \geq 2\gamma\theta\mathrm{Var}(H).$$

The higher order contributions are upper bounded in a manner similar to above.

6.3.5 Improvement for general k-local Hamiltonians

Let $G = (V, E)$ be a hypergraph with hyperedges of size at most k and $n = |V|$ qubits on its vertices. We denote the number of hyperedges that contain $i \in V$ by $\deg(i)$ and assume $\deg(i) \leq d$ for all $i \in V$. Consider a k-local Hamiltonian $H = \sum_{R \in E} h_R$ where each local term h_R acts non-trivially on a subset $R \in E$ of qubits with $|R| \leq k$ and $\|h_R\| \leq 1$. Here without loss of generality we assume the input product states is $|v\rangle = |0^n\rangle$. We use a similar argument as in the proof of Theorem 170 to relate the improvement after applying an extension of the quantum circuit $V(\theta)$ in Eq. (6.4) to the variance $\mathrm{Var}(H)$. To this end, we write

$$\mathrm{Var}(H) = \sum_{t=1}^{k} \langle 0^n | H Q_t H | 0^n \rangle$$

where Q_t are the projector onto the computational basis states with Hamming weight t. Note that operators Q_t with Hamming weight $> k$ do not contribute. It holds that there exists a t such that

$$\langle 0^n | H Q_t H | 0^n \rangle \geq \frac{1}{k} \mathrm{Var}(H).$$

Depending on t, our choice of circuit $V(\theta)$ is a generalization of what we had before in Theorem 170. The set S contains all the collection of t different vertices $\{j_1, j_2, \ldots, j_t\}$ which fully reside in the support of at least one local term h_R of the Hamiltonian H. That is, there exists an R such that $\{j_1, j_2, \ldots, j_t\} \subseteq \mathrm{supp}(h_R)$. Define $V(\theta) = e^{-i\theta L}$ where

$$L = \sum_{\{j_1, j_2, \ldots, j_t\} \in S} (-1)^{a_{j_1, \ldots, j_t}} P_{j_1} P_{j_2} \ldots P_{j_t}.$$

Here each P_j is a single-qubit Pauli operator with the property $\langle 0 | P_j | 0 \rangle = 0$ that acts nontrivially only on qubit j and $a_{j_1, \ldots, j_t} \in \{0, 1\}$ is chosen so that

$$i \langle 0^n | (-1)^{a_{j_1, \ldots, j_t}} [P_{j_1} P_{j_2} \ldots P_{j_t}, H] | 0^n \rangle = |\langle 0^n | [P_{j_1} P_{j_2} \ldots P_{j_t}, H] | 0^n \rangle|.$$

Let $|\theta\rangle = e^{-i\theta L} |0^n\rangle$. Then

$$\langle \theta | H | \theta \rangle = \langle 0^n | H | 0^n \rangle + \theta \sum_{\{j_1, \ldots, j_t\} \in S} |\langle 0^n | [P_{j_1} P_{i_2} \ldots P_{j_t}, H] | 0^n \rangle| + \mathrm{Err},$$

where the higher order terms Err can be bounded as

$$\begin{aligned} |\mathrm{Err}| &= \left| \sum_{m \geq 2} \frac{i^m \theta^m}{m!} \langle 0^n | [L, H]_m | 0^n \rangle \right| \\ &\leq |E| \sum_{m \geq 2} \frac{\theta^m}{m!} \left(2kd \binom{k}{t-1} \right)^m \\ &\leq \left(2kd \binom{k}{t-1} \right)^2 \theta^2 |E| e^{2kd \binom{k}{t-1} \theta}. \end{aligned} \tag{6.60}$$

In the second line, we used the fact that $[L,H]_m = \sum_R [L,h_R]_m$ and each term $[L,h_R]$ can be expanded as a sum of at most $\left(kd\binom{k}{t-1}\right)^m$ non-zero terms each of norm at most 2^m. This is because the operators $P_{j_1}, P_{j_2}, \ldots, P_{j_t}$ commute with each other for different choices of $\{j_1, j_2, \ldots, j_t\}$ and only those that overlap with the support of h_R may contribute. The number of such operators (i.e. $|S \cap \mathrm{supp}(h_R)|$) can be bounded by $kd\binom{k}{t-1}$ as follows: There are at most k vertices in $\mathrm{supp}(h_R)$ and each vertex is in the support of $\leq d$ other terms in the Hamiltonian. Given a vertex $j \in \mathrm{supp}(h_R)$ and an overlapping Hamiltonian term $h_{R'}$ such that $j \in \mathrm{supp}(h_{R'})$, there are $\binom{k}{t-1}$ choices of vertices $\{j_1, j_2, \ldots, j_t\} \subseteq \mathrm{supp}(h_{R'})$ that contain j. Hence, from the definition of set S follows that $|S \cap \mathrm{supp}(h_R)| \leq kd\binom{k}{t-1}$ (one can obtain tighter bounds using $\langle 0|P_j|0\rangle = 0$). Now define

$$\beta = \frac{1}{|E|} \sum_{\{j_1,\ldots,j_t\} \in S} |\langle 0^n|[P_{j_1} P_{j_2} \ldots P_{j_t}, H]|0^n\rangle|. \tag{6.61}$$

It holds that $\beta \leq 2\binom{k}{t}$. To see this, note that $|\langle 0^n|[P_{j_1} P_{j_2} \ldots P_{j_t}, H]|0^n\rangle| \leq \sum_R |\langle 0^n|[P_{j_1} P_{j_2} \ldots P_{j_t}, h_R]|0^n\rangle|$. Using the assumption $\langle 0|P_j|0\rangle = 0$, it follows that the only choices of vertices $\{j_1, \ldots, j_t\}$ that may contribute in $|\langle 0^n|[P_{j_1} P_{j_2} \ldots P_{j_t}, h_R]|0^n\rangle|$ are those which are fully contained in $\mathrm{supp}(h_R)$. The number of such vertices is bounded by $\binom{k}{t}$. Using $|\langle 0^n|[P_{j_1} P_{j_2} \ldots P_{j_t}, h_R]|0^n\rangle| \leq 2$, we arrive at the claimed bound $\beta \leq 2\binom{k}{t}$. We have

$$\langle\theta|H|\theta\rangle = \langle 0^n|H|0^n\rangle + |E| \left(\theta\beta - \left(2kd\binom{k}{t-1}\right)^2 \theta^2 e^{2kd\binom{k}{t-1}\theta} \right).$$

Choosing $\theta = O\left(\frac{\beta}{k^2 d^2 \binom{k}{t-1}^2}\right)$ gives

$$\langle\theta|H|\theta\rangle \geq 0^n|H|0^n\rangle + \Omega\left(\frac{|E|\beta^2}{k^2 d^2 \binom{k}{t-1}^2}\right). \tag{6.62}$$

Now let β_1 be given by Eq. (6.61) with $P_{j_1} P_{j_2} \ldots P_{j_t} = X_{j_1} \otimes X_{j_2} \otimes \cdots \otimes X_{j_t}$ for all $\{j_1, \ldots, j_t\} \in S$. Define $|\hat{e}_{j_1,\ldots,j_t}\rangle = X_{j_1} \otimes X_{j_2} \otimes \cdots \otimes X_{j_t}|0^n\rangle$ and the operator

$$p = \begin{pmatrix} 0 & e^{-i\frac{\pi}{2t}} \\ e^{i\frac{\pi}{2t}} & 0 \end{pmatrix}.$$

