

Quantum many-body systems in thermal equilibrium

Álvaro M. Alhambra^{1,*}

¹*Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Straße 1, D-85748 Garching, Germany*

The thermal or equilibrium ensemble is one of the most ubiquitous states of matter. For models comprised of many locally interacting quantum particles, it describes a wide range of physical situations, relevant to condensed matter physics, high energy physics, quantum chemistry and quantum computing, among others. We give a pedagogical overview of some of the most important universal features about the physics and complexity of these states, which have the locality of the Hamiltonian at its core. We focus on mathematically rigorous statements, many of them inspired by ideas and tools from quantum information theory. These include bounds on their correlations, the form of the subsystems, various statistical properties, and the performance of classical and quantum algorithms. We also include a summary of a few of the most important technical tools, as well as some self-contained proofs.

Parts of these notes were the basis for a lecture series within the “Quantum Thermodynamics Summer School 2021” during August 2021 in Les Diablerets, Switzerland.

CONTENTS

I. Introduction	2
A. Scope and content	4
II. Mathematical preliminaries and notation	5
A. Operator norms	5
B. Information-theoretic quantities	6
C. Lattice notation	6
D. Asymptotic notation	7
III. An overview of technical tools	7
A. Connected cluster expansion	8
B. Thermal locality estimates	10
C. Quantum belief propagation	12
D. Selected trace inequalities	13
IV. Correlations	14
A. Correlations between neighbouring regions: Thermal area law	15
B. Decay of long-range correlations	16
C. A refined correlation decay: Conditional mutual information	18
V. Locality of temperature	19
A. Local indistinguishability	20

* alvaro.alhambra@mpq.mpg.de

1. Proof in 1D	21
B. Hamiltonian of mean force	23
VI. Statistical properties	23
A. Jaynes' maximum entropy principle	24
B. Measurement statistics and concentration bounds	25
C. Equivalence of ensembles	26
VII. Algorithms and complexity of thermal states	29
A. An efficient classical algorithm for the 1D partition function	29
B. Hardness of approximating partition functions	30
C. Tensor network methods	31
D. Quantum algorithms for preparing thermal states	33
General considerations	33
Algorithms based on purifications	34
Quantum Metropolis Sampling	34
Variational algorithms	35
Efficient algorithms from physical features	36
VIII. Commuting Hamiltonians	36
IX. Conclusions and open questions	38
Acknowledgements	40
A. Miscellaneous proofs	40
Locality of operator E_A	40
Proof of Eq. (53)	41
Proof of Eq. (109)	41
References	42

I. INTRODUCTION

We are currently at the dawn of the age of synthetic quantum matter. Increasingly better experiments on a variety of quantum platforms are improving in size and controllability at unprecedented rates, aided by the current impulse of quantum information science and technology. This gives very good prospects to the exploration of the physics of complex quantum many body systems. Our aspiration to better understand these systems is very well motivated from a scientific perspective, but also potentially from the industrial one: unlocking the potential of complex quantum systems may bring surprising advances to the engineering of new materials or chemical compounds in the future. It may also yield computational tools with unprecedented capabilities for a still unknown range of applications.

Many of the most commonly studied materials and current experimental platforms are described by an arrangement of quantum particles in some sort of lattice configuration. Due to the spatial decay of electromagnetic forces, each of these particles only interacts appreciably with their immediate vicinity, which causes the interactions between them to be local.

In these notes, we focus on the properties of these important systems when at thermal equilibrium, so that they are accurately described by the so-called thermal or Gibbs state. We review and explain some of their most important universal properties, covered from a mathematical perspective. That is, we focus on statements that can be proven about states of the Gibbs form

$$\rho_\beta = \frac{e^{-\beta H}}{Z} = \frac{1}{Z} \sum_l e^{-\beta E_l} |E_l\rangle\langle E_l|, \quad (1)$$

where $H = \sum_l E_l |E_l\rangle\langle E_l|$ is the Hamiltonian, β is the inverse temperature and $Z \equiv \text{Tr}[e^{-\beta H}]$ is the partition function. The Hamiltonian describes the interaction between the N particles, which will be restricted the interactions to be short-ranged or local. A “local Hamiltonian” is a Hermitian operator H in the finite-dimensional Hilbert space of N d -dimensional particles $(\mathbb{C}^d)^{\otimes N}$. It is defined as a sum of terms

$$H = \sum_i h_i \otimes \mathbb{I}, \quad (2)$$

each of which has support (i.e. acts non-trivially) on at most k particles, and bounded strength, such that $\max_i \|h_i\| = h$ (for a definition of the operator norm $\|\cdot\|$ see Sec. II A below). In what follows we just write the terms as h_i for simplicity. These constitute the individual interactions, which are typically arranged in a lattice of a small dimension. A simple example is e.g. the transverse-field Ising model in one dimension with open boundary conditions

$$H_{\text{Ising}} = \sum_{j=1}^{N-1} (J\sigma_j^X \sigma_{j+1}^X + \Delta\sigma_j^Z) + \Delta\sigma_N^Z. \quad (3)$$

Here, $k = 2$ and the interactions are arranged on a 1D chain.

The idea of a local Hamiltonian is very general, and involves many different models describing a wide range of situations, of interest for many fields of physics, chemistry and computer science. The only thing they have in common is the *locality* of the interactions. We thus aim to understand mathematically how this fact alone constrains both the physics and the computational complexity, when combined with thermal fluctuations.

The thermal states of these general local Hamiltonians are interesting for a wide variety of reasons, some of which are:

- It is one of the most **ubiquitous** states of quantum matter: typical experiments happen at finite temperature, where the quantum system at hand is weakly coupled to some external radiation field that drives it to the thermal state.
- The thermal state is also important when studying not just systems with an external bath, but also in the evolution of **isolated** quantum systems, even when their global state is pure: in very generic cases, these end up being “their own bath”, and the individual subsystems thermalize to the Gibbs ensemble [1, 2].
- From a general condensed matter/material science standpoint, we are very interested in numerous questions about the physics at **finite temperature**: How are conserved quantities (e.g. charge, energy) propagated in a state close to equilibrium? How does the system respond to small or large perturbations away from equilibrium?
- Systems at thermal equilibrium (both quantum and classical) display interesting **phase transitions** in certain (low) temperature regimes (e.g. classical Ising model in 2D). It is thus relevant to study what are their universal properties both in and away from the critical points.

- They are also important from the point of view of **quantum phases of matter** and topological order: it has been widely established that thermal states of local models in dimension $D - 1$ appear in the entanglement spectrum of D - dimensional ground states [3]. As such, understanding their structure should also help us in elucidating the low energy correlations of many interesting systems.
- These states are also important for **computation**. For instance, being able to sample from the thermal distribution of local models is a typical subroutine for certain classical and quantum algorithms [4-7]. They are also a very naturally occurring data structure in both classical and quantum machine learning [8-12] (often under the name of *Boltzmann machines*).
- It is known from **quantum computational complexity** that the low energy subspace of local Hamiltonians is able to encode the solution to very hard computational problems: finding the ground state energy is **QMA complete** [13]. Thus, it is widely believed that even a quantum computer should not be able to do it in polynomial time. This then at least also applies to the thermal state at very low temperature, and motivates the study of how the complexity changes as the temperature rises [14, 15].

There are many different specific aspects that one could explore, but here we focus on the following, which we believe to be of particular importance:

- The correlations between the particles at different parts of the lattice, and how they are distributed depending on its geometry.
- The form of the subsystems that a thermal state can take, and how are they related to few-particle Gibbs states.
- The statistical physics properties of these systems at equilibrium, including Jaynes' principle, concentration bounds and equivalence of ensembles.
- The efficiency of classical and quantum algorithms for the generation and manipulation of thermal states, and the computation of expectation values and partition functions.

The general topic of these notes, and the particular results explained here, are a small part of the exciting past, present and future efforts to understand the physics and complexity of quantum many-body systems. We hope to contribute to the understanding and cross-fertilization of the many different angles that the quantum many-body problem can take. See also e.g. [1, 16] for previous references with partially overlapping content.

A. Scope and content

Throughout these notes, we cover statements that have a precise mathematical formulation, many of them motivated by a quantum information theoretic perspective. This notably includes a short exposition of a few key technical tools in Sec. III. These have not previously appeared together, but rather separately explained in the literature with various levels of detail, depending on the context and usage. We hope that this encourages new, potentially unexpected, applications thereof.

Along the sections with the actual physical and computational results, we write the proofs of some of the simpler or more important ones explained throughout. This includes at least one

main result per section, which should serve as a pedagogical example. For the rest, some of which have more detailed or involved derivations, we refer the reader to the original works cited along the text. One of our main hopes is that after reading these notes even the more technical works will be more easily accessible to a wider range of researchers. Because of this, rather than the traditional Theorem-Proof structure of most mathematical physics writing, we have chosen a more streamlined style for the presentation which allows for more physical explanations and intuitions of the steps. This will hopefully contribute to a wider readability.

There are no new results in these notes, and the proofs are either the same or slightly simplified versions of previous ones in the literature. The relevant references are included, but this does not mean that all of the previous relevant ones are listed here: we are certainly missing to mention a very large body of work. This includes many relevant papers on mathematical physics, but also a lot of important physics literature that covers these topics from perspectives that are beyond our scope: based on numerical methods, theory work on experimental implementations, as well as all experimental results.

II. MATHEMATICAL PRELIMINARIES AND NOTATION

A. Operator norms

A basic but very important mathematical tool in this context are the Schatten p -norms for operators, as well as the different inequalities between them. These norms are maps from the space of operators to \mathbb{R} , as $M \rightarrow \|M\|_p$, that obey the following properties:

- Homogeneous: If α is a scalar, $\|\alpha M\|_p = |\alpha| \|M\|_p$.
- Positive: $\|M\|_p \geq 0$.
- Definite: $\|M\|_p = 0 \iff M = 0$.
- Triangle inequality: $\|M + N\|_p \leq \|M\|_p + \|N\|_p$.

For a given operator M with singular values $\{\lambda_l^M\}$ and $p \in [1, \infty)$, they are defined as

$$\|M\|_p \equiv \text{Tr}[|M|^p]^{\frac{1}{p}} = \left(\sum_l (\lambda_l^M)^p \right)^{\frac{1}{p}}. \quad (4)$$

The more important ones are the operator norm $\|M\| \equiv \|M\|_\infty = \max_l |\lambda_l^M|$, the Hilbert-Schmidt 2-norm $\|M\|_2 = \text{Tr}[MM^\dagger]^{1/2}$ and the 1-norm or trace norm

$$\|M\|_1 = \max_{\|N\| \leq 1} \text{Tr}[MN]. \quad (5)$$

Thus $|\text{Tr}[M]| \leq \|M\|_1$, with equality for positive operators. For quantum states, $\text{Tr}[\rho] = \|\rho\|_1 = 1$.

Typically we measure the “strength” of an observable with the operator norm, and the closeness of two quantum states with the trace norm $\|\rho - \sigma\|_1$, since it is related to the probability of distinguishing them under measurements. The 2-norm, on the other hand, is often the easiest one to compute in practice. Also note the very important Hölder’s inequality

$$\|MN\|_p \leq \|M\|_{q_1} \|N\|_{q_2}, \quad (6)$$

which holds for $\frac{1}{p} = \frac{1}{q_1} + \frac{1}{q_2}$ (e.g. $p = q_1 = 1, q_2 = \infty$). A particularly useful corollary is the Cauchy-Schwarz inequality, when $q_1 = q_2 = 2$ and $p = 1$,

$$|\text{Tr}[M^\dagger N]|^2 \leq \text{Tr}[MM^\dagger] \text{Tr}[NN^\dagger]. \quad (7)$$

B. Information-theoretic quantities

Let us define the von Neumann entropy of a quantum state ρ [17]

$$S(\rho) = -\text{Tr}[\rho \log(\rho)], \quad (8)$$

which, roughly speaking, quantifies the uncertainty we have about the particular state. It is bounded by $0 \leq S(\rho) \leq \log d$. The lower bound is obtained by choosing ρ pure, and the upper bound by the identity $\rho = \mathbb{I}/d$. Another important quantity is the Umegaki relative entropy

$$D(\rho|\sigma) = \text{Tr}[\rho(\log \rho - \log \sigma)], \quad (9)$$

which is a measure of distinguishability of quantum states. It obeys Pinsker's inequality

$$D(\rho|\sigma) \geq \frac{1}{2} \|\rho - \sigma\|_1^2, \quad (10)$$

which relates the relative entropy with the 1-norm. It is strictly positive for $\rho \neq \sigma$, and vanishes otherwise. It is also closely related to the non-equilibrium free energy

$$D(\rho|\rho_\beta) = \beta \text{Tr}[\rho H] - S(\rho) + \log Z \equiv \beta F_\beta(\rho) + \log Z, \quad (11)$$

which also shows that the equilibrium free energy is $F_\beta(\rho_\beta) = -\beta^{-1} \log Z$.