Let β_2 be given by Eq. (6.61) with $P_{j_1} P_{j_2} \ldots P_{j_t} = p_{j_1} \otimes p_{j_2} \otimes \cdots \otimes p_{j_t}$ for all $\{j_1, \ldots, j_t\} \in S$. Then,

one can see that

$$\begin{aligned}\frac{\beta_1+\beta_2}{2} &= \frac{1}{|E|}\sum_{\{j_1,\ldots,j_t\}\in S} |\mathrm{Im}(\langle \hat{e}_{j_1,\ldots,j_t}|H|0^n\rangle)| + |\mathrm{Re}(\langle \hat{e}_{j_1,\ldots,j_t}|H|0^n\rangle)| \\ &\geq \frac{1}{|E|}\sum_{\{j_1,\ldots,j_t\}\in S} d\cdot\left(\frac{|\mathrm{Im}(\langle \hat{e}_{j_1,\ldots,j_t}|H|0^n\rangle)|^2}{d^2} + \frac{|\mathrm{Re}(\langle \hat{e}_{j_1,\ldots,j_t}|H|0^n\rangle)|^2}{d^2}\right) \\ &= \frac{1}{d|E|}|\langle 0^n|HQ_tH|0^n\rangle| \\ &\geq \frac{1}{kd|E|}\mathrm{Var}(H).\end{aligned}$$

This implies either β_1 or β_2 is larger than $\frac{1}{kd|E|}\mathrm{Var}(H)$. By plugging this into Eq. (6.62) we arrive at

$$\langle\theta|H|\theta\rangle \geq \langle 0^n|H|0^n\rangle + \Omega\left(\frac{\mathrm{Var}(H)^2}{k^4d^4\binom{k}{t-1}^2|E|}\right).$$

This bound is minimized by allowing $t = \lceil k/2+1\rceil$ which results in the following overall lower bound on the improvement to the energy of the input state $|0^n\rangle$:

$$\langle\theta|H|\theta\rangle \geq \langle 0^n|H|0^n\rangle + \Omega\left(\frac{\mathrm{Var}(H)^2}{2^{O(k)}d^4|E|}\right). \tag{6.63}$$

We note that the $\Omega(1/d^4)$ dependence of (6.63) on the degree is quadratically worse than the bound we obtained for 2-local Hamiltonians in Theorem 170; It would be interesting to recover the $\Omega(1/d^2)$ scaling in this case.

6.3.6 Local classical algorithm for triangle-free graphs

Theorem 181. *Consider a two-local Hamiltonian $H = \sum_{\{i,j\}\in E} h_{ij}$ where $G = (V,E)$ is a d-regular triangle-free graph. There is an efficient randomized algorithm that computes a product state $|v\rangle = \otimes_{i=1}^n |v_i\rangle$ satisfying*

$$\mathbb{E}_v\langle v|H|v\rangle \geq \frac{1}{4}\mathrm{Tr}(H) + \Omega\left(\frac{\mathrm{quad}(H)}{\sqrt{d}}\right). \tag{6.64}$$

Note that in Eq. (6.64) the first term on the right hand side is equal to the expected energy of H with respect to a random state $\rho = I/2^n$.

Proof of Theorem 181. It will be convenient to work in a local Pauli basis X, Y or Z chosen at random and independently for each qubit. We write h_{ij} in this randomly chosen product basis. Let us define $w_{ij} = \mathrm{Tr}(h_{ij})/4$ and

$$u_{ij}^x = \frac{1}{4}\mathrm{Tr}(h_{ij}X_iX_j) \quad u_{ij}^y = \frac{1}{4}\mathrm{Tr}(h_{ij}Y_iY_j) \quad u_{ij}^z = \frac{1}{4}\mathrm{Tr}(h_{ij}Z_iZ_j).$$

Due to the random choice of basis we have

$$\mathbb{E}[(u_{ij}^a)^2] = \frac{1}{9}\mathrm{quad}(h_{ij}) \qquad a \in \{x, y, z\}. \tag{6.65}$$

Following [BMO+15, Has19, HM17], we start with a random i.i.d assignment of pure product $|v\rangle = \otimes_{i=1}^n |v_i\rangle$ states to the vertices. We then select a subset of vertices A uniformly at random. For any vertex $i \in A$, let $N(i) = \{\{i, j\} \in E : j \notin A\}$ be the neighboring edges that contain exactly one vertex in A (i.e. vertex i). The remaining edges that are not in $\cup_{i\in A} N(i)$ either connect two vertices that are not in A or connect two vertices in A. We denote the former by M and the latter by M'.

The initial random pure state at each vertex ρ_i can be represented by $\rho_i = \frac{1}{2}(\mathbb{1} + r_i^x X_i + r_i^y Y_i + r_i^z Z_i)$, where $(r_i^x, r_i^y, r_i^z) \in \mathbb{R}^3$ is the Bloch vector with norm $|r_i^x|^2 + |r_i^y|^2 + |r_i^z|^2 = 1$. For a vertex $i \in A$, the total energy of the edges $N(i)$ is given by

$$\begin{aligned}
&\sum_{j:\{i,j\}\in N(i)} \mathrm{tr}[h_{ij}\rho_i \otimes \rho_j] \\
&= \sum_{j:\{i,j\}\in N(i)} w_{ij} + \sum_{j:\{i,j\}\in N(i)} (u_{ij}^x r_i^x r_j^x + u_{ij}^y r_i^y r_j^y + u_{ij}^z r_i^z r_j^z) + \sum_{j:\{i,j\}\in N(i)} D_{ij}(\vec{r}_i, \vec{r}_j)
\end{aligned} \tag{6.66}$$

where

$$D_{ij}(\vec{r}_i, \vec{r}_j) = \sum_{a\neq b} c_{ij}^{ab} r_i^a r_j^b + \sum_{a\in\{x,y,z\}} (d_{ij}^a r_i^a + e_{ij}^a r_j^a).$$

for some coefficients $c_{ij}^{ab}, d_{ij}^a, e_{ij}^a$. Using Cauchy–Schwarz inequality, we see that the first two terms in Eq. (6.66) can be maximized by applying a local unitary on each vertex $i \in A$ which rotates the state ρ_i to a state $\tilde{\rho}_i$ with the Bloch vector

$$R_i^a = \Big(\sum_{j:\{i,j\}\in N(i)} u_{ij}^a r_j^a\Big)\Big(\Big(\sum_{j:\{i,j\}\in N(i)} u_{ij}^x r_j^x\Big)^2 + \Big(\sum_{j:\{i,j\}\in N(i)} u_{ij}^y r_j^y\Big)^2 + \Big(\sum_{j:\{i,j\}\in N(i)} u_{ij}^z r_j^z\Big)^2\Big)^{-1/2} \tag{6.67}$$

for $a \in \{x, y, z\}$. When the denominator of Eq. (6.67) is zero, the vector $\vec{R}$ is chosen uniformly at random. Hence, we get

$$\begin{aligned}
&\sum_{j:\{i,j\}\in N(i)} \mathrm{tr}[h_{ij}\tilde{\rho}_i \otimes \rho_j] \\
&= \sum_{\{i,j\}\in N(i)} w_{ij} + \Big(\Big(\sum_{j:\{i,j\}\in N(i)} u_{ij}^x r_j^x\Big)^2 + \Big(\sum_{j:\{i,j\}\in N(i)} u_{ij}^y r_j^y\Big)^2 + \Big(\sum_{j:\{i,j\}\in N(i)} u_{ij}^z r_j^z\Big)^2\Big)^{1/2} \\
&+ \sum_{j:\{i,j\}\in N(i)} D_{ij}(\vec{R}_i, \vec{r}_j),
\end{aligned} \tag{6.68}$$

A property of this construction is that $\mathbb{E}[R_i^a] = 0$ for $a \in \{x, y, z\}$ and R_i, R_j are independent of each other for $\{i, j\} \in M'$. This follows from the triangle-freeness, the definition of the set $N(i)$, and the initial uniform i.i.d. distribution of the state of vertices. Moreover, we have

$$\mathbb{E}[r_j^a r_k^b] = \mathbb{E}[R_j^a r_k^b] = 0 \quad \text{whenever} \quad a \neq b \quad .$$