From these quantities we can also define the quantum mutual information, which, given a bipartite state ρ^{AB} on subsystems A and B , with $\text{Tr}_B[\rho^{AB}] = \rho_A$, $\text{Tr}_A[\rho^{AB}] = \rho_B$, quantifies the correlations between A and B as

$$I(A : B)_\rho = S(\rho^A) + S(\rho^B) - S(\rho^{AB}) = D(\rho^{AB}|\rho^A \otimes \rho^B). \quad (12)$$

For instance, it is zero if and only if $\rho^{AB} = \rho^A \otimes \rho^B$. For all these three functions we can also define their corresponding Rényi generalizations. See [18, 19] for details.

A further, perhaps more refined quantity is the *conditional mutual information* (CMI), defined as

$$\begin{aligned} I(A : C|B)_\rho &= S(\rho^{AB}) + S(\rho^{BC}) - S(\rho^{ABC}) - S(\rho^B) \\ &= I(A : BC)_\rho - I(A : B)_\rho. \end{aligned} \quad (13)$$

This perhaps less known quantity is behind many non-trivial statements in quantum information theory (see Section 11.7 in [20] for more details). In a nutshell, it measures how much A and C share correlations that are *not* mediated by B . That is, if this quantity is small, most of the correlations between A and C (which may be weak) are in reality correlations between A and B and B and C .

C. Lattice notation

In what follows we need some technical definitions regarding the properties of the Hamiltonian and the lattice. The lattice is a hypergraph which we denote by $\Lambda = \{V, E\}$ with vertex set V and hyperedges E . To each vertex we associate a Hilbert space of dimension d , \mathbb{C}^d . The number of particles is $N = |V|$, and the number of hyperedges is $|E|$. We can separate the vertices into subregions, such as Λ_A , and we denote with $\partial_A \in \Lambda_A$ the sites at the boundary of that region (that is, with at least one hyperedge connecting to $\setminus \Lambda_A$), of which there are $|\partial_A|$. For simplicity, we often refer to regions as A, B, \dots instead of $\Lambda_A, \Lambda_B, \dots$. We also need the notion of “distance”

between two regions, $\text{dist}(A, B)$, defined as the smallest number of overlapping hyperedges that connect a vertex of A with a vertex of B .

To define the Hamiltonian, we associate local interactions to hyperedges, such that $H = \sum_{i \in E} h_i$. For an operator h_i , the set of vertices on which it has non-trivial support is $\text{supp}(h_i)$. We have already specified that each h_i is such that $|\text{supp}(h_i)| \leq k$ (that is, the hyperedges have size at most k), so that for constant k , $N \propto |E|$. We also note that $\|h_i\| \leq h$ and introduce the following quantity

$$J = \max_{x \in V} \sum_{i: x \in \text{supp}(h_i)} \|h_i\|, \quad (15)$$

that is, J upper bounds the norm of the interactions that act on an individual vertex.

D. Asymptotic notation

The so-called asymptotic or Bachmann–Landau notation succinctly describes the asymptotic behaviour of a function when the argument grows large. It is typically used when in a particular expression there are constant factors that we are happy to omit, that are unnecessarily cumbersome, or when we only have partial knowledge of the asymptotic behaviour. We say that, given a function $f(N) \geq 0$:

- $f(N) = \mathcal{O}(g(N))$ if there are constants $M, N_0 > 0$ such that $\forall N > N_0, f(N) \leq Mg(N)$.
- $f(N) = \tilde{\mathcal{O}}(g(N))$ is similar to $\mathcal{O}(g(N))$ but with possible additional poly-logarithmic factors, so that instead $\forall N > N_0, f(N) \leq Mg(N)\text{polylog}(g(N))$.
- $f(N) = o(g(N))$ if for every $\varepsilon > 0$ there exists a $N_0 > 0$ such that $\forall N > N_0, f(N) \leq \varepsilon g(N)$.
- $f(N) = \Omega(g(N))$ if there are constants $M, N_0 > 0$ such that $\forall N > N_0, f(N) \geq Mg(N)$.

These are the most commonly used symbols of this notation, all of which appear below.

III. AN OVERVIEW OF TECHNICAL TOOLS

When studying quantum thermal states from a mathematical point of view, what we often need is some way of simplifying the operator $e^{-\beta H}$, in a way that makes the particular problem at hand mathematically tractable. This is usually achieved by expressing the relevant function of $e^{-\beta H}$ in simpler terms. Potential issues that complicate this are:

1. The exponential of a local operator is not a local operator, due to the high order terms in the expansion, and could in principle be arbitrarily complicated.
2. The individual elements in the Hamiltonian Eq. (2) do not commute with each other. Thus we cannot divide the exponential of the Hamiltonian into smaller pieces by iterating simple identities like $e^{-\beta(H_1+H_2)} \stackrel{?}{=} e^{-\beta H_1} e^{-\beta H_2}$.

The locality of the Hamiltonian helps make these two problems often not as serious as they could be in general situations. There is a number of tools to deal with this, and we now describe some of the most relevant ones. Below, we explain how the cluster expansion in Sec. III A helps with issue 1, while there are at least two different techniques in Sec. III B and III C that help us with issue 2.

A. Connected cluster expansion

This is a powerful set of ideas whose origins can be traced back to a wide set of the classic (and classical) literature on mathematical physics and statistical mechanics (see e.g. [21]). It has traditionally been used to prove the analyticity of the partition function at high temperatures, as well as the existence of computationally efficient approximation schemes to it. To illustrate the main ideas, we outline the main steps needed to prove those results, following e.g. [22] and [23]. These works are recent, but this is a decades-old and very studied tool. See [22] in particular for the reference to several important classic works.

Let us start with the logarithm of the partition function in any dimension $\log Z \equiv \log \text{Tr}[e^{-\beta H}]$, and consider its Taylor expansion around $\beta = 0$

$$\log Z = \sum_m \frac{\beta^m}{m!} K_m. \quad (16)$$

One can then ask: what is the radius of convergence of this Taylor series? More precisely, we would like to know whether there is some β^* such that for $0 \leq \beta < \beta^*$ we have that:

- The function $\log Z$ is analytic.
- The m -th derivative at $\beta = 0$ is such that

$$\left| \frac{d^m \log Z}{d\beta^m} \right| = |K_m| \leq C_1 N (\beta/\beta^*)^m m!, \quad (17)$$

for some constant C_1 .

- The truncated Taylor series gives a good approximation as

$$\left| \log Z - \sum_{m=0}^M \frac{\beta^m}{m!} K_m \right| \leq C_1 N \frac{(\beta/\beta^*)^{(M+1)}}{1 - (\beta/\beta^*)}. \quad (18)$$

The results we now describe prove these through an analysis of the individual terms in the expansion, which can be written in terms of the so-called *connected clusters*. A cluster is a multiset (that is, a set counting multiplicities) of Hamiltonian terms h_i (or alternatively, of hyperedges $\{i \in E\}$), which can appear more than once. A given cluster \mathbf{W} has size $|\mathbf{W}|$ equal to the number of elements in the multiset (counting multiplicities $\mu_i^{\mathbf{W}}$, so that $|\mathbf{W}| = \sum_{\{i \in \mathbf{W}\}} \mu_i^{\mathbf{W}}$). Moreover, \mathbf{W} is connected if the hypergraph with hyperedges $i \in \mathbf{W}$ is connected. Let us define the set of all clusters of size at most m with \mathcal{C}_m , and the set of all connected clusters as \mathcal{G}_m .

Now, let us define the Hamiltonian with auxiliary variables $\{\lambda_i\}$ as $H(\lambda) = \sum_i \lambda_i h_i$. We use this to introduce the cluster derivative

$$\mathcal{D}_{\mathbf{W}} = \prod_{i \in \mathbf{W}} \left(\frac{\partial}{\partial \lambda_i} \right)^{\mu_i^{\mathbf{W}}} \Bigg|_{\lambda=0}. \quad (19)$$

Here, the subscript $\lambda = 0$ means to set $\lambda_i = 0$ for all i after taking the derivatives. It is not hard to see that we can write

$$\beta^m K_m(\lambda) = \sum_{\mathbf{W} \in \mathcal{C}_m} \mathcal{D}_{\mathbf{W}} \log \text{Tr}[e^{-\beta H(\lambda)}], \quad (20)$$

where now the moments K_m depend on the $\{\lambda_i\}$.

Moreover, if $\mathbf{W} \notin \mathcal{G}_m$, we have $\mathbf{W} = \mathbf{W}_1 \cup \mathbf{W}_2$ where $\mathbf{W}_1, \mathbf{W}_2$ are non-overlapping clusters. Let us define $h_{\mathbf{W}_1}, h_{\mathbf{W}_2}$ as the Hamiltonian terms in those clusters, so that $[h_{\mathbf{W}_1}, h_{\mathbf{W}_2}] = 0$. We then have

$$\mathcal{D}_{\mathbf{W}} \log \text{Tr}[e^{-\beta H(\lambda)}] = \mathcal{D}_{\mathbf{W}} \log \text{Tr}[e^{-\beta(h_{\mathbf{W}_1}(\lambda) + h_{\mathbf{W}_2}(\lambda))}] \quad (21)$$

$$= \mathcal{D}_{\mathbf{W}} \log \text{Tr}[e^{-\beta h_{\mathbf{W}_1}(\lambda)}] + \mathcal{D}_{\mathbf{W}} \log \text{Tr}[e^{-\beta h_{\mathbf{W}_2}(\lambda)}] \quad (22)$$

$$= 0. \quad (23)$$

This means we can write the moments in terms of connected clusters only

$$\beta^m K_m(\lambda) = \sum_{\mathbf{W} \in \mathcal{G}_m} \mathcal{D}_{\mathbf{W}} \log \text{Tr}[e^{-\beta H(\lambda)}]. \quad (24)$$

To relate this back to the original moments, we simply need to set all of the $\lambda_i = 1$, as $K_m(1) = K_m$.

The fact that only connected clusters contribute is an important simplification, since it reduces their number dramatically. Then, a way to show the convergence of the series (see [23, 24]) is to prove the following:

- The number of connected clusters of size m is bounded by $C_1 N c_1^m$ for some constant c_1 [25, 26].
- The size of each cluster derivative for a cluster of size $|\mathbf{W}| = m$ is at most

$$\left| \mathcal{D}_{\mathbf{W}} \log \text{Tr}[e^{-\beta H(\lambda)}] \right| \leq (\beta c_2)^m m! \quad (25)$$

for some constant c_2 [23, 24].

The constants here can usually be taken to be simple functions of the lattice parameters k, J, h and of some property of the interaction graph (e.g. in [23] it is the degree of the dual graph). These facts together imply that $|K_m| \leq c_1 c_2 \equiv (\beta^*)^{-1} \simeq \mathcal{O}(1)$, so that the partition function is analytic for $\beta < \beta^*$ and that it is well approximated by its Taylor series. Other, more direct and general proofs also exist in the literature (see in particular [27] and other following works), which also give results of the form of Eq. (18) for some $(\beta^*) \simeq \mathcal{O}(1)$.

The next step is to prove that the individual Taylor terms can be computed efficiently. This requires two separate steps:

- The set of all clusters of size m can be enumerated in time $\text{poly}(N) \times \exp(\mathcal{O}(m))$ [23, 28]
- Each cluster derivative can be computed exactly in time $\text{poly}(N) \times \exp(\mathcal{O}(m))$ [22, 23].

We can thus add all the contributions from the different derivatives to obtain K_m in time $\text{poly}(N) \times \exp(\mathcal{O}(m))$. This, together with Eq. (18), implies that by calculating up to a degree $M = \mathcal{O}(\log(N/\epsilon) \times \log(\beta^*/\beta))$ there exists an ϵ -close additive approximation to $\log Z$ that can be computed in time $\text{poly}(N, \epsilon^{-1})$.

It is not an accident that this only works for high enough temperatures. We do not expect to be able to prove many general statements at all temperatures, due to the presence of thermal phase transitions (such as the one of the classical 2D Ising model), and to the fact that the ground state energy is computationally hard to estimate [13]. Let us note, however, that there are specific models in the literature for which the convergence can be guaranteed for larger ranges of temperatures (see e.g. [29] and references therein). See Sec. VII B for more details.

The log-partition function is the most common quantity to calculate with this method. This same technique also allows for e.g. the computation of expectation values such as $\text{Tr}[h_i e^{-\beta H} / Z]$

by differentiating by an extra λ_i in the cluster derivative. It can be also applied to other similar objects. A good example are characteristic functions of the form $\text{Tr}[e^{\alpha A} e^{-\beta H} / Z]$ for some local observable A , which allows for the derivation of probability theory statements, as explained further in Sec. [VIB](#).

B. Thermal locality estimates

We now show the first method to decompose the thermal state into a product of smaller operators, despite the non-commutativity, which is related to the idea of operator growth. Consider an operator A with local support on some small region on the lattice. For simplicity, let us consider that this region is such that $|\text{supp}(A)| \leq k$.

An interesting quantity to study is the operator evolved in Euclidean or imaginary time β under the Hamiltonian H ,

$$A(i\beta) = e^{-\beta H} A e^{\beta H}. \quad (26)$$

This is in analogy with the Heisenberg-picture operator $A(t) = e^{-itH} A e^{itH}$, which can be understood in terms of the well-known Lieb-Robinson bounds [\[30\]](#), that state that the support of $A(t)$ is mostly confined to a linear lightcone. In many situations, one will want to choose A here to be one of the h_i operators.

It then makes sense to ask the following question: what is the locality of the Euclidean-evolved operator $A(i\beta)$? Perhaps surprisingly, this can be dramatically different to the real-time case: there is no general linear growth with the inverse temperature β , but a much wilder dependence on it. The main difference is that e^{-itH} is a unitary matrix, while $e^{-\beta H}$ is not. This means that results that exploit unitarity, such as the aforementioned Lieb-Robinson bounds, do not apply straightforwardly. Our best way forward seems then to analyze $A(i\beta)$ in terms of nested commutators

$$A(i\beta) = \sum_{m=0}^{\infty} \frac{(-\beta)^m}{m!} [H, [H, \dots, [H, A] \dots]] \equiv \sum_{m=0}^{\infty} \beta^m C_m(A). \quad (27)$$

It is easy to see that the m -th term in this expansion has support on a connected region whose furthestmost point is a distance m away from A . The question then becomes: how does this expansion in terms of β converge?

It can be shown that, for general interaction graphs, [\[21, 31\]](#)

$$\|C_m(A)\| \leq k \|A\| (2Jk)^m. \quad (28)$$

This statement is very much related to the bound on the number of connected clusters in Sec. [III A](#) above, since only connected clusters contribute to the nested commutators. Eq. [\(28\)](#) implies that as long as $\beta < (2Jk)$, the expansion can be controlled as a geometric series, from which we obtain

$$\|A(i\beta)\| \leq k \|A\| \frac{1}{1 - 2\beta Jk} \quad (29)$$

$$\|A(i\beta) - \sum_{m=0}^M \beta^m C_m(A)\| \leq k \|A\| \frac{(2\beta Jk)^{M+1}}{1 - 2\beta Jk}. \quad (30)$$

Given that the m -th nested commutator can have support on at most $k \times m$ sites, the latter equation means that $A(i\beta)$ is, roughly speaking, localized within the subset of vertices a distance at most $k \times m$ away from $\text{supp}(A)$. It is known that in general one cannot extend this result to temperatures

lower than $\beta \simeq \mathcal{O}(1)$, since there exists an example of a 2D lattice in which the terms in the nested commutators in Eq. (27) add up constructively, in a way that the norm of $A(i\beta)$ diverges [32].

On the other hand, it has been known for some time [33] that when the lattice is a one-dimensional chain, the nested commutators grow more slowly, so that this type of convergence happens for all temperatures. For simplicity, we show explicitly the result for $k = 2$ combining [34] and [32], which is

$$\|A(i\beta)\| \leq \|A\| f(\beta, J) \exp(f(\beta, J)) \quad (31)$$

$$\|A(i\beta) - \sum_{m=0}^M \beta^m C_m(A)\| \leq 15\|A\| e^{-(M+1)} \quad \forall M > g(\beta, J). \quad (32)$$

Here, we have defined $f(\beta, J) \equiv 16\beta J \exp(1 + 8\beta J)$ and $g(\beta, J) \equiv \exp(240e^2\beta J) - 1$ [35]. The intuitive reason for these is that the geometric bound of Eq. (28) can be improved in this case as [32] (again, for $k = 2$)

$$\|C_m(A)\| \leq 15\|A\| \left(\frac{240eJ}{\log(m+1)} \right)^m. \quad (33)$$

Notice that, because of the logarithm, the series in Eq. (27) is not geometric, and converges for all β . For further explanations of these points see also [36].

So far, we have described how does $A(i\beta)$ approximate its Taylor expansion. An alternative approximation commonly considered is to the operator $e^{-\beta H_{\Lambda_m}} A e^{\beta H_{\Lambda_m}}$, where $H_{\Lambda_m} = \sum_{\text{supp}(h_i) \in \Lambda_m} h_i$ and Λ_m is some region of the full lattice Λ . One can then consider how the norm

$$\|A(i\beta) - e^{-\beta H_{\Lambda_m}} A e^{\beta H_{\Lambda_m}}\| \quad (34)$$

decays with m in terms of how Λ_m is defined (typically, some hyper-sphere of the lattice). The analysis and convergence turn out to be almost the same as the one for the moments $C_m(A)$ above. The reason is that the difference between $\sum_{m=0}^M \beta^m C_m(A)$ and $e^{-\beta H_{\Lambda_m}} A e^{\beta H_{\Lambda_m}}$ are essentially the higher order terms in β of the latter, which are also suppressed. See e.g. [34] for a detailed analysis of the 1D case or e.g. Lemma 20 in [37] for a proof in higher dimensions.

One of the main reasons why both of these approximations are interesting is that they are related to the following propagator

$$E_A \equiv e^{-\beta(H+A)} e^{\beta H} = \mathcal{T} e^{-\int_0^\beta ds e^{-sH} A e^{sH}}, \quad (35)$$

where \mathcal{T} denotes the usual time-ordered integral. This is such that, for $\beta < (2Jk)^{-1}$

$$e^{-\beta(H+A)} = E_A e^{-\beta H}. \quad (36)$$

This operator can be analyzed through a usual Dyson series in terms of powers of $e^{-xH} A e^{xH}$. From this it can be shown that E_A has bounded norm as it follows from Eq. (29) and (35) that

$$\|E_A\| \leq \exp\left(\int_0^\beta ds \|e^{-sH} A e^{sH}\|\right) = \left(\frac{1}{1-2\beta Jk}\right)^{k\|A\|}. \quad (37)$$

In Appendix A we also show that it is approximately localized in the same way as $e^{-xH} A e^{xH}$ is. This means that there exist an operator $E_A(l)$ with support restricted to a distance at most l away from A such that for $\beta < (2Jk)^{-1}$

$$\|E_A - E_A(l)\| \leq \beta k \|A\| \frac{(2\beta Jk)^{l+1}}{1-2\beta Jk}. \quad (38)$$

Also, notice that if $[H, A] = 0$, then $E_A = e^{-\beta A}$. With the right choice of H, A , the operator E_A can thus be thought of as a “transfer operator”.

Alternatively, one can also define the following operator

$$E'_A \equiv e^{-\beta(H+A)} e^{\beta H} e^{\beta A}, \quad (39)$$

with the difference that H, A are now treated on equal footing. In this case, E'_A is just the multiplicative error term in the first order Trotter product formula, which can be similarly analyzed through the expansion of $A(i\beta)$ (see the thorough analysis of Trotter errors in [38] for more details). These Trotter errors are most commonly analyzed in the context of digital quantum simulation [39], for which it is often convenient to go to higher orders in the decomposition.

We finish this subsection outlining a result in 1D related to this discussion, which follows from bounds on the quantity in Eq. (34). It appeared first in [33], and it features in Sections V A 1 and VII A. Let us define $E_A^l = e^{-\beta(H_l+A)} e^{\beta H_l}$, where H_l are the interaction terms a distance at most l away from $\text{supp}(A)$. It can be shown that

$$\|E_A\| \leq C_1 \quad (40)$$

$$\|E_A - E_A^l\| \leq C_2 \frac{q^l}{(l+1)!}, \quad (41)$$

where C_1, C_2 and $q > 1$ are constants depending on k, J, β which we do not show explicitly for simplicity, although notice that C_1 will be essentially the exponential of Eq. (31). The proof is similar to that of Appendix A, together with a bound on Eq. (34). We refer the reader to e.g. [33, 34] for further details in 1D, and a similar discussion also applies in higher dimensions and high temperatures.

C. Quantum belief propagation

An idea related to the previous locality estimates appeared first [40], and has more recently featured in several results [12, 37, 41–44]. It also allows us to decompose the thermal state as a product of smaller, localized operators, which makes certain calculations more tractable. This is part of a celebrated series of works including the decay of correlations for gapped ground states [45] or the area law of entanglement in one dimension [46] which show how Lieb-Robinson bounds (a dynamical statement) can be used to prove static properties about ground and thermal states. See [47, 48] for overviews.

We follow the presentation from [41, 42] (see also [49]). The idea is to consider the “perturbed” Hamiltonian $H(s) = H + sA$ and the following derivative

$$\frac{de^{-\beta H(s)}}{ds} = -\frac{\beta}{2} \left\{ e^{-\beta H(s)}, \Phi_\beta^{H(s)}(A) \right\} \quad (42)$$

where, if $H(s) = \sum_i E_i(s) |i(s)\rangle \langle i(s)|$ is the energy eigenbasis,

$$\Phi_\beta^{H(s)}(A)_{ij} = \langle i(s) | A | j(s) \rangle \tilde{f}_\beta(E_i(s) - E_j(s)), \quad (43)$$

where $\tilde{f}_\beta(\omega) = \frac{\tanh(\beta\omega/2)}{\beta\omega/2}$. With $f_\beta(t) = \frac{4}{\beta\pi} \log\left(\frac{e^{\pi|t|\beta+1}}{e^{\pi|t|\beta-1}}\right)$ the Fourier transform of $\tilde{f}_\beta(\omega)$ (see Appendix B of [12]), we can also write

$$\Phi_\beta^{H(s)}(A) = \int_{-\infty}^{\infty} dt f_\beta(t) e^{-itH(s)} A e^{itH(s)}. \quad (44)$$

Since $\|e^{-itH(s)} A e^{itH(s)}\| = \|A\|$, $\Phi_\beta^{H(s)}(A)$ has bounded operator norm, and can also be approximated by a localized operator by using Lieb-Robinson bounds, which reads

$$\| [M_C, e^{-itH} N_D e^{itH}] \| \leq c \|M_C\| \|N_D\| \min\{|C|, |D|\} e^{c'(vt - \text{dist}(C,D))}, \quad (45)$$

where M_C, N_D are operators on regions C, D on the lattice, separated by distance $\text{dist}(C, D)$.

We now integrate Eq. (42) between $s = 0$ and $s = 1$ to obtain

$$e^{-\beta(H+A)} = O_A e^{-\beta H} O_A^\dagger, \quad (46)$$

where

$$O_A = \mathcal{T} e^{-\frac{\beta}{2} \int_0^1 ds \Phi_\beta^{H(s)}(A)}. \quad (47)$$

Similarly to Eq. (35) above (see Appendix A), this operator has a bounded norm, and also is very close to a localized one. More precisely, we have that

$$\|O_A\| \leq e^{\frac{\beta}{2} \int_0^1 ds \|\Phi_\beta^{H(s)}(A)\|} = e^{\frac{\beta}{2} \|A\|}, \quad (48)$$

and also, that there exists an operator O_A^l defined as

$$O_A^l \equiv \mathcal{T} e^{-\frac{\beta}{2} \int_0^1 ds \Phi_\beta^{H_l(s)}(A)}, \quad (49)$$

with $\Phi_\beta^{H_l(s)}(A)$ defined as the Hamiltonian terms of $H(s)$ with support on a ball of radius l centered at $\text{supp}(A)$. This is such that

$$\|O_A - O_A^l\| \leq \frac{c' \beta \|A\|}{2} e^{(1+c') \frac{\beta}{2} \|A\|} e^{-\frac{c'l}{1+c'v \frac{\beta}{\pi}}}. \quad (50)$$

These bounds can be compared to Eq. (40) and (41), which are of a very similar nature. There are, however, two important differences between O_A and E_A in Eq. (35) above:

- Since it is based on the Lieb-Robinson bound, the operator O_A is well-behaved in all lattices and at all temperatures, in the sense that it has a bounded norm and is approximately localized. This is as opposed to E_A , which is likely a large operator in high dimensions and low temperatures.
- On the other hand, to recover $e^{-\beta(H+A)}$ from $e^{-\beta H}$ we require left and right multiplication with O_A, O_A^\dagger , as opposed to Eq. (36), which may be problematic in some applications. In particular, we should not expect it to be a key ingredient in proving results that do not hold at all temperatures such as the decay of correlations or the analyticity of the partition function.