Using these observations, the expected value of the total energy after the local improvements is

$$\begin{aligned}
&\mathbb{E}\left[\sum_{\{i,j\}\in M'} \mathrm{tr}[h_{ij}\tilde{\rho}_i \otimes \tilde{\rho}_j]\right] + \mathbb{E}\left[\sum_{\{i,j\}\in M} \mathrm{tr}[h_{ij}\rho_i \otimes \rho_j]\right] + \mathbb{E}\left[\sum_{i\in A}\sum_{j:\{i,j\}\in N(i)} \mathrm{tr}[h_{ij}\tilde{\rho}_i \otimes \rho_j]\right] \\
&= \sum_{\{i,j\}\in E} w_{ij} + \mathbb{E}\left[\sum_{i\in A}\left((\sum_{j:\{i,j\}\in N(i)} u_{ij}^x r_j^x)^2 + (\sum_{j:\{i,j\}\in N(i)} u_{ij}^y r_j^y)^2 + (\sum_{j:\{i,j\}\in N(i)} u_{ij}^z r_j^z)^2\right)^{1/2}\right] \\
&\geq \sum_{\{i,j\}\in E} w_{ij} + \mathbb{E}\left[\sum_{i\in A} |\sum_{j:\{i,j\}\in N(i)} u_{ij}^z r_j^z|\right] \qquad (6.69)
\end{aligned}$$

The first term $\sum_{\{i,j\}\in E} w_{ij}$ corresponds to the expected energy when the product states are chosen uniformly at random. The second term is a lower bound on the improvement achieved by the local updates which we now show is at least $\Omega(|E|/\sqrt{d})$.

For a fixed choice of the set $A \subseteq V$, we define the random variable $\xi_i = \sum_{j:\{i,j\}\in N(i)} u_{ij}^z r_j^z$. Using the "second moment method," for $t \in [0,1]$, we get

$$\Pr\left[|\xi_i| \geq t\sqrt{\mathbb{E}[\xi^2]}\right] \geq (1-t^2)^2 \frac{\mathbb{E}[\xi^2]^2}{\mathbb{E}[\xi^4]} \qquad (6.70)$$

One way to sample uniform pure states over Bloch sphere is to uniformly draw $\phi \sim [0, 2\pi]$, $r_j^z \sim [-1,1]$ and set $r_j^x = \sqrt{1-(r_j^z)^2}\cos\phi$ and $r_j^y = \sqrt{1-(r_j^z)^2}\sin\phi$. Given this we have $\mathbb{E}[r_j^z] = 0$, $\mathbb{E}[(r_j^z)^2] = 1/3$, $\mathbb{E}[(r_j^z)^3] = 0$, and $\mathbb{E}[(r_j^z)^4] = 1/5$. Hence, using Corollary 9.6 of [O'D14], we have $\mathbb{E}[\xi^4] \leq 9 \cdot \mathbb{E}[\xi^2]^2$. Plugging this in (6.70) implies that for a fixed choice of the set A, the expectation with respect to the random distribution of the initial product states for an arbitrary choice of $t \in [0,1]$ is

$$\begin{aligned}
\mathbb{E}\left[|\sum_{j:\{i,j\}\in N(i)} u_{ij}^z r_j^z|\right] &\geq \frac{1}{9}\cdot t(1-t^2)^2 \cdot \mathbb{E}[\xi^2]^{1/2} \\
&= \frac{1}{9}\cdot t(1-t^2)^2 \cdot \left(\sum_{j:\{i,j\}\in N(i)} (u_{ij}^z)^2 \mathbb{E}\left[(r_j^z)^2\right]\right)^{1/2} \\
&\geq \frac{1}{9\sqrt{3}}\cdot t(1-t^2)^2 \cdot \left(\sum_{j:\{i,j\}\in N(i)} (u_{ij}^z)^2\right)^{1/2}. \qquad (6.71)
\end{aligned}$$

Finally, we calculate the expectation with respect to the set $A \subseteq V$. Note that the set $N(i)$ is also a random variable determined by the set A. Conditioned on the event that the vertex $i \in A$ and using Theorem 9.24 of [O'D14], we have

$$\Pr\left[\sum_{j:\{i,j\}\in N(i)} (u_{ij}^z)^2 \geq \frac{1}{2}\sum_{j:\{i,j\}\in E} (u_{ij}^z)^2\right] \geq \frac{1}{4e^2}.$$

Thus, we get

$$\mathbb{E}\left[\sum_{i\in A}\left(\sum_{j:\{i,j\}\in N(i)}(u_{ij}^z)^2\right)^{1/2}\right] \geq \frac{1}{8\sqrt{2}e^2}\sum_{i\in V}\left(\sum_{j:\{i,j\}\in E}(u_{ij}^z)^2\right)^{1/2}$$
$$\geq \frac{1}{4\sqrt{2}e^2}\cdot\frac{1}{\sqrt{d}}\sum_{\{i,j\}\in E}(u_{ij}^z)^2. \tag{6.72}$$

Finally, taking the expectation over the random choice of local basis and using Eq. (6.65) we get

$$\mathbb{E}\left[\sum_{i\in A}\left(\sum_{j:\{i,j\}\in N(i)}(u_{ij}^z)^2\right)^{1/2}\right] \geq \frac{1}{36\sqrt{2}e^2}\cdot\frac{1}{\sqrt{d}}\sum_{\{i,j\}\in E}\mathrm{quad}(h_{ij})$$
$$\geq \frac{1}{36\sqrt{2}e^2}\cdot\frac{1}{\sqrt{d}}\mathrm{quad}(H) \tag{6.73}$$

We arrive at (6.64) by plugging this into (6.71) and using (6.69).

□

Bibliography

[AAB+19] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando GSL Brandao, David A Buell, et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 2019. 4.1

[AAG21] Anurag Anshu, Itai Arad, and David Gosset. An area law for 2d frustration-free spin systems. *arXiv preprint arXiv:2103.02492*, 2021. 3.1.1

[AAJ16] Anurag Anshu, Itai Arad, and Aditya Jain. How local is the information in tensor networks of matrix product states or projected entangled pairs states. *Physical Review B*, 94(19):195143, 2016. 3.4

[AAKS21] Anurag Anshu, Srinivasan Arunachalam, Tomotaka Kuwahara, and Mehdi Soleimanifar. Sample-efficient learning of interacting quantum systems. *Nature Physics*, 17(8):931–935, 2021. 1.8.1, 4

[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, New York, NY, USA, 2009. Association for Computing Machinery. 3.3.4, 3.3.4, 3.8.1, 3.8.1, 3.8.1

[AAR+18] Mohammad H Amin, Evgeny Andriyash, Jason Rolfe, Bohdan Kulchytskyy, and Roger Melko. Quantum Boltzmann machine. *Physical Review X*, 8(2):021050, 2018. 4.4.1

[AAV13] Dorit Aharonov, Itai Arad, and Thomas Vidick. Guest column: The quantum PCP conjecture. *ACM SIGACT News*, 44(2):47–79, 2013. 6.1

[AAV16] Anurag Anshu, Itai Arad, and Thomas Vidick. Simple proof of the detectability lemma and spectral gap amplification. *Physical Review B*, 93:205142, May 2016. 3.8.1, 58, 3.8.1

[Abr19] Nilin Abrahamsen. A polynomial-time algorithm for ground states of spin trees. *arXiv preprint arXiv:1907.04862*, 2019. 3.1.1

[AE11] Dorit Aharonov and Lior Eldar. On the complexity of commuting local Hamiltonians, and tight conditions for topological order in such systems. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science—FOCS 2011*, pages 334–343. IEEE Computer Soc., Los Alamitos, CA, 2011. 4.4.2

[AGM20] Anurag Anshu, David Gosset, and Karen Morenz. Beyond product state approximations for a quantum analogue of max cut. In *15th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2020)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020. 6.1, 6.2.1, 6.2.1