D. Selected trace inequalities

In the past two subsections we have explored how to analyze perturbations to the Hamiltonian in the Gibbs operator $e^{-\beta(H+A)}$ via E_A and O_A . When considering traces, simpler identities hold. We exemplify this with two with very elementary implications and proofs, which can be found in (at least) [50]. The first one is about the stability of partition functions. Let H_1, H_2 be Hermitian operators. Then we have

$$|\log \text{Tr}[e^{H_1+H_2}] - \log \text{Tr}[e^{H_1}]| \leq \|H_2\|. \quad (51)$$

The proof is just as follows:

$$|\log \text{Tr}[e^{H_1+H_2}] - \log \text{Tr}[e^{H_1}]| = \left| \int_0^1 \frac{d}{dt} \log \text{Tr} [e^{H_1+tH_2}] dt \right| \leq \int_0^1 \left| \frac{\text{Tr}[H_2 e^{H_1+tH_2}]}{\text{Tr}[e^{H_1+tH_2}]} \right| dt \leq \|H_2\|, \quad (52)$$

where in the last inequality we have simply used Hölder's inequality Eq. (6) with $q_1 = 1, q_2 = \infty$. If we take e.g. $H_1 = -\beta H, H_2 = -\beta A$, this implies that changing the Hamiltonian by A changes the log-partition function at most by $\beta \|A\|$.

The second is a similar result that holds for expectation values of positive operators. Let H_1, H_2 be as before, and let $C > 0$. Then

$$|\log \text{Tr}[C e^{H_1+H_2}] - \log \text{Tr}[C e^{H_1}]| \leq \int_0^1 dt \int_{-1/2}^{1/2} ds \|e^{-s(H_1+tH_2)} H_2 e^{s(H_1+tH_2)}\|. \quad (53)$$

The proof can be found in Appendix A. This norm can then be bounded with the results from Sec. III B, to scale as $\propto \|H_2\|$. The resulting expression can for instance be used for analyzing characteristic functions of observables F by taking $C = e^{\alpha F}$ for $\alpha \in \mathbb{R}$ [50].

More generally, in the practice of mathematical quantum physics, whether it is from the many body, the QI, or any other perspective, many important proof ingredients take the form of inequalities, either between operators, traces, or norms (such as those already mentioned in Sec. II A). There are too many to give a reasonably complete overview here but we refer the reader to e.g. [51, 52].

IV. CORRELATIONS

One of the more important questions when studying many body systems is: how and how much are the different parts correlated? Intuitively, the stronger these correlations, and the longer their range, the more complex a state is - the reason being that we cannot think of the large system as a collection of simpler, weakly correlated parts. The obvious extreme example is that of an uncorrelated gas, in which the particles do not interact and have independent distributions.

For thermal states with local interactions, we can expect that locality will make the state far from generic, in a way that constraints its complexity. Intuitively, it should cause the correlations to be "localized", meaning that particles are only correlated with their vicinity as given by the lattice geometry. For a rough intuition, consider the first terms of the Taylor series

$$e^{-\beta H} = \mathbb{I} - \beta \sum_i h_i + \frac{\beta^2}{2} \sum_{i,j} h_i h_j + \dots \quad (54)$$

That is, at very high temperatures we approach the trivial uncorrelated state $\propto \mathbb{I}$ and the leading order term includes only k -local couplings, with only higher order terms coupling far away particles. We thus expect that the correlations between particles will generally be weaker *i)* the higher the temperature and *ii)* the larger their distance on the interaction graph.

One of the main motivation is that, as we will see in later sections, the situations in which the correlations are weaker or short range roughly correspond to those in which there are better algorithms for the description of thermal states. This is perhaps most clearly the case in the context of tensor network methods. We now proceed to describe (and even prove) the more important ways in which these correlations are constrained.

A. Correlations between neighbouring regions: Thermal area law

One of the more important statements about correlations in quantum many-body systems is the area law. This roughly states that a measure of correlations between two adjacent regions is upper-bounded by a number proportional to the size of their mutual boundary.

Traditionally, this has been mostly studied in the context of ground states, which are pure. There, the measure of correlations is the entanglement entropy, or some Rényi version of it. In that context, an area law for the entanglement entropy is believed to hold for all ground states of models with a gap [53]. This can be proven in 1D [46, 54] and in some cases in 2D [55]. The interest in it is largely due to its relation to other phenomena, such as phase transitions [56], the decay of long-range correlations [57] or the effectiveness of certain tensor network algorithms [58, 59].

For thermal states, a very general area law can be shown to hold for systems in any dimension, at all temperatures. We now give a short proof of this statement, and then discuss its significance (see [60] for the original reference). In this case, since it is a mixed state, an appropriate measure of correlations is the mutual information in Eq. (12).

Let us partition our interaction graph into two subsets of particles A, B , with a thermal state ρ_β^{AB} . We start with the very simple thermodynamic observation that the free energy F of the thermal state is lower than that of any other state (this follows from Eq. (11)), and in particular

$$F_\beta(\rho_\beta^{AB}) \leq F_\beta(\rho_\beta^A \otimes \rho_\beta^B). \quad (55)$$

Writing out the free energy explicitly as $F_\beta(\rho) = \text{Tr}[\rho H] - \beta^{-1}S(\rho)$ and rearranging yields

$$S(\rho_\beta^A \otimes \rho_\beta^B) - S(\rho_\beta^{AB}) \leq \beta (\text{Tr}[H\rho_\beta^A \otimes \rho_\beta^B] - \text{Tr}[H\rho_\beta^{AB}]). \quad (56)$$

Given that the entropy is additive $S(\rho \otimes \sigma) = S(\rho) + S(\sigma)$ notice that the LHS is exactly the mutual information $I(A : B)_{\rho_\beta^{AB}}$ from Eq. (12). Since the Hamiltonian is local, we can write it as

$$H = H_A + H_B + H_I, \quad (57)$$

where H_A, H_B have support only on A, B respectively, and H_I is the interaction between them (with support on both). By definition, the expectation values of H_A and H_B coincide on both states $\text{Tr}[(H_A + H_B)\rho_\beta^A \otimes \rho_\beta^B] = \text{Tr}[(H_A + H_B)\rho_\beta^{AB}]$, so that

$$\beta (\text{Tr}[H\rho_\beta^A \otimes \rho_\beta^B] - \text{Tr}[H\rho_\beta^{AB}]) = \beta (\text{Tr}[H_I\rho_\beta^A \otimes \rho_\beta^B] - \text{Tr}[H_I\rho_\beta^{AB}]). \quad (58)$$

Now we can use a few of the operator inequalities from Section II A to obtain

$$\text{Tr}[H_I\rho_\beta^A \otimes \rho_\beta^B] - \text{Tr}[H_I\rho_\beta^{AB}] \leq \|H_I(\rho_\beta^A \otimes \rho_\beta^B - \rho_\beta^{AB})\|_1 \quad (59)$$

$$\leq \|H_I\| \times \|\rho_\beta^A \otimes \rho_\beta^B - \rho_\beta^{AB}\|_1 \quad (60)$$

$$\leq \|H_I\| \times (\|\rho_\beta^A \otimes \rho_\beta^B\|_1 + \|\rho_\beta^{AB}\|_1) = 2\|H_I\|. \quad (61)$$

Putting Eq. (56) and (59) together we have the final result

$$I(A : B)_{\rho_\beta^{AB}} \leq 2\beta\|H_I\|. \quad (62)$$

This is the *area law* for the mutual information of a thermal state: it implies that the strength of the correlations of systems A, B cannot depend on their size, but that it grows at most as their common boundary. For a local Hamiltonian, we have that

$$\|H_I\| \leq 2kh|\partial_{AB}|, \quad (63)$$

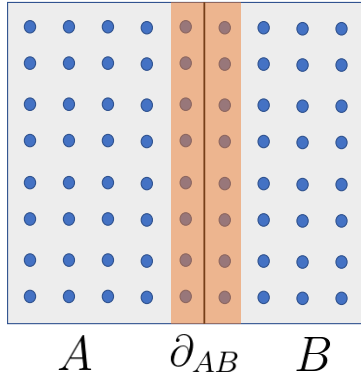


Figure 1: The correlations between regions A and B grow at most as their mutual boundary ∂_{AB} .

where $|\partial_{AB}|$ is the number of particles at the boundary of regions A, B , as defined by the interaction graph. Notice that with $|\partial_{AB}|$ we do not mean the size of the boundary of systems A, B together, but the number of elements of ∂_A that are connected to ∂_B by hyperedges. We show this schematically in Fig. 1. This is to be contrasted with the most general upper bound on the mutual information, which is $I(A : B) \leq \min\{\log(d_A), \log(d_B)\}$ (since $\log d_A \propto |A|$ this would be a “volume law” instead).

What this strongly suggests (although it does not quite prove) is that the correlations between A and B are localized around the mutual boundary, and that the bulks of A and B are mostly uncorrelated. That is, the only relevant information about A that B contains is about the region of A that is near their boundary.

This statement, as can be seen from the proof, holds for all temperatures and all interaction graphs, which is likely as general as it can be. The drawback of that generality, however, is that it will be unable to signal important phenomena that happens only at specific temperature ranges, such as thermal phase transitions, or an efficient classical or quantum simulability. Other more specific versions of the thermal area law in the literature may have more potential in this regard. It can be shown for a wide range of settings for different measures of correlations, such as the entanglement negativity [61], and Renyí generalizations of the mutual information [19], which follow from the locality estimates in Sec. III B.

Let us finish by noting that the temperature dependence of Eq. (62) can be improved to $\tilde{\mathcal{O}}(\beta^{2/3})$ [37]. This can be proven with a variety of techniques, including those of Sec. III B and Sec. III C, as well as methods originated in the study of ground states [54]. This dependence is not far from optimal, since there exists a 1D model for which the scaling of the MI is at least $\mathcal{O}(\beta^{1/5})$ at low temperatures [62].

This idea also suggests that the scaling of the mutual information with β in the low temperature regime is related to the computational complexity of the ground space of the models. However, many important physical models have a very different temperature dependence, such as $\log(\beta+1)$ [63, 64]. Classical systems, on the other hand, have an upper bound that is independent of the temperature, as $I(A : B)_{\rho_\beta^{AB}} \leq |\partial_{AB}| \log d$ [60].

B. Decay of long-range correlations

An important statement about thermal states is that often their spatially separated parts are very weakly correlated. Let C, D be regions such that their distance is $\text{dist}(C, D)$ (see Fig. 2). We

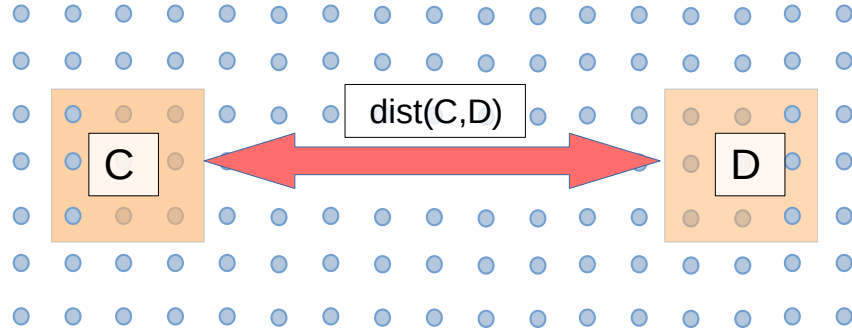


Figure 2: Regions C, D in the lattice are separated by a distance $\text{dist}(C, D)$. The mutual information between these two regions typically decays exponentially with their distance.

focus on certain measure of correlations evaluated at the marginals on these regions $\text{Tr}_{\setminus(CD)}[\rho_\beta] = \rho_\beta^{CD}$. For instance, taking the mutual information, we expect that in general

$$I(C : D)_{\rho_\beta^{CD}} \leq f(\text{dist}(C, D)), \quad (64)$$

where f is some rapidly decaying function. In fact, it is most often the case that $f(l) \leq K|\partial_C||\partial_D|e^{-l/\xi}$, where $K > 0$ is some constant, $\partial_{C,D}$ is the size of the boundary of each region, and ξ is the thermal *correlation length* that depends on the temperature and other parameters, but not on l or the system's size. This can be proven in the two following scenarios:

- For any k -local interaction graph above a threshold temperature $\beta < \beta^*$, where β^* depends on parameters of the Hamiltonian (but not on its size). This has been proven through the cluster expansion technique outlined in Sec. III A [37, 65, 66].
- For 1D systems at all temperatures [67]. This has been proven using the locality estimates from III B, and in particular the properties of the operator E_A in Eq. 36.

The proofs are slightly involved and beyond the scope of these notes, so we refer the reader to the original references.

A more commonly stated but weaker condition is the decay of the connected two-point correlators. This usually takes the form

$$\frac{|\text{Tr}[\rho_\beta M_C \otimes N_D] - \text{Tr}[\rho_\beta M_C] \text{Tr}[\rho_\beta N_D]|}{\|M_C\| \|N_D\|} \leq K|\partial_C||\partial_D|e^{-\text{dist}(C,D)/\xi}, \quad (65)$$

where here M_C and N_D have support on regions C, D , respectively. This follows simply from the decay of the mutual information and Pinsker's inequality applied to the marginal on regions C, D .

This general property of correlation decay has been shown, at least in some cases, to be equivalent to the analyticity of the partition function [44], although the close connection between the two is better understood classically [68]. Both these properties in turn are related to the absence of phase thermal phase transitions, which implies no long-range order. At those phase transitions, the correlation function diverges and the partition function becomes non-analytic. Since there are known phase transitions at finite temperature (e.g. 2D classical Ising model), the exponential decay does not hold for all thermal states at all temperatures.

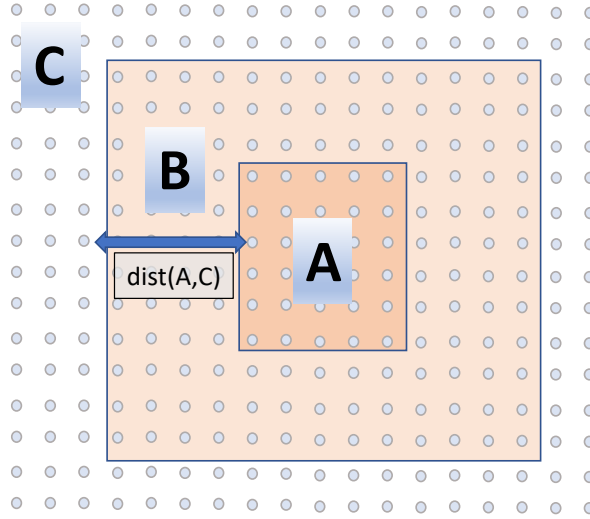


Figure 3: In this configuration, the region B shields A from C , such that the minimum distance between A and C is given by the shortest path from A to C through B . The CMI $I(A : C|B)$ decays exponentially with that distance.

The decay of correlations is an important fact. It shows that the different parts of the system behave almost completely independently. A state with this property should then share many large-scale features with an uncorrelated gas, in which the particles are not interacting at all. As such, this property has as a wealth of related physical consequences. For instance, it is associated with basic statistical physics facts covered in Sec. VI, in particular the validity of the central limit theorem and related results on concentration properties of thermal states [69, 70] and the phenomenon of equivalence of ensembles [69, 71, 72]. It also features in the proof of *local indistinguishability* in Sec. V A 1.

C. A refined correlation decay: Conditional mutual information

In our discussion of the thermal area law in Sec. IV A, we mentioned that the bound by itself does not quite imply that the correlations in a system are localized, in the sense that a particular subsystem is only appreciably correlated with its vicinity. There is, however, a significantly stronger statement about correlations that does imply it in a clear way.

This is the property of being an approximate *quantum Markov state* [73], which is defined in terms of the decay of the CMI in Eq. (13). For this property, we need to consider three regions A, B, C such that B shields A from C . A simple example of this is given in Fig. 3, with an illustration of the 1D case in Fig. 4.

Since this quantifies how many of the correlations between A and C are not mediated through B , we thus expect that it becomes small as the size of B grows, and A, C are further apart. This is perhaps the strongest sense in which correlations can be localized. For instance, the decay of the mutual information at long distances already follows from the statement from [74], by choosing $B = \emptyset$ to be the empty set, so that $I(A : C|B) = I(A : C)$. The following two results are known

- In one dimension, by choosing A, B, C to be adjacent regions (see Fig. 4), [42] shows that

$$I(A : C|B)_{\rho_\beta} \leq c_1 |B| e^{-c_2 \sqrt{|B|}}. \quad (66)$$

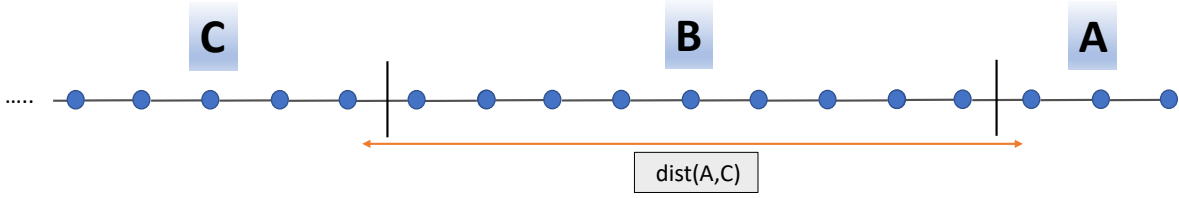


Figure 4: In this chain, the middle region B shields A from C , and their distance is related to the size of B .

The key technique to show this is the quantum belief propagation from Sec. III C, but the locality bounds from Sec. III B are also sufficient.

- In larger dimensions, and at high temperatures [74] $\beta \leq \beta^*$, (see Fig. 3)

$$I(A : C|B)_{\rho_\beta} \leq k_1 \min\{|\partial A|, |\partial C|\} \left(\frac{\beta}{\beta^*}\right)^{-k_2 \times \text{dist}(A,C)}. \quad (67)$$

This is shown by using the connected cluster expansion in Sec. III A, applied to the CMI instead of the log-partition function.

For 1D, the converse is also true: any state with a sufficiently fast decaying CMI approximates the thermal state of some local Hamiltonian [42]. Also in 1D it is expected that the decay is $e^{-\Omega(|B|)}$ as opposed to Eq. (66), which may be important for certain applications [42, 43, 75].

The significance of a fast decay of the CMI is highlighted by some further quantum information tools, in particular, the idea of the Petz map [76, 77]. An important result in this regard states that, given a tripartite state $\rho = \rho^{ABC}$, there exists a CPTP map $\mathcal{N}(\cdot)_{B \rightarrow BC}$ (that is, acting on B , and with output on BC) such that [78, 79]

$$I(A : C|B)_\rho \geq 2 \|\rho^{ABC} - \mathcal{N}(\rho^{AB})_{B \rightarrow BC}\|_1. \quad (68)$$

The map \mathcal{N} usually goes under the name of *recovery map*. See [80] for an overview of this topic and the proof techniques involved.

When applied to thermal states, the fast decay of the CMI thus guarantees that the state on A, B, C can be reconstructed from ρ_{AB} by acting locally on B (and importantly, not on A), such that $\mathcal{I}_A \otimes \mathcal{R}_{B \rightarrow BC}(\rho^{AB}) \simeq \rho^{ABC}$, with $\mathcal{R}_{B \rightarrow BC}$ some CP map taking only B as input. This gives a way of sequentially preparing the whole thermal state from its smaller components, which can potentially be used e.g. for quantum algorithms (see Sec. VII).

The results described here can be seen as the quantum analogues of a much stronger statement that holds for classical probability distributions: the Hammersley-Clifford theorem [81]. This says that a classical Gibbs state of a local Hamiltonian with interaction graph Λ is also a so-called Markov random field defined in terms of the graph Λ . This means that for classical Hamiltonians, the CMI as defined here is always zero. In Sec. VIII we will see that a similar result also holds for *commuting* quantum Hamiltonians.

V. LOCALITY OF TEMPERATURE

In the previous section we focused the correlations between different parts. Now, we move the spotlight to features of individual subsystems. That is, if we divide the system into A and its

complement $\setminus A$, what does $\text{Tr}_{\setminus A}[\rho_\beta]$ look like? In the rest of the section we drop the subscript β for simplicity of notation.

We can illustrate the situation with the trivial case: if the particles are non-interacting, it holds that the marginal on A is the thermal state of H_A which is the Hamiltonian that acts only on subsystem A . That is

$$\rho^A \equiv \text{Tr}_{\setminus A}[\rho] = \frac{e^{-\beta H_A}}{Z_A}. \quad (69)$$

However, this only holds because the sites are independent. The question is: how does Eq. (69) change when we introduce local (and potentially strong) interactions? Can we identify the state of a subsystem with some thermal state? How different is it from $\frac{e^{-\beta H_A}}{Z_A}$? This general question sometimes goes under the name of *locality of temperature* [65].

There are (to the author's knowledge) two different but related answers to this: the idea of *local indistinguishability* and also the notion of *Hamiltonian of mean force*. We now explain both of them, elaborate on their significance, and also give a proof of the simplest instance of the first (in 1D). These results largely simplify the study of local properties of thermal states, in that they show how local properties are largely independent of the bulk and can be calculated just by computing a small subsystem [65, 82].

A. Local indistinguishability

Given the discussion in the previous section on the decay of correlations, we expect that the state of a local subsystem does not depend much on the parts that are far away enough from it. A possible way to phrase this is that the local marginal ρ_A is indistinguishable from the marginal of a much smaller thermal state, with a Hamiltonian that acts only in the vicinity of A . We now make this intuition precise. Let us refer to the partition in Fig. 3 or Fig. 4, and write the Hamiltonian with the following terms:

$$H = H_A + H_{AB} + H_B + H_{BC} + H_C. \quad (70)$$

We now have the full thermal state ρ , as well as a thermal state supported on A, B defined as

$$\rho_0^{AB} = \frac{e^{-\beta(H_A+H_B+H_{AB})}}{Z_{AB}}, \quad (71)$$

that is, without the terms in H that have support in C (note that C may comprise most of the system). One can also think of this as the marginal of the thermal state $\rho_0^{AB} \otimes \rho_0^C \equiv e^{-\beta(H_{AB}+H_C)}/Z_{AB}Z_C$ in which we have removed the interactions H_{BC} between AB and C . Notice that $\rho_0^{AB} \neq \rho^{AB}$ due to the presence of H_{BC} . This is, however, just a small local term.

The main idea is that if B is large enough, these two states are almost indistinguishable on A . Let us assume that the connected correlations from Eq. (65) decay with function $f(\text{dist}(C, D))$. Then, the following upper bound holds for some constant $K > 0$ [43]

$$\|\text{Tr}_{BC}[\rho] - \text{Tr}_B[\rho_0^{AB}]\|_1 \leq K|\partial\delta_C| \left(f(\text{dist}(A, C)) + e^{-\frac{c'\text{dist}(A, C)}{1+c'v\frac{\beta}{\pi}}} \right). \quad (72)$$

The first term in the RHS comes from the decay of correlations assumption. The second comes from using the QBP technique in Sec. III C. The exponential decay of this quantity thus holds

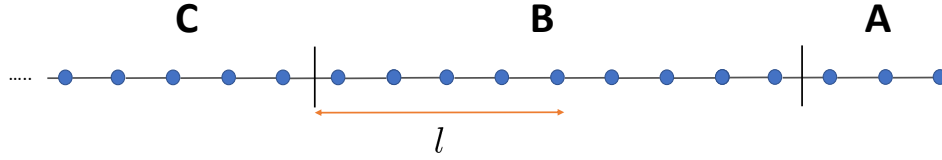


Figure 5: Choice of regions for the proof in Sec. V A 1, and depiction of the distance l on which the operator E_{BC}^l acts within B .

whenever both the correlations decay fast enough, and Lieb-Robinson bounds hold. An alternative proof for high temperatures using the cluster expansion can also be found in [65].

A straightforward consequence of this is that we do not need to know the whole state to compute local quantities. If we care about some kind of local order parameter, or want to compute currents or else between some part and its surroundings, we can calculate them without having to diagonalize a huge matrix of size $\exp(N)$, but rather just focus on a much smaller region. This has various implications for the construction of classical and quantum algorithms, as we describe in Sec. VII.