[AGMKS21] Anurag Anshu, David Gosset, Karen J. Morenz Korol, and Mehdi Soleimanifar. Improved approximation algorithms for bounded-degree local hamiltonians. *Phys. Rev. Lett.*, 127:250502, Dec 2021. 1.8.2, 6

[AHL+14] Dorit Aharonov, Aram W. Harrow, Zeph Landau, Daniel Nagaj, Mario Szegedy, and Umesh Vazirani. Local tests of global entanglement and a counterexample to the generalized area law. In *Proceedings of the 2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, FOCS '14, page 246–255, USA, 2014. IEEE Computer Society. 2.1, 3, 3.1.1, 3.1.2, 3.4, 3.7.2, 54, 3.7.3, 1

[AHS20] Anurag Anshu, Aram W Harrow, and Mehdi Soleimanifar. From communication complexity to an entanglement spread area law in the ground state of gapped local Hamiltonians. *To appear in Nature Physics. Preprint available at arXiv:2004.15009*, 2020. 1.8.1, 2.1, 3

[AKL16] Itai Arad, Tomotaka Kuwahara, and Zeph Landau. Connecting global and local energy distributions in quantum spin models on a lattice. *Journal of Statistical Mechanics: Theory and Experiment*, 2016(3):033301, March 2016. 1.7, 3.8.2, 85, 4.6.4, 86, 4.9.2, 4.10.3, 4.10.3

[AKLV13] Itai Arad, Alexei Kitaev, Zeph Landau, and Umesh Vazirani. An area law and sub-exponential algorithm for 1D systems, 2013. arXiv: 1301.1162. 2.1, 3.1.1, 3.2.2, 3.3.4, 3.3.4, 3.8, 3.8.2, 65, 3.8.3

[ALM+98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, 1998. 6.1

[ALV12] Itai Arad, Zeph Landau, and Umesh Vazirani. Improved one-dimensional area law for frustration-free systems. *Physical Review B*, 85:195145, May 2012. 2.1, 3.1.1, 3.2.2, 3.3.4, 3.4, 3.6, 3.8.1

[ALVV17] Itai Arad, Zeph Landau, Umesh Vazirani, and Thomas Vidick. Rigorous RG algorithms and area laws for low energy eigenstates in 1D. *Comm. Math. Phys.*, 356(1):65–105, 2017. 2.1, 3.3.4

[AN22] Anurag Anshu and Chinmay Nirkhe. Circuit lower bounds for low-energy states of quantum code hamiltonians. In *13th Innovations in Theoretical Computer Science Conference (ITCS 2022)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2022. 6.2.2

[Ara69] Huzihiro Araki. *Gibbs states of a one dimensional quantum lattice. Communications in Mathematical Physics*, 14(2):120–157, Jun 1969. 3.4, 4.3.3, 5.1.3

[AS98] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: a new characterization of NP. *J. ACM*, 45(1):70–122, 1998. 6.1

[BAL19] Eyal Bairey, Itai Arad, and Netanel H Lindner. Learning a local Hamiltonian from local measurements. *Physical Review Letters*, 122(2):020504, 2019. 4.1, 4.4.1

[Bar15] Alexander Barvinok. Computing the partition function for cliques in a graph. *Theory Comput.*, 11:339–355, 2015. 5.1.2

[Bar16a] Alexander Barvinok. *Combinatorics and complexity of partition functions*, volume 30 of *Algorithms and Combinatorics*. Springer, Cham, 2016. 5, 5.1.1, 5.1.1, 5.1.3, 5.3, 135, 5.3

[Bar16b] Alexander Barvinok. Computing the permanent of (some) complex matrices. *Found. Comput. Math.*, 16(2):329–342, 2016. 5.1.2

[BBT07] 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. *arXiv preprint arXiv:0705.1115*, 2007. 6.1

[BCC+15] Dominic W Berry, Andrew M Childs, Richard Cleve, Robin Kothari, and Rolando D Somma. Simulating hamiltonian dynamics with a truncated taylor series. *Physical review letters*, 114(9):090502, 2015. 3.3.2, 3.7.1, 3.7.2, 2

[BCWdW01] Harry Buhrman, Richard Cleve, John Watrous, and Ronald de Wolf. Quantum fingerprinting. *Phys. Rev. Lett.*, 87:167902, Sep 2001. 2.1.1

[BDOT08] Sergey Bravyi, David P. DiVincenzo, Roberto Oliveira, and Barbara M. Terhal. The complexity of stoquastic local Hamiltonian problems. *Quantum Inf. Comput.*, 8(5):361–385, 2008. 3.4, 5.7

[Ber07] Dominic W Berry. Implementation of multipartite unitary operations with limited resources. *Physical Review A*, 75(3):032349, 2007. 3.4

[BG17] Sergey Bravyi and David Gosset. Polynomial-time classical simulation of quantum ferromagnets. *Physical Review Letters*, 119(10):100503, 2017. 4.4.2, 5, 5.1.1, 5.1.3, 5.7, 5.7.2

[BGKT19] Sergey Bravyi, David Gosset, Robert König, and Kristan Temme. Approximation algorithms for quantum many-body problems. *Journal of Mathematical Physics*, 60(3):032203, 2019. 6.1

[BGM21] Sergey Bravyi, David Gosset, and Ramis Movassagh. Classical algorithms for quantum mean values. *Nature Physics*, 17(3):337–341, 2021. 2.1, 7, 6.1

[BGP+20] Eyal Bairey, Chu Guo, Dario Poletti, Netanel H Lindner, and Itai Arad. Learning the dynamics of open quantum systems from their steady states. *New Journal of Physics*, 2020. 4.1

[BH16] Fernando GSL Brandao and Aram W Harrow. Product-state approximations to quantum states. *Communications in Mathematical Physics*, 342(1):47–80, 2016. 6.1

[BHV06] S. Bravyi, M. B. Hastings, and F. Verstraete. Lieb-robinson bounds and the generation of correlations and topological quantum order. *Phys. Rev. Lett.*, 97:050401, Jul 2006. 4.10.4

[BK16] Fernando GSL Brandão and Michael J Kastoryano. Finite correlation length implies efficient preparation of quantum thermal states. *Communications in Mathematical Physics*, pages 1–16, 2016. 4.4.2, 5.1.1, 127, 5.1.3, 3

[BKF22] Fernando GS L Brandão, Richard Kueng, and Daniel Stilck França. Faster quantum and classical SDP approximations for quadratic binary optimization. *Quantum*, 6:625, 2022. 4.1

[BKKT20] Sergey Bravyi, Alexander Kliesch, Robert Koenig, and Eugene Tang. Obstacles to variational quantum optimization from symmetry protection. *Phys. Rev. Lett.*, 125:260505, Dec 2020. 6.1

[BKL+19] Fernando G. S. L. Brandão, Amir Kalev, Tongyang Li, Cedric Yen-Yu Lin, Krysta M. Svore, and Xiaodi Wu. Quantum SDP Solvers: Large Speed-Ups, Optimality, and Applications to Quantum Learning. In *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*, volume 132 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 27:1–27:14. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019. 4.3.1, 4.4.2, 4.8.1

[BMBO20] Xavier Bonet-Monroig, Ryan Babbush, and Thomas E. O'Brien. Nearly optimal measurement scheduling for partial tomography of quantum states. *Phys. Rev. X*, 10:031064, Sep 2020. 4.8.2

[BMO+15] Boaz Barak, Ankur Moitra, Ryan O'Donnell, Prasad Raghavendra, Oded Regev, David Steurer, Luca Trevisan, Aravindan Vijayaraghavan, David Witmer, and John Wright. Beating the random assignment on constraint satisfaction problems of bounded degree. In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, volume 40 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages 110–123. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2015. 6.2.3, 6.2.3, 6.3.6

[Bra15] Sergey Bravyi. Monte Carlo simulation of stoquastic Hamiltonians. *Quantum Inf. Comput.*, 15(13-14):1122–1140, 2015. 5.1.3, 5.7, 5.7.2

[Bre15] Guy Bresler. Efficiently learning Ising models on arbitrary graphs. In *STOC'15—Proceedings of the 2015 ACM Symposium on Theory of Computing*, pages 771–782. ACM, New York, 2015. 4.1, 4.4.1

[BS17a] Alexander Barvinok and Pablo Soberón. Computing the partition function for graph homomorphisms. *Combinatorica*, 37(4):633–650, 2017. 5.1.1

[BS17b] Fernando GSL Brandao and Krysta M Svore. Quantum speed-ups for solving semidefinite programs. In *Foundations of Computer Science (FOCS), 2017 IEEE 58th Annual Symposium on*, pages 415–426. IEEE, 2017. 4.1

[BSK+17] Hannes Bernien, Sylvain Schwartz, Alexander Keesling, Harry Levine, Ahmed Omran, Hannes Pichler, Soonwon Choi, Alexander S Zibrov, Manuel Endres, Markus Greiner, et al. Probing many-body dynamics on a 51-atom quantum simulator. *Nature*, 551(7682):579–584, 2017. 4.1

[BV03] Sergey Bravyi and Mikhail Vyalyi. Commutative version of the k-local Hamiltonian problem and common eigenspace problem. *arXiv: quant-ph/0308021*, 2003. 4.4.2

[CC04] Pasquale Calabrese and John Cardy. Entanglement entropy and quantum field theory. *Journal of Statistical Mechanics: Theory and Experiment*, 2004(06):P06002, 2004. 3.4

[CC22] Nolan J Coble and Matthew Coudron. Quasi-polynomial time approximation of output probabilities of geometrically-local, shallow quantum circuits. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 598–609. IEEE, 2022. 2.1

[CGJV19] Andrea Coladangelo, Alex B. Grilo, Stacey Jeffery, and Thomas Vidick. Verifier-on-a-leash: new schemes for verifiable delegated quantum computation, with quasilinear resources. In *Advances in cryptology—EUROCRYPT 2019. Part III*, volume 11478 of *Lecture Notes in Comput. Sci.*, pages 247–277. Springer, Cham, 2019. 2.1

[CH19] Matthew Coudron and Aram W. Harrow. Universality of EPR pairs in entanglement-assisted communication complexity, and the communication cost of state conversion. In *34th Computational Complexity Conference*, volume 137 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 20, 25. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2019. 3.1.1, 3.1.2

[CHSH69] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23:880–884, Oct 1969. 2.1

[CHW07] Andrew M. Childs, Aram W. Harrow, and Pawel Wocjan. Weak Fourier-Schur sampling, the hidden subgroup problem, and the quantum collision problem. In *STACS 2007*, volume 4393 of *Lecture Notes in Comput. Sci.*, pages 598–609. Springer, Berlin, 2007. 2.3.2, 24

[CKS14] Anushya Chandran, Vedika Khemani, and S. L. Sondhi. How universal is the entanglement spectrum? *Phys. Rev. Lett.*, 113:060501, Aug 2014. 3.1.1

[CL68] C. Chow and C. Liu. Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory*, 14(3):462–467, 1968. 4.1

[CPF+10] Marcus Cramer, Martin B Plenio, Steven T Flammia, Rolando Somma, David Gross, Stephen D Bartlett, Olivier Landon-Cardinal, David Poulin, and Yi-Kai Liu. Efficient quantum state tomography. *Nature communications*, 1(1):1–7, 2010. 2.1, 2.1.2

[CPGW15] Toby S Cubitt, David Perez-Garcia, and Michael M Wolf. Undecidability of the spectral gap. *Nature*, 528(7581):207, 2015. 5.1.1

[CPSV11] J. Ignacio Cirac, Didier Poilblanc, Norbert Schuch, and Frank Verstraete. Entanglement spectrum and boundary theories with projected entangled-pair states. *Phys. Rev. B*, 83:245134, Jun 2011. 3.1.1, 3.4

[CS17] Anirban Narayan Chowdhury and Rolando D. Somma. Quantum algorithms for Gibbs sampling and hitting-time estimation. *Quantum Inf. Comput.*, 17(1-2):41–64, 2017. 5.1.3

[CSS18] Anirban Narayan Chowdhury, Yigit Subasi, and Rolando D Somma. Improved implementation of reflection operators. *arXiv:1803.02466*, 2018. 3.7.3

[CST+21] Andrew M Childs, Yuan Su, Minh C Tran, Nathan Wiebe, and Shuchen Zhu. Theory of trotter error with commutator scaling. *Physical Review X*, 11(1):011020, 2021. 3.4

[CW20] Jordan Cotler and Frank Wilczek. Quantum overlapping tomography. *Physical Review Letters*, 124(10):100401, 2020. 4.8.2

[Din07] Irit Dinur. The PCP theorem by gap amplification. *J. ACM*, 54(3):Art. 12, 44, 2007. 6.1

[Dob96] Roland Dobrushin. Estimates of semi-invariants for the ising model at low temperatures. *Translations of the American Mathematical Society-Series 2*, 177:59–82, 1996. 5.4

[DS85] Roland L Dobrushin and Senya B Shlosman. Completely analytical gibbs fields. In *Statistical physics and dynamical systems*, pages 371–403. Springer, 1985. 1.8.2

[DS87] Roland L Dobrushin and Senya B Shlosman. Completely analytical interactions: constructive description. *Journal of Statistical Physics*, 46(5-6):983–1014, 1987. 5, 5.1.1, 5.1.1, 5.1.2, 5.1.2, 5.1.3, 5.5, 145, 5.5, 5.5, 5.6

[DSVW04] Martin Dyer, Alistair Sinclair, Eric Vigoda, and Dror Weitz. Mixing in time and space for lattice spin systems: a combinatorial view. *Random Structures Algorithms*, 24(4):461–479, 2004. 5.1.1, 5.1.3

[DVZ18] Marcello Dalmonte, Benoît Vermersch, and Peter Zoller. Quantum simulation and spectroscopy of entanglement Hamiltonians. *Nature Physics*, 14(8):827–831, 2018. 3.1.1, 3.4

[EHF19] Tim J Evans, Robin Harper, and Steven T Flammia. Scalable bayesian Hamiltonian learning. *arXiv:1912.07636*, 2019. 4.1, 4.4.1

[EM17] Lior Eldar and Saeed Mehraban. Approximating the permanent of a random matrix with vanishing mean. *arXiv preprint arXiv:1711.09457*, 2017. 7

[EY36] Carl Eckart and Gale Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936. 18

[FGG14] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm. *arXiv preprint arXiv:1411.4028*, 2014. 6.1, 6.2.1

[FGG15] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm applied to a bounded occurrence constraint problem, 2015. 6.2.3

[FGG20] Edward Farhi, David Gamarnik, and Sam Gutmann. The quantum approximate optimization algorithm needs to see the whole graph: A typical case, 2020. 6.1

[FH14] Michael H. Freedman and Matthew B. Hastings. Quantum systems on non-k-hyperfinite complexes: A generalization of classical statistical mechanics on expander graphs. *Quantum Info. Comput.*, 14(1–2):144–180, January 2014. 6.2.2

[Fis65] ME Fisher. The nature of critical points. *Lect. Theor. Phys.*, 7:1–159, 1965. 5.1.1, 5.1.2

[FU15] Jürg Fröhlich and Daniel Ueltschi. *Some properties of correlations of quantum lattice systems in thermal equilibrium. Journal of Mathematical Physics*, 56(5):053302, 2015. 4.3.3

[Gam03] Theodore Gamelin. *Complex analysis*. Springer Science & Business Media, 2003. 5.2.4

[Gao15] Jingliang Gao. Quantum union bounds for sequential projective measurements. *Phys. Rev. A*, 92:052331, Nov 2015. 3.8.1

[GK12] Sevag Gharibian and Julia Kempe. Approximation algorithms for qma-complete problems. *SIAM Journal on Computing*, 41(4):1028–1050, 2012. 6.1