1. Proof in 1D

We now show the full proof of this result in the case of one dimension. The more general one, however, is essentially the same and can be found in [43]. It uses previously mentioned results, and shares some steps and ideas that appear in other perhaps more fundamental questions including, for instance, the proof of the absence of phase transitions in 1D [33, 44] or of decay of correlations [33, 67]. It will also be a key ingredient in the algorithm of Sec. VII A. We believe these reasons make it of pedagogical interest.

We focus on the restricted setting of a chain, that we divide into three parts A, B, C , such that B is in the middle and A is a small subsystem at the end of the chain, as in Fig. 5. The aim is a small upper bound on

$$\|\text{Tr}_{BC}[\rho] - \text{Tr}_B[\rho_0^{AB}]\|_1 = \max_{\|N_A\| \leq 1} |\text{Tr}[N_A(\rho - \rho_0^{AB} \otimes \rho_0^C)]|, \quad (73)$$

where N_A has support on A only, and the equality comes from the definition of the 1-norm. Now, let us define the following two operators

- $E_{BC} = e^{\beta H} e^{-\beta(H-H_{BC})}$,
- $E_{BC}^l = e^{\beta(H_C^l + H_B^l + H_{BC})} e^{-\beta(H_B^l + H_C^l)}$, where H_B^l and H_C^l are the terms of H_B and H_C that are a distance at least l from the boundary terms H_{BC} .

The second one E_{BC}^l is the same as E_{BC} but restricting the terms that appear in the exponents to be in the vicinity of H_{BC} . The parameter l is free, so we can choose to our convenience. We refer now to the result from [33] in Eq. 40 and 41, from which it follows that

$$\|E_{BC}\| \leq C_1 \quad (74)$$

$$\|E_{BC} - E_{BC}^l\| \leq C_2 \frac{q^{1+l}}{(l+1)!} \quad (75)$$

That is, the operator E_{BC} has bounded norm and, since we can approximate it by E_{BC}^l with some $l < \text{dist}(A, C)$, its support on region A is super-exponentially suppressed in l (due to the factorial,

which always dominates over q^l). In what follows, we choose $l = |B|/2$. Notice that by definition $\rho_0^{AB} \otimes \rho_0^C = \frac{Z}{Z_{AB}Z_C} \rho E_{BC}$.

Let us define N_A^* to be the operator that optimizes the RHS of Eq. (73). With the triangle inequality we can write

$$|\text{Tr}[N_A^*(\rho - \rho_0^{AB} \otimes \rho_0^C)]| \leq \left| \text{Tr}[N_A^*(\rho - \frac{Z}{Z_{AB}Z_C} \rho E_{BC}^l)] \right| + \left| \text{Tr}[N_A^*(\frac{Z}{Z_{AB}Z_C} \rho E_{BC}^l - \rho_0^{AB} \otimes \rho_0^C)] \right|. \quad (76)$$

Let us now upper-bound these two terms independently. The second can be bounded with Eq. (75) and Hölder's inequality applied twice.

$$\left| \text{Tr}[N_A^* \frac{Z}{Z_{AB}Z_C} \rho E_{BC}^l - \rho_0^{AB} \otimes \rho_0^C] \right| = \left| \text{Tr}[N_A^* \frac{Z}{Z_{AB}Z_C} \rho (E_{BC}^l - E_{BC})] \right| \quad (77)$$

$$\leq \frac{Z}{Z_{AB}Z_C} \|N_A^*\| \|\rho\|_1 \|E_{BC}^l - E_{BC}\| \quad (78)$$

$$\leq \frac{Z}{Z_{AB}Z_C} \times C_2 \frac{q^{1+l}}{(1+l)!}. \quad (79)$$

Given Eq. (51), $\max\{\frac{Z}{Z_{AB}Z_C}, \frac{Z_{AB}Z_C}{Z}\} \leq e^{\beta \|H_{BC}\|}$, which is a constant that only depends on β, k, J . Thus, this second term is super-exponentially suppressed in $|B|$.

For the first term, we require the decay of correlations property Eq. (65) (which, as explained above, always holds in 1D). Since $l = |B|/2$,

$$\left| \text{Tr}[N_A^* \rho E_{BC}^l] - \text{Tr}[N_A^* \rho] \text{Tr}[\rho E_{BC}^l] \right| \leq K e^{-\frac{|B|}{2\xi}} \|E_{BC}^l\| \leq 2KC_1 e^{-\frac{|B|}{2\xi}}, \quad (80)$$

where for the last inequality we used $\|E_{BC}^l\| \leq \|E_{BC}\| + \|E_{BC}^l - E_{BC}\| \leq 2C_1$, which holds for sufficiently large l . We can now write

$$\left| \text{Tr}[N_A^*(\rho - \frac{Z}{Z_{AB}Z_C} \rho) E_{BC}^l] \right| \leq \left| \text{Tr}[N_A^* \rho] - \frac{Z}{Z_{AB}Z_C} \text{Tr}[N_A^* \rho] \text{Tr}[\rho E_{BC}^l] \right| + 2KC_1 e^{-\frac{|B|}{2\xi}} \quad (81)$$

$$\leq \left(1 - \frac{Z}{Z_{AB}Z_C} \text{Tr}[\rho E_{BC}^l] \right) + 2KC_1 e^{-\frac{|B|}{2\xi}}, \quad (82)$$

where we used the triangle inequality in the first line, and Hölder's inequality $\text{Tr}[N_A^* \rho] \leq \|N_A^*\| \leq 1$ to get to the second. Finally, we can use Eq. (75) again after another application of Hölder's inequality

$$|\text{Tr}[\rho E_{BC}^l] - \text{Tr}[\rho E_{BC}]| \leq \|E_{BC} - E_{BC}^l\| \leq C_2 \frac{q^{1+l}}{(1+l)!}, \quad (83)$$

and since $\text{Tr}[\rho E_{BC}] = \frac{Z}{Z_{AB}Z_C} \leq e^{\beta \|H_{BC}\|}$ we obtain

$$\left| \text{Tr}[N_A^*(\rho - \frac{Z}{Z_{AB}Z_C} \rho E_{BC}^l)] \right| \leq C_2 e^{\beta \|H_{BC}\|} \frac{q^{1+l}}{(1+l)!} + 2KC_1 e^{-\frac{|B|}{2\xi}}. \quad (84)$$

This finishes the proof. Putting everything together, we see that we have upper-bounded our target quantity in Eq. (73) by a small number related to the error term in the decay of correlations and Araki's result. Without writing the constants explicitly, and just on the leading exponential error, the final result is stated as

$$\|\text{Tr}_{BC}[\rho] - \text{Tr}_B[\rho_0^{AB}]\|_1 \leq e^{-\Omega(|B|)}, \quad (85)$$

where $\Omega(x)$ is defined in Sec. II D.

For simplicity, we have only dealt with the case of a 1D chain, where A is at the end of it. To generalize the proof, one just needs to define an analogous partition ABC in higher dimensions (see Fig. 3 for 2D) and then remove all the different interaction terms from H_{BC} one by one. Here, we have done it with the operator E_{BC} , but this can also be done with the (suitably defined) QBP operator O_A from Sec. III C, and the result Eq. (72) is basically unchanged.

B. Hamiltonian of mean force

We have seen that the marginal on A is close to that of a smaller thermal state of the same Hamiltonian. Inspired by the non-interacting case, we can also think of a different potential feature of ρ_A : is it the thermal state of *some* Hamiltonian on A , different from H_A ? This is obviously the case, since we can always define

$$\tilde{H}_A \equiv \beta^{-1} \log \text{Tr}_{\setminus A}[e^{-\beta H}]. \quad (86)$$

This is the so-called Hamiltonian of mean force [83]. The relevant non-trivial question is: how does this Hamiltonian compare to the “bare” Hamiltonian H_A , which disregards the interactions of A with the rest of the system? Also, is the resulting \tilde{H}_A local in some way? That is, we would like to understand the norm and locality of the operator $\Phi_A \equiv \tilde{H}_A - H_A$.

This turns out to be a difficult problem, which is related to the quantum Markov property and the decay of the conditional mutual information from Sec. IV C. We now briefly describe a known result for high temperatures from [74], whose proof involves involves the connected cluster expansion applied to Eq. (86).

Since the interactions are local, it makes sense that, if the size of A is much larger than the number of nearest neighbours k , most of the weight of Φ_A is localized around its boundary with the rest of the system, of size $|\partial A|$. The precise question is: can we approximate Φ_A with another operator Φ_A^l that only has support on sites a distance l away from the boundary? Theorem 2 in [74] shows that, for any temperature β above a threshold one $\beta^* > 0$, one can define a Φ_A^l such that

$$\|\Phi_A - \Phi_A^l\| \leq \frac{e}{4\beta} \frac{(\beta/\beta^*)^{l/k}}{1 - \frac{\beta}{\beta^*}} |\partial A|. \quad (87)$$

That is, Φ_A can be exponentially well approximated with an operator localized around the boundary. See Fig. 6 for an illustration.

In many of the previously discussed results, such as the decay of correlations, we saw that they hold either in one dimension or at high temperatures. A similar result is expected to hold in 1D, but this is a so far open problem. See [84] for a recent overview on this topic for a different set of models, and its implications.

VI. STATISTICAL PROPERTIES

We now explain and prove some important statistical features of thermal states. These are central statements of the field of statistical physics and characterize the ensembles involved: the thermal (or canonical) and the microcanonical, as well as the grand canonical or others, when relevant. In contrast to the results of other sections, all those shown here (as well as their proofs) apply equally to classical models.

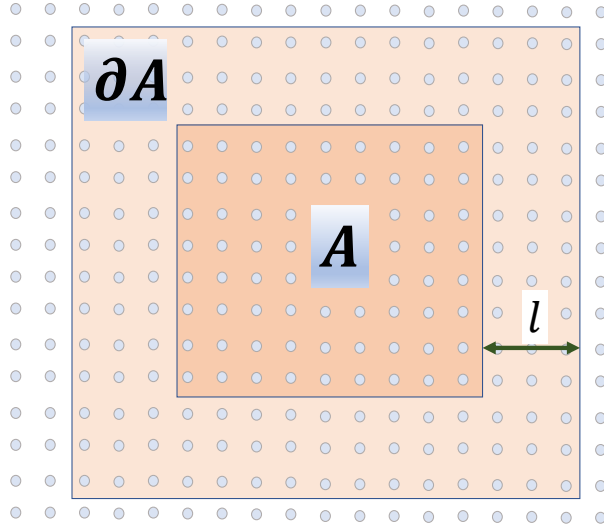


Figure 6: Illustration of the regions of the Hamiltonian of mean force. The correction term Φ_A is exponentially well approximated by $\Phi_{A,l}$, which has support on the region ∂A only.

A. Jaynes' maximum entropy principle

A well-known property that uniquely characterizes thermal states is the so-called maximum entropy principle. This specifies that of all the states with a given energy (or the expectation value of some other quantity) they are the state of largest possible entropy. To see this, let us choose $\rho \neq \rho_\beta$ such that $\text{Tr}[\rho H] = \text{Tr}[\rho_\beta H]$. Then,

$$S(\rho_\beta) - S(\rho) = \text{Tr}[\rho \log \rho] + \beta \text{Tr}[\rho_\beta H] + \log Z \quad (88)$$

$$= \text{Tr}[\rho \log \rho] + \beta \text{Tr}[\rho H] + \log Z \quad (89)$$

$$= \text{Tr}[\rho \log \rho] - \text{Tr}[\rho \log \rho_\beta] \quad (90)$$

$$= D(\rho || \rho_\beta) > 0. \quad (91)$$

Notice that these steps are unchanged if instead of considering just the Hamiltonian H we take into account a higher number of charges Q_i with their chemical potentials μ_i , and the state $\exp(-\sum_j \mu_j Q_j) / \text{Tr}[\exp(-\sum_j \mu_j Q_j)]$.

This simple principle is often interpreted as follows: if there is some state of which we only have partial information (in this case, its average energy), it is very often a good guess to assume it is the thermal state of that energy. Since it is the state with maximum entropy (which we can associate with “maximum ignorance”), its choice makes the fewest assumptions about the structure of the actual state at hand. This idea is often applied in fields like statistical inference and optimization problems in which one might only, including certain quantum algorithms [4–7]. It can also be seen as a variational definition that uniquely singles out thermal states. This allows for the application of this principle in different types of algorithms for finding or characterizing them [12, 85].