[GP19] Sevag Gharibian and Ojas Parekh. Almost optimal classical approximation algorithms for a quantum generalization of Max-Cut. In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, volume 145 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 31, 17. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2019. 6.1, 6.2.1

[Gre69] William Greenberg. Correlation functionals of infinite volume quantum spin systems. *Communications in Mathematical Physics*, 11(4):314–320, 1969. 5.4

[Gro79] Leonard Gross. *Decay of correlations in classical lattice models at high temperature. Communications in Mathematical Physics*, 68(1):9–27, Feb 1979. 4.3.3

[Har05] Aram Harrow. *Applications of coherent classical communication and the Schur transform to quantum information theory*. PhD thesis, Massachusetts Institute of Technology, 2005. 13, 2.3.2

[Has00] Johan Hastad. On bounded occurrence constraint satisfaction. *Inform. Process. Lett.*, 74(1-2):1–6, 2000. 6.2.3

[Has01] Johan Hastad. Some optimal inapproximability results. *J. ACM*, 48(4):798–859, 2001. 6.2.3

[Has04] M B Hastings. Lieb-Schultz-Mattis in Higher Dimensions. *Phys. Rev. B*, 69(104431), 2004. 1.7, 3.4, 3.4

[Has06] Matthew B Hastings. Solving gapped hamiltonians locally. *Physical Review B*, 73(8):085115, 2006. 5.1.2, 5.1.3, 5.4, 143

[Has07a] Matthew B Hastings. An area law for one-dimensional quantum systems. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(08):P08024, 2007. 2.1, 3.1.1

[Has07b] Matthew B Hastings. Quantum belief propagation: An algorithm for thermal quantum systems. *Physical Review B*, 76(20):201102, 2007. 1.7, 4.6.3, 84, 5.1.2, 5.2.3

[Has19] Matthew B Hastings. Classical and quantum bounded depth approximation algorithms. *arXiv preprint arXiv:1905.07047*, 2019. 6.2.3, 6.2.3, 6.3.6

[HC71] J. M. Hammersley and P. E. Clifford. Markov field on finite graphs and lattices. 1971. 4.3.3

[HGG09] Jonathan Huang, Carlos Guestrin, and Leonidas Guibas. Fourier theoretic probabilistic inference over permutations. *The Journal of Machine Learning Research*, 10:997–1070, 2009. 2.5

[HK06] Matthew B. Hastings and Tohru Koma. Spectral gap and exponential decay of correlations. *Commun. Math. Phys.*, 265(3):781–804, Aug 2006. 3.4

[HKK08] Masahito Hayashi, Akinori Kawachi, and Hirotada Kobayashi. Quantum measurements for hidden subgroup problems with optimal sample complexity. *Quantum Information and Computation*, 8:0345–0358, 2008. 4.7

[HKP20] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, Oct 2020. 4.8.2

[HKT21] Jeongwan Haah, Robin Kothari, and Ewin Tang. Optimal learning of quantum hamiltonians from high-temperature gibbs states. *arXiv preprint arXiv:2108.04842*, 2021. 4.4.2

[HL11] A. W. Harrow and D. W. Leung. A communication-efficient nonlocal measurement with application to communication complexity and bipartite gate capacities. *IEEE Trans. Inf. Theory*, 57(8):5504–5508, 2011. 3.1.2

[HLM17] Aram W. Harrow, Cedric Yen-Yu Lin, and Ashley Montanaro. Sequential measurements, disturbance and property testing. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1598–1611. SIAM, Philadelphia, PA, 2017. 2.1

[HLP20] Sean Hallgren, Eunou Lee, and Ojas Parekh. An approximation algorithm for the max-2-local hamiltonian problem. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2020)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020. 6.1

[HM13] Aram W. Harrow and Ashley Montanaro. Testing product states, quantum merlin-arthur games and tensor optimization. *J. ACM*, 60(1):3:1–3:43, February 2013. 1.8.1, 2, 2.1, 2.1.1, 2.1.1, 2-1a, 2.1.1, 7, 2.1.1, 2.1.1, 2.1.1, 2-2, 2.1.1, 2.1.1, 2.2.1, 2.2.2

[HM17] Aram W Harrow and Ashley Montanaro. Extremal eigenvalues of local hamiltonians. *Quantum*, 1:6, 2017. 6.1, 6.2.3, 6.3.6

[HMS20] Aram W. Harrow, Saeed Mehraban, and Mehdi Soleimanifar. Classical algorithms, correlation decay, and complex zeros of partition functions of quantum many-body systems. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 378–386, 2020. 1.8.2, 4.4.2, 5

[HS+86] Geoffrey E Hinton, Terrence J Sejnowski, et al. Learning and relearning in Boltzmann machines. *Parallel distributed processing: Explorations in the microstructure of cognition*, 1(282-317):2, 1986. 4.1

[HW03] Patrick Hayden and Andreas Winter. Communication cost of entanglement transformations. *Physical Review A*, 67:012326, Jan 2003. 3.1.2, 3.1.2, 3.1.2

[HW12] A. W. Harrow and A. Winter. How many copies are needed for state discrimination? *IEEE Transactions on Information Theory*, 58(1):1–2, 2012. 4.7

[HW20] Aram W Harrow and Annie Y Wei. Adaptive quantum simulated annealing for bayesian inference and estimating partition functions. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 193–212. SIAM, 2020. 4.1

[Jay57a] Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957. 4.3.1

[Jay57b] Edwin T Jaynes. Information theory and statistical mechanics. ii. *Physical Review*, 108(2):171, 1957. 4.3.1, 4.3.1, 4.8.1

[Jay82] Edwin T Jaynes. On the rationale of maximum-entropy methods. *Proceedings of the IEEE*, 70(9):939–952, 1982. 4.3.1

[JNV+20] Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen. MIP*=RE. *arXiv preprint arXiv:2001.04383*, 2020. 2.1

[JS93] Mark Jerrum and Alistair Sinclair. Polynomial-time approximation algorithms for the Ising model. *SIAM J. Comput.*, 22(5):1087–1116, 1993. 5.1.3

[KBa16] Michael J. Kastoryano and Fernando G. S. L. Brandão. Quantum Gibbs samplers: the commuting case. *Comm. Math. Phys.*, 344(3):915–957, 2016. 5.1.3

[KBa19a] Kohtaro Kato and Fernando G. S. L. Brandão. Locality of edge states and entanglement spectrum from strong subadditivity. *Phys. Rev. B*, 99:195124, May 2019. 3.1.1, 3.4

[KBa19b] Kohtaro Kato and Fernando G. S. L. Brandão. Quantum approximate Markov chains are thermal. *Comm. Math. Phys.*, 370(1):117–149, 2019. 4.3.4, 4.6.3, 5.1.3, 5.2.3

[KGK+14] M Kliesch, C Gogolin, MJ Kastoryano, A Riera, and J Eisert. Locality of temperature. *Physical Review X*, 4(3):031019, 2014. 4.3.3, 5.1.2, 5.1.3, 5.4, 142, 143, 3

[Kim12] Isaac H. Kim. Perturbative analysis of topological entanglement entropy from conditional independence. *Phys. Rev. B*, 86:245116, Dec 2012. 1.7

[Kim17] Isaac H Kim. Markovian matrix product density operators: Efficient computation of global entropy. *arXiv preprint arXiv:1709.07828*, 2017. 4.6.3, 5.1.3

[KKBa20] Tomotaka Kuwahara, Kohtaro Kato, and Fernando G. S. L. Brandão. Clustering of conditional mutual information for quantum gibbs states above a threshold temperature. *Physical Review Letters*, 124:220601, 2020. 4.3.4, 4.4.2, 5.1.3, 2

[KM17] Adam R. Klivans and Raghu Meka. Learning graphical models using multiplicative weights. In *58th Annual IEEE Symposium on Foundations of Computer Science—FOCS*, pages 343–354. 2017. 4.1, 4.4.1, 4.4.2