B. Measurement statistics and concentration bounds

In Sec. IV we saw how in many instances of thermal states, the different subsystems do not necessarily have strong correlations. This has a number of consequences, and we now explore an important one that shows that their large-scale statistical properties resemble those of non-interacting/statistically independent systems. These are *concentration bounds*, akin to the (perhaps more widely known) central limit theorem. The setting is as follows: let us consider a k -local observable $A = \sum_j A_j$, such that A_j has support on at most k sites. The best example is certainly the energy, but also other properties like magnetization $\sum_j \sigma_j^Z$.

The expectation value of any such observable can be thought of as a macroscopic property of the system (such as the average magnetization of the material). While we expect that there will be thermal fluctuations around that average value, our intuition from thermodynamics tells us that any such large-scale property should have a definite value, almost free of fluctuations. This is due to one of the most basic ideas from probability theory: the measurement statistics of sums of independent random variables greatly concentrate around the average. The main conclusion is that if we measure an observable A on a thermal state, the outcome will be very close to the average $\langle A \rangle_\beta$ with overwhelmingly large probability. That is, the distribution

$$P_{A,\beta}(x) = \text{Tr}[\rho \delta(x - A)]. \quad (92)$$

which is the probability of obtaining outcome x when measuring A , is highly peaked around the average $\langle A \rangle_\beta = \text{Tr}[\rho_\beta A]$. This has important implications for the validity of thermodynamic descriptions of these systems, in that averaged macroscopic quantities characterize the large system of many particles whose properties we do not know with any certainty.

In the theory of probability, there are various types of concentration bounds. Their proofs most often involve constraining the characteristic function $\langle e^{\lambda A} \rangle_\beta$, where λ may be real or imaginary. An important one that we now describe is the Chernoff-Hoeffding inequality. In this case it reads

$$P_{A,\beta}(|x - \langle A \rangle_\beta| > \delta) \leq 2 \exp\left(-\frac{\delta^2}{4c\bar{A}}\right), \quad (93)$$

where $\bar{A} \equiv \sum_j \|A_j\|$. This then says that if $\delta^2 \gg c\bar{A}$, the probability of measuring A to be away from $\langle A \rangle_\beta$ by at least δ is exceedingly small. The most common proof technique is via a bound on the characteristic function of the form

$$\log \langle e^{\tau(A - \langle A \rangle_\beta)} \rangle_\beta \leq c\tau^2 \bar{A}, \quad (94)$$

for some constant c . From this it follows that

$$P_{A,\beta}(x - \langle A \rangle_\beta > \delta) = \int_{x - \langle A \rangle_\beta > \delta} \text{Tr}[\rho \delta(x - A)] = \int_{x - \langle A \rangle_\beta > \delta} \text{Tr}[\rho e^{\tau(A - \langle A \rangle_\beta)} e^{-\tau(A - \langle A \rangle_\beta)} \delta(x - A)] \quad (95)$$

$$\leq e^{-\tau\delta} \text{Tr}[\rho_\beta e^{\tau(A - \langle A \rangle_\beta)}] \leq \exp(-\tau\delta + c\tau^2 \bar{A}). \quad (96)$$

One can follow the same steps for the range $\langle A \rangle_\beta - x > \delta$. Then, choosing $\tau = \delta/(2c\bar{A})$ yields Eq. (93).

This result can be very easily shown for independent random variables or for independent spins. For interacting spins, Eq. (94) was shown in [24] with the cluster expansion technique from Sec. III A, which holds for all dimensions and all temperatures $\beta < \beta^*$. The main result from [70] proves a slightly weaker version of Eq. (93) with a different technique, only assuming the decay of correlations from Sec. IV B.

A related important type of concentration bound is given by large deviation theory. This is the branch of probability theory concerned with understanding the likelihood of very rare events, and has a long history as one of the most important mathematical frameworks for studying statistical physics. For instance it gives a way of describing the equilibrium properties of large ensembles (as is also the case here), or for predicting the long-time behaviour of non-equilibrium processes such as Brownian motion. See [86] for an excellent overview of the main results and their consequences for classical systems.

The basic idea is that given any set of measurement outcomes \mathcal{A} , we would like to identify whether there always exists a *rate function* $I_{\mathcal{A}}$ such that

$$\lim_{N \rightarrow \infty} -\frac{\log P_{A,\beta}(x \in \mathcal{A})}{N} = I_{\mathcal{A}}. \quad (97)$$

If this is the case, the dominant behaviour of $P_{A,\beta}(x \in \mathcal{A})$ is essentially a decaying exponential $P_{A,\beta}(x \in \mathcal{A}) \simeq e^{-NI_{\mathcal{A}} + o(N)}$, unless $I_{\mathcal{A}} = 0$. This means that, in the thermodynamic limit, the measurement statistics of A are extremely peaked around the points where the rate function vanishes $I_{\mathcal{A}} = 0$.

This is slightly stronger than the Chernoff-Hoeffding inequality, in that it can in principle give an exact expression of the probability distribution for large enough N . However, we do not always know how large an N is “enough”, and for finite N , it often does not give an expression as explicit as Eq. (93).

Again, the proof strategy most often involves the characteristic function. In particular, the Gärtner-Ellis theorem states that a sufficient condition is that the function

$$g(\tau) = \lim_{N \rightarrow \infty} \frac{\log \langle e^{\tau A} \rangle_{\beta}}{N} \quad (98)$$

exists and is differentiable. This has been shown using the cluster expansion in [87] for 1-local observables, and upper bounds on the rate for general observables have been shown using the locality estimates from Sec. III B in [50, 88]. The full large deviation principle was shown in 1D in [89]. An alternative proof can be found in [90].

There are various other interesting statements coming from probability theory that apply in this type of setting. Another interesting one is the Berry-Esseen theorem [69], which can be thought of as a refinement of the central limit theorem for finite sample size (which in this case is the system size N). The proof, which only assumes the decay of correlation property, can be found in [91].

C. Equivalence of ensembles

We now prove an important statement in the study of statistical physics, which goes back all the way to Boltzmann and Gibbs. In large systems, the average macroscopic properties of both the thermal or canonical state, and of the microcanonical ensemble, are essentially the same. This means that both canonical and ergodic averages coincide in the thermodynamic limit, and shows that the particular ensemble used for calculations does not matter much.

There are various similar statements in the literature [24, 69, 71, 72, 92–94], but the proof that we now show follows that of [24, 71, 72] and relies on the concentration results from the previous section. Let us define the extensive observable $A = \sum_j A_j$ (such as e.g. the total magnetization $\sum_j^N \sigma_j^Z$) with thermal/canonical average $\langle A \rangle_{\beta}$ which for simplicity we will set to $\langle A \rangle_{\beta} = 0$, while

the microcanonical average is

$$\langle A \rangle_{E,\Delta} = \frac{1}{D_N(E, \Delta)} \sum_{E_j \in (E-\Delta, E)} \langle E_j | A | E_j \rangle, \quad (99)$$

where E is the energy and Δ the width of the microcanonical window (which might depend on N), and $|E_j\rangle$ is the energy eigenstate of energy E_j . $D_N(E, \Delta)$ is a normalization constant counting the number of eigenstates within the window. This motivates the following probability distribution

$$P_{E,\Delta}(x) = \frac{1}{D_N(E, \Delta)} \sum_{E_j \in (E-\Delta, E)} \delta(x - \langle E_j | A | E_j \rangle), \quad (100)$$

which gives the probability of measuring $x = \langle E_j | A | E_j \rangle$ when sampling eigenstates from the microcanonical ensemble.

First, we need to determine what is the energy that corresponds to temperature β and thus characterizes the microcanonical ensemble. Given the temperature β , the microcanonical energy E_0 is such that

$$E_0(\Delta, \beta) \equiv \operatorname{argmax}_E D_N(E, \Delta) e^{-\beta E}. \quad (101)$$

Assuming that the width is significantly different than the energy scales of the system, $\Delta \ll \langle H \rangle_\beta$ (which is most typically the case), this roughly implies that E_0 is the energy of the microstates $\{|E_j\rangle\}$ that have the dominant weight in the canonical ensemble (when the density of states is weighted by the factor $e^{-\beta E}$). We have written the dependence on β, Δ explicitly in Eq. (101), but let us now drop them for simplicity of notation.

We start by upper bounding the m -th (even) moment of $P_{U,\Delta}(x)$

$$\int_{-\infty}^{\infty} x^m P_{E_0,\Delta}(x) = \frac{1}{D_N(E_0, \Delta)} \sum_{E_j \in (E_0-\Delta, E_0)} |\langle E_j | A | E_j \rangle|^m \quad (102)$$

$$\leq \frac{1}{D_N(E_0, \Delta)} \sum_{E_j \in (E_0-\Delta, E_0)} |\langle E_j | A^m | E_j \rangle| = \langle A^m \rangle_{E_0,\Delta}, \quad (103)$$

where we used the convexity of x^m with m even. The bound can easily be expressed in terms of a canonical average as, since A^m is positive,

$$\langle A^m \rangle_{E_0,\Delta} = \frac{1}{D_N(E_0, \Delta)} \sum_{E_j \in (E_0-\Delta, E_0)} \langle E_j | A^m | E_j \rangle \leq \frac{e^{\beta E_0}}{D_N(E_0, \Delta)} \sum_{E_j \in (E_0-\Delta, E_0)} e^{-\beta E_j} \langle E_j | A^m | E_j \rangle \quad (104)$$

$$\leq \frac{e^{\beta E_0}}{D_N(E_0, \Delta)} \sum_{E_j \in (-\infty, \infty)} e^{-\beta E_j} \langle E_j | A^m | E_j \rangle = \frac{Z e^{\beta E_0}}{D_N(E_0, \Delta)} \langle A^m \rangle_\beta. \quad (105)$$

The factor $\frac{Z e^{\beta E_0}}{D_N(E_0, \Delta)}$ can now be upper bounded using the definition of the microcanonical ensemble and the concentration bound. Let us define the following modified partition function $\tilde{Z} \equiv \sum_{|E_j - E_0| \leq \delta} e^{-\beta E_j}$. If we also set $\delta = KN^{1/2}$ with $K = \mathcal{O}(1)$ it follows from Eq. (93) that

$$\frac{\tilde{Z}}{Z} = 1 - P_{H,\beta}(|x - \langle H \rangle_\beta| \geq \delta) \geq 1 - 2 \exp\left(-\frac{\delta^2}{4cJN}\right) \geq 1/2. \quad (106)$$

Now divide the energy range in the sum in equal parts of width $\Delta^* \equiv \min\{\Delta, \beta^{-1}\}$, such that the largest energy of each interval is E_ν , so that $E_{\nu+1} = E_\nu + \Delta^*$ and

$$\tilde{Z} \leq \sum_{\substack{\nu \in \mathbb{Z} \\ |E_\nu - E_0| \leq \Delta^* + \delta}} D_N(E_\nu, \Delta^*) e^{-\beta(E_\nu - \Delta^*)} \leq e^{\beta\Delta^*} \left(\frac{2\delta}{\Delta^*} + 2 \right) \max_\nu D_N(E_\nu, \Delta^*) e^{-\beta E_\nu} \quad (107)$$

$$\leq \frac{1}{2} K' \frac{N^{1/2}}{\Delta^*} D_N(E_0, \Delta) e^{-\beta E_0}, \quad (108)$$

with $K' = \mathcal{O}(1)$, where the last inequality follows from the fact that $D_N(E_0, \Delta)$ is monotonic on Δ . We thus have $\int_{-\infty}^{\infty} x^m P_{E_0, \Delta}(x) \leq K' \frac{N^{1/2}}{\Delta^*} \langle A^m \rangle_\beta$. To finish this part of the proof we thus bound $\langle A^m \rangle_\beta$. It was shown in [24, 72] that the concentration inequality Eq. (93) implies that

$$\langle A^m \rangle_\beta \leq (4c\bar{A})^{m/2} \left(\frac{m}{2} \right)!. \quad (109)$$

For completeness, we reproduce the proof in Appendix A.