[KMS16] Tomotaka Kuwahara, Takashi Mori, and Keiji Saito. Floquet-magnus theory and generic transient dynamics in periodically driven many-body quantum systems. *Annals of Physics*, 367:96 – 124, 2016. 4.10.3

[KMT+17] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M Chow, and Jay M Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*, 549(7671):242–246, 2017. 6.1

[KP86] Roman Kotecký and David Preiss. Cluster expansion for abstract polymer models. *Communications in Mathematical Physics*, 103(3):491–498, 1986. 5.4

[Kro19] Hari Krovi. An efficient high dimensional quantum Schur transform. *Quantum*, 3:122, February 2019. 13

[KS18] Tomotaka Kuwahara and Keiji Saito. Polynomial-time classical simulation for one-dimensional quantum gibbs states. *arXiv preprint arXiv:1807.08424*, 2018. 5.1.3

[KS20a] Tomotaka Kuwahara and Keiji Saito. Area law of noncritical ground states in 1d long-range interacting systems. *Nature communications*, 11(1):1–7, 2020. 3.3.4, 3.8.2, 62

[KS20b] Tomotaka Kuwahara and Keiji Saito. Gaussian concentration bound and ensemble equivalence in generic quantum many-body systems including long-range interactions. *Annals of Physics*, 421:168278, 2020. 3.4

[Kuw16] Tomotaka Kuwahara. Asymptotic behavior of macroscopic observables in generic spin systems. *Journal of Statistical Mechanics: Theory and Experiment*, 2016(5):053103, 2016. 4.10.3

[LH08] Hui Li and F. D. M. Haldane. Entanglement spectrum as a generalization of entanglement entropy: Identification of topological order in non-abelian fractional quantum hall effect states. *Phys. Rev. Lett.*, 101:010504, Jul 2008. 3.1.1, 3.4

[Llo96] Seth Lloyd. Universal quantum simulators. *Science*, pages 1073–1078, 1996. 6.1

[LMH+17] BP Lanyon, C Maier, Milan Holzäpfel, Tillmann Baumgratz, C Hempel, P Jurcevic, Ish Dhand, AS Buyskikh, AJ Daley, Marcus Cramer, et al. Efficient tomography of a quantum many-body system. *Nature Physics*, 13(12):1158–1162, 2017. 2.1.2

[LP99] Hoi-Kwong Lo and Sandu Popescu. Classical communication cost of entanglement manipulation: Is entanglement an interconvertible resource? *Physical Review Letters*, 83(7):1459, 1999. 3.4

[LP08] Matthew S Leifer and David Poulin. Quantum graphical models and belief propagation. *Annals of Physics*, 323(8):1899–1946, 2008. 4.1, 4.3.4

[LR72] Elliott H. Lieb and Derek W. Robinson. The finite group velocity of quantum spin systems. *Communications in Mathematical Physics*, 28(3):251–257, 1972. 1.7, 81

[LSS19a] Jingcheng Liu, Alistair Sinclair, and Piyush Srivastava. Fisher zeros and correlation decay in the Ising model. In *10th Innovations in Theoretical Computer Science*, volume 124 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 55, 8. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2019. 5.1.1, 5.1.1, 5.1.3

[LSS19b] Jingcheng Liu, Alistair Sinclair, and Piyush Srivastava. The Ising partition function: zeros and deterministic approximation. *J. Stat. Phys.*, 174(2):287–315, 2019. 5.1.3, 5

[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. 2.1

[LW05] Michael A Levin and Xiao-Gang Wen. String-net condensation: A physical mechanism for topological phases. *Physical Review B*, 71(4):045110, 2005. 5.1.1

[LW18] Guang Hao Low and Nathan Wiebe. Hamiltonian simulation in the interaction picture. *arXiv:1805.00675*, 2018. 3.3.2, 3.3.2, 3.7.1, 53, 3.7.2

[LY52] Tsung-Dao Lee and Chen-Ning Yang. Statistical theory of equations of state and phase transitions. ii. lattice gas and ising model. *Physical Review*, 87(3):410, 1952. 5.1.1, 5.1.2, 5.7

[Mah18] Urmila Mahadev. Classical verification of quantum computations. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 259–267, 2018. 2.1

[MB19] Ryan L Mann and Michael J Bremner. Approximation algorithms for complex-valued ising models on bounded degree graphs. *Quantum*, 3:162, 2019. 7

[MBS+18] Jarrod R McClean, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature communications*, 9(1):1–6, 2018. 6.1

[MdW13] Ashley Montanaro and Ronald de Wolf. A survey of quantum property testing. *arXiv preprint arXiv:1310.2035*, 2013. 2, 2.1, 2.1, 2.1.1

[MKB05] Florian Mintert, Marek Kuś, and Andreas Buchleitner. Concurrence of mixed multipartite quantum states. *Physical review letters*, 95(26):260502, 2005. 2.1.1

[Mon15a] Andrea Montanari. Computational implications of reducing data to sufficient statistics. *Electron. J. Stat.*, 9(2):2370–2390, 2015. 4.3.3, 4.4.2

[Mon15b] Ashley Montanaro. Quantum speedup of monte carlo methods. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2181):20150301, 2015. 4.1

[MST+20] Mario Motta, Chong Sun, Adrian TK Tan, Matthew J O'Rourke, Erika Ye, Austin J Minnich, Fernando GSL Brandão, and Garnet Kin-Lic Chan. Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution. *Nature Physics*, 16(2):205–210, 2020. 4.1

[MSVC15] Andras Molnar, Norbert Schuch, Frank Verstraete, and J Ignacio Cirac. Approximating gibbs states of local Hamiltonians efficiently with projected entangled pair states. *Physical Review B*, 91(4):045138, 2015. 5.1.3

[NLPD+19] John Napp, Rolando L La Placa, Alexander M Dalzell, Fernando GSL Brandao, and Aram W Harrow. Efficient classical simulation of random shallow 2d quantum circuits. *arXiv preprint arXiv:2001.00021*, 2019. 2.1

[NS06] Bruno Nachtergaele and Robert Sims. Lieb-robinson bounds and the exponential clustering theorem. *Commun. Math. Phys.*, 265(1):119–130, Jul 2006. 3.4, 3.4

[NS09] Bruno Nachtergaele and Robert Sims. Locality estimates for quantum spin systems. In Vladas Sidoravičius, editor, *New Trends in Mathematical Physics*, pages 591–614, Dordrecht, 2009. Springer Netherlands. 1.7, 81

[O'D14] Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014. 6.3.6, 6.3.6

[OW15] Ryan O'Donnell and John Wright. Quantum spectrum testing. In *STOC'15—Proceedings of the 2015 ACM Symposium on Theory of Computing*, pages 529–538. ACM, New York, 2015. 2.1, 2.1.2, 13, 2.4, 2.4, 27, 2.4

[Par82] Yong Moon Park. The cluster expansion for classical and quantum lattice systems. *Journal of Statistical Physics*, 27(3):553–576, 1982. 5.4

[PGPH20] David Pérez-García and Antonio Pérez-Hernández. Locality estimes for complex time evolution in 1d. *arXiv preprint arXiv:2004.10516*, 2020. 3.4

[PGVWC08] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac. Peps as unique ground states of local Hamiltonians. *Quantum Info. Comput.*, 8(6):650–663, July 2008. 3.4

[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*, 5:4213, 2014. 6.1

[PR18] Han Peters and Guus Regts. Location of zeros for the partition function of the ising model on bounded degree graphs. *arXiv preprint arXiv:1810.01699*, 2018. 5.1.3, 5

[PT20] Ojas Parekh and Kevin Thompson. Beating random assignment for approximating quantum 2-local hamiltonian problems, 2020. 6.1

[PTBO10] Frank Pollmann, Ari M. Turner, Erez Berg, and Masaki Oshikawa. Entanglement spectrum of a topological phase in one dimension. *Phys. Rev. B*, 81:064439, Feb 2010. 3.1.1

[PW09] David Poulin and Pawel Wocjan. Sampling from the thermal quantum gibbs state and evaluating partition functions with a quantum computer. *Physical review letters*, 103(22):220502, 2009. 5.1.3

[PY95] Yong Moon Park and Hyun Jae Yoo. *Uniqueness and clustering properties of Gibbs states for classical and quantum unbounded spin systems. Journal of Statistical Physics*, 80(1):223–271, Jul 1995. 4.3.3

[QR19] Xiao-Liang Qi and Daniel Ranard. Determining a local Hamiltonian from a single eigenstate. *Quantum*, 3:159, 2019. 4.1, 4.4.1

[RUV13] Ben W Reichardt, Falk Unger, and Umesh Vazirani. Classical command of quantum systems. *Nature*, 496(7446):456–460, 2013. 2.1

[SBM+11] Jonathan Simon, Waseem S Bakr, Ruichao Ma, M Eric Tai, Philipp M Preiss, and Markus Greiner. Quantum simulation of antiferromagnetic spin chains in an optical lattice. *Nature*, 472(7343):307–312, 2011. 4.1

[SF71] Masuo Suzuki and Michael E Fisher. Zeros of the partition function for the heisenberg, ferroelectric, and general ising models. *Journal of Mathematical Physics*, 12(2):235–246, 1971. 5.1.2, 5.7, 161, 163, 5.7.1, 165

[SK14] Brian Swingle and Isaac H. Kim. Reconstructing quantum states from local data. *Phys. Rev. Lett.*, 113:260501, Dec 2014. 4.1, 4.3.1

[Sly10] Allan Sly. Computational transition at the uniqueness threshold. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science—FOCS 2010*, pages 287–296. IEEE Computer Soc., Los Alamitos, CA, 2010. 5.1

[SML+11] Alireza Shabani, Masoud Mohseni, Seth Lloyd, Robert L Kosut, and Herschel Rabitz. Estimation of many-body quantum Hamiltonians via compressive sensing. *Physical Review A*, 84(1):012107, 2011. 4.1

[SPCPG13] Norbert Schuch, Didier Poilblanc, J. Ignacio Cirac, and David Pérez-García. Topological order in the projected entangled-pair states formalism: Transfer operator and boundary Hamiltonians. *Phys. Rev. Lett.*, 111:090501, Aug 2013. 3.1.1, 3.4

[SS12] Allan Sly and Nike Sun. The computational hardness of counting in two-spin models on d-regular graphs. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science—FOCS 2012*, pages 361–369. IEEE Computer Soc., Los Alamitos, CA, 2012. 5.1, 5.1.2

[SSR+14] C Senko, J Smith, P Richerme, A Lee, WC Campbell, and C Monroe. Coherent imaging spectroscopy of a quantum many-body spin system. *Science*, 345(6195):430–433, 2014. 4.1

[SST14] Alistair Sinclair, Piyush Srivastava, and Marc Thurley. Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs. *J. Stat. Phys.*, 155(4):666–686, 2014. 5.1, 5.1.3

[Suz69] Masuo Suzuki. On the distribution of zeros for the Heisenberg model. *Progress of Theoretical Physics*, 41(6):1438–1449, 1969. 166

[SW22] Mehdi Soleimanifar and John Wright. Testing matrix product states. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1679–1701, 2022. 1.8.1, 2

[Tou15] Dave Touchette. Quantum information complexity. In *Proceedings of the 47th Annual ACM on Symposium on Theory of Computing*, STOC '15, pages 317–326, New York, NY, USA, 2015. ACM. 3.4

[Uel04] Daniel Ueltschi. *Cluster expansions and correlation functions. Moscow Mathematical Journal*, 4(2):511–522, 2004. 4.3.3

[vAGGdW20] J. van Apeldoorn, A. Gilyén, S. Gribling, and R. de Wolf. Quantum SDP-solvers: Better upper and lower bounds. *Quantum*, 4:230, 2020. 4.1

[Vaz13] Vijay V Vazirani. *Approximation algorithms*. Springer Science & Business Media, 2013. 6.1

[VC06] F. Verstraete and J. I. Cirac. Matrix product states represent ground states faithfully. *Phys. Rev. B*, 73:094423, Mar 2006. 19

[Vid03] Guifré Vidal. Efficient classical simulation of slightly entangled quantum computations. *Phys. Rev. Lett.*, 91:147902, Oct 2003. 2.1

[VMC08] F. Verstraete, V. Murg, and J.I. 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. 3.4

[VMLC16] Marc Vuffray, Sidhant Misra, Andrey Lokhov, and Michael Chertkov. Interaction screening: Efficient and sample-optimal learning of ising models. In *Advances in Neural Information Processing Systems*, pages 2595–2603, 2016. 4.1, 4.3.3, 4.4.1, 4.4.2

[VMN+19] Guillaume Verdon, Jacob Marks, Sasha Nanda, Stefan Leichenauer, and Jack Hidary. Quantum Hamiltonian-based models and the variational quantum thermalizer algorithm. *arXiv:1910.02071*, 2019. 4.4.1

[VV19] Umesh Vazirani and Thomas Vidick. Fully device independent quantum key distribution. *Commun. ACM*, 62(4):133, March 2019. 2.1

[Wei04] Dror Weitz. *Mixing in time and space for discrete spin systems*. ProQuest LLC, Ann Arbor, MI, 2004. Thesis (Ph.D.)–University of California, Berkeley. 5.1.1, 5.1.3

[Wei06] Dror Weitz. Counting independent sets up to the tree threshold. In *STOC'06: Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, pages 140–149. ACM, New York, 2006. 5.1, 5.1.1, 5.1.3

[WGFC14a] Nathan Wiebe, Christopher Granade, Christopher Ferrie, and David Cory. Quantum Hamiltonian learning using imperfect quantum resources. *Physical Review A*, 89(4):042314, 2014. 4.1, 4.4.1

[WGFC14b] Nathan Wiebe, Christopher Granade, Christopher Ferrie, and David G Cory. Hamiltonian learning and certification using quantum resources. *Physical Review Letters*, 112(19):190501, 2014. 4.1, 4.4.1

[WJ08] Martin J. Wainwright and Michael I. Jordan. Graphical models, exponential families, and variational inference. *Found. Trends Mach. Learn.*, 1(1-2):1–305, 2008. 4.1

[WKS16] Nathan Wiebe, Ashish Kapoor, and Krysta M. Svore. Quantum deep learning. *Quantum Info. Comput.*, 16(7–8):541–587, May 2016. 4.1

[WPS+17] Jianwei Wang, Stefano Paesani, Raffaele Santagati, Sebastian Knauer, Antonio A Gentile, Nathan Wiebe, Maurangelo Petruzzella, John G O'Brien, Jeremy L. Rarity, Anthony Laing, et al. Experimental quantum hamiltonian learning. *Nature Physics*, 13(6):551–555, 2017. 4.1

[Wri16] John Wright. *How to learn a quantum state*. PhD thesis, Carnegie Mellon University, 2016. 2.3.2

[Yu21] Nengkun Yu. Sample efficient identity testing and independence testing of quantum states. In *12th Innovations in Theoretical Computer Science Conference*, volume 185 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. 11, 20. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2021. 2.1

[ZPH+17] Jiehang Zhang, Guido Pagano, Paul W Hess, Antonis Kyprianidis, Patrick Becker, Harvey Kaplan, Alexey V Gorshkov, Z-X Gong, and Christopher Monroe. Observation of a many-body dynamical phase transition with a 53-qubit quantum simulator. *Nature*, 551(7682):601–604, 2017. 4.1

[ZSW20] Yiqing Zhou, E. Miles Stoudenmire, and Xavier Waintal. What limits the simulation of quantum computers? *Phys. Rev. X*, 10:041038, Nov 2020. 2.1