We are now in a position to bound the tail of $P_{E, \Delta}(x)$ as

$$P_{E, \Delta}(x \geq x_0) = \int_{x_0}^{\infty} P_{E, \Delta}(x) dx \leq \frac{1}{x_0^m} \int_{-\infty}^{\infty} x^m P_{E_0, \Delta}(x) \quad (110)$$

$$\leq K' \frac{N^{1/2}}{\Delta^*} \left(\frac{4c\bar{A}}{x_0^2} \right)^{m/2} \left(\frac{m}{2} \right)! \leq K' \frac{N^{1/2}}{\Delta^*} \left(\frac{4mc\bar{A}}{x_0^2} \right)^{m/2}. \quad (111)$$

Thus, choosing $m = \lfloor \frac{x_0^2}{4ce\bar{A}} \rfloor$, and repeating for $x \leq -x_0$, leads to (let us now bring back the average $\langle A \rangle_\beta$ explicitly, previously taken to be zero)

$$P_{E, \Delta}(|x - \langle A \rangle_\beta| \geq x_0) \leq 2eK' \frac{N^{1/2}}{\Delta^*} \exp\left(-\frac{x_0^2}{8ce\bar{A}}\right). \quad (112)$$

We are almost done. We now bound the difference between canonical and microcanonical as

$$|\langle A \rangle_{E, \Delta} - \langle A \rangle_\beta| \leq \sum_j \frac{|\langle E_j | A | E_j \rangle - \langle A \rangle_\beta|}{D_{E_0, \Delta}} \leq \int_{|x| \leq \bar{A}} P_{E, \Delta}(x - \langle A \rangle_\beta)(x - \langle A \rangle_\beta) dx \quad (113)$$

$$\leq x_0 + 2\bar{A} P_{E, \Delta}(|x - \langle A \rangle_\beta| \geq x_0), \quad (114)$$

and so choosing $x_0 = \sqrt{8ce\bar{A}} \log(4\bar{A}eK' \frac{N^{1/2}}{\Delta^*})$, the fact that $\bar{A} \propto N$ yields, for some constant K'' ,

$$\frac{1}{N} |\langle A \rangle_{E, \Delta} - \langle A \rangle_\beta| \leq \frac{K'' \log \frac{N^{3/2}}{\Delta^*}}{N^{1/2}}, \quad (115)$$

so that the difference vanishes in the thermodynamic limit. Notice that $\Delta^* \equiv \min\{\Delta, \beta^{-1}\}$, so that in principle even rather low temperatures and very small (up to exponentially small) microcanonical windows are allowed. This is the final result. It states that average properties are essentially the same, provided the condition Eq. (101) holds, and that the width Δ is not too small. The fact that it can be up to exponentially small in system size is rather strong, and related to weak statements of the eigenstate thermalization hypothesis (see [72, 95]).

VII. ALGORITHMS AND COMPLEXITY OF THERMAL STATES

When addressing specific problems in many-body physics, we would most often like to understand whether they are fundamentally complex or not, in the precise sense established by theoretical computer science. This can typically be done in two complementary ways:

- By showing that there exists an algorithm with a provable performance and run-time. Additionally, it is interesting if the algorithm can be explicitly constructed, and implemented in practice.
- By establishing that a problem, or a set of them, belong to or are complete or hard for a certain complexity class.

This applies to both classical and quantum computation, and their respective complexity classes.

Problems related to quantum thermal states can also be studied under this light. The relevant ones include most notably the estimation of the partition function, or the generation of either approximations to the thermal states (in quantum computers) or their classical representations (in classical computers).

As an illustrative example of what can be proven, we start with a simple explicit algorithm that approximates the quantum partition functions in 1D in polynomial time [96]. We then briefly review some other important known results about the hardness of approximating partition functions. The rest of the section includes an explanation of the current best tensor network results, which are provably efficient in a wide range of situations, and a short review of quantum algorithms for preparing thermal states.

A. An efficient classical algorithm for the 1D partition function

Using some of the results from the previous sections, we now show that, assuming that $\beta = \mathcal{O}(1)$, we can efficiently approximate the partition function in 1D. This is done with an algorithm with runtime $\text{poly}(N, \varepsilon^{-1})$ that outputs Z' , where

$$|\log Z' - \log Z| \leq \mathcal{O}(\varepsilon). \quad (116)$$

This section follows the result and proof strategy from [96], with some minor modifications.

In one dimension, let us consider the partial Hamiltonian $H_j = \sum_{i=1}^{j-1} h_i$, which includes the first $j - 1$ interaction terms as counted from the left, starting from the leftmost h_1 . Then, define the partial partition function

$$Z_i = \text{Tr}[e^{-\beta(H_i+h_i)}] = \text{Tr}[O_{h_i} e^{-\beta(H_i)} O_{h_i}^\dagger] \equiv \text{Tr}[e^{-\beta(H_i)} A_i], \quad (117)$$

where O_{h_i} is the quantum belief propagation from Sec. III C and $A_i = O_{h_i}^\dagger O_{h_i}$. Now, rewriting Eq. (117) notice the simple iterative relation

$$Z_i = Z_{i-1} \text{Tr}[\rho_i A_i], \quad (118)$$

where $\rho_i = e^{-\beta H_i} / Z_{i-1}$. Thus we can write

$$Z = d^N \prod_{i=1}^{|E|} \text{Tr}[\rho_i A_i], \quad (119)$$

where $Z \equiv Z_{|E|}$ and $d^N = Z_0$. The key now is to use results from Sec. III C to approximate A_i , and local indistinguishability from Sec. V A 1. Let $A_i^l \equiv (O_{h_i}^l)^\dagger O_{h_i}^l$, so that

$$\|A_i - A_i^l\| = \|A_i - O_{h_i}^\dagger O_{h_i}^l + O_{h_i}^\dagger O_{h_i}^l - A_i^l\| \quad (120)$$

$$\leq 2\|O_{h_i}\| \|O_{h_i} - O_{h_i}^l\| \quad (121)$$

$$\leq e^{\mathcal{O}(\beta h)} e^{-\frac{c'l}{1+c'v\frac{\beta}{\pi}}}, \quad (122)$$

where in the first line we used the triangle inequality and in the second we used both Eq. (48) and (50). Now, let us label by Λ_{l^*} to be the rightmost region of the chain of length l^* in which H_{i+1} has support. Choose $l^* \in \mathbb{R}$ so that $A_i^{l^*}$ has support in the right side of Λ_{2l^*} and define $\rho_i^{(l^*)} = e^{-\beta H_{\Lambda_{l^*}}} / \text{Tr} e^{-\beta H_{\Lambda_{l^*}}}$, where $H_{\Lambda_{l^*}} = \sum_{\text{supp}(h_i) \in \Lambda_{l^*}} h_i$.

The expectation value can be approximated as

$$\left| \text{Tr}[\rho_i A_i] - \text{Tr}[\rho_i^{(2l^*)} A_i^{l^*}] \right| \leq \left| \text{Tr}[\rho_i A_i] - \text{Tr}[\rho_i A_i^{l^*}] \right| + \left| \text{Tr}[\rho_i A_i^{l^*}] - \text{Tr}[\rho_i^{(2l^*)} A_i^{l^*}] \right| \quad (123)$$

$$\leq \|A_i - A_i^{l^*}\| + \|A_i^{l^*}\| \times \|\text{Tr}_{\setminus \Lambda_{l^*}}[\rho_i] - \text{Tr}_{\setminus \Lambda_{2l^*} \setminus \Lambda_{l^*}}[\rho_i^{(2l^*)}]\|_1. \quad (124)$$

This follows from the triangle inequality. The partial trace $\setminus \Lambda_{l^*}$ is over the support of ρ_i excluding Λ_{l^*} . Eq. (124) now has a form that we can upper bound. Since $\|A_i^{l^*}\| \leq \|A_i - A_i^{l^*}\| + \|A_i\| \leq e^{\mathcal{O}(\beta h)}$, we can use Eq. (122) to bound the first term, and Eq. (85) with $|B| = l^*$ to bound the second. With these, we conclude that there exists constants c_1, c_2 depending on all the constants involved (i.e. β, h, J, k, c', v) such that

$$\left| \text{Tr}[\rho_i A_i] - \text{Tr}[\rho_i^{(2l^*)} A_i^{l^*}] \right| \leq c_1 e^{-c_2 l^*}. \quad (125)$$

The key feature of $\text{Tr}[\rho_i^{(2l^*)} A_i^{l^*}]$ is that it is an expectation value of an operator whose form we known explicitly, as per Eq. (49), evaluated in a thermal state of size $2l^*$. This can be computed exactly (or rather, with a subleading error) in a time $\exp(\mathcal{O}(l^*))$. Let us now choose a precision ε/N in Eq. (125), so that $l^* = \mathcal{O}(\log N/\varepsilon^{-1})$. This way, we have

$$Z' \equiv d^N \prod_{i=1}^{|E|} \text{Tr}[\rho_i^{(2l^*)} A_i^{l^*}] = Z \prod_{i=1}^{|E|} \left(1 + \frac{\varepsilon}{N \text{Tr}[\rho_i A_i]} \right) = Z (1 + \mathcal{O}(\varepsilon)). \quad (126)$$

The last equation comes from the fact that $N \propto |E|$ and that all eigenvalues of A_i are $\mathcal{O}(1)$, as per the definition in Eq. (47). The algorithm thus consists of exactly calculating the numbers $\{\text{Tr}[\rho_i^{(2l^*)} A_i^{l^*}]\}$ exactly, and then multiplying them, so that

$$|\log Z' - \log Z| \leq \mathcal{O}(\varepsilon). \quad (127)$$

Since there are $|E| \propto N$ terms in Z' , and each takes time $\text{poly}(N \times \varepsilon^{-1})$, the final runtime is $\text{poly}(N, \varepsilon^{-1})$, as desired.

B. Hardness of approximating partition functions

In the previous section we have seen how the partition function can be approximated in 1D in the sense of Eq. (116) as long as the temperature is $\beta = \mathcal{O}(1)$. Moreover, through the cluster expansion we briefly explained in Sec. III A how it can be approximated for any local model as long as $\beta < \beta^*$, where β^* is some fixed constant.

On the other hand, in the limit of $\beta \rightarrow \infty$, the log-partition function equals the energy of the ground state. For classical models, approximating this to a certain precision is an NP-complete problem. For local quantum Hamiltonians, it is QMA hard. This means that there should be no efficient classical or quantum algorithm to approximate log-partition functions for low enough temperatures, both for classical and quantum models. In fact, it is known that the classical problem is only slightly harder than NP [97] [98], and that it is at least $\#P$ hard if complex interactions are allowed [99]. For the quantum case, the exact complexity class to which this belongs or is complete for is not yet clear (see [100] for more details and results).

There is, however, the expectation that for certain classes of interesting models we can still compute the partition function efficiently, even with classical algorithms and at very low temperatures. Beyond the cluster expansion, a prominent technique are Quantum Monte Carlo methods, which are however restricted to Hamiltonians without the so-called “sign problem” (often referred to as *stoquastic* [101]). These allow for a sampling procedure that approximates the log-partition function. In some restricted cases this also comes with guarantees of fast convergence [102–104].

Other results on particular classes of models recently studied are e.g. those of [100] a class of dense Hamiltonians was shown to have an efficient algorithm, while in [29] a class of quantum models close to classical ones was shown to have a convergent cluster expansion even at low temperatures. There also exists quantum algorithms for approximating general partition functions [100, 105] in the sense of Eq. (116), with exponential run-time $\sim \sqrt{d^N}/Z$, as well as other related algorithms with a slightly weaker notion of approximation [106, 107].

Another relevant angle of this problem, explored in [44] and a number of works in the classical literature, is the direct connection of the (classical) hardness of approximating the partition function with the physics of phase transitions. The idea is based on an important result for the classical Ising model [68], and can be summarized as: a physical phase transition in the system may come together with a computational phase transition in which approximating $\log Z$ becomes fundamentally harder. It has been shown that in quite a general setting [44], the analyticity of the log-partition function implies the existence of an efficient algorithm, akin to what was explained in Sec. III A. There are important open questions, however. For instance we may expect that away from a phase transition, whenever the correlations decay exponentially, the partition function will be analytic and can be approximated efficiently.

C. Tensor network methods

Tensor network (TN) techniques are perhaps the most successful way of classically computing physical properties of quantum systems in 1D, and sometimes 2D. This success not only comes from the amount of numerical results obtained with them through the years, but also from the fact that we have theoretical guarantees for the performance of many TN algorithms. This includes most notably the regime of low energy physics [46, 58, 108–110] and, as we now review, that of finite temperature too. A detailed explanation of TN is beyond the scope of these notes, so we encourage the reader to first check the numerous introductory reviews on the subject e.g. [111–114].

The aim is to obtain a TN representation M_D of the object $e^{-\beta H}$ such that $\|e^{-\beta H} - M_D\|_1 \leq \varepsilon Z$, which then allows us to compute all thermal expectation values up to error ε as per Eq. (5). The index D labels the bond dimension which, roughly speaking, quantifies the complexity of representing M_D . A TN of bond dimension D requires a memory $\propto n \times D^2$ to be stored. In a nutshell, the approximation operator M_D should be made out of a sum or low-depth product of operators with smaller support i.e. of size at most $\propto \log D$. This is graphically described for 1D in Fig. 7.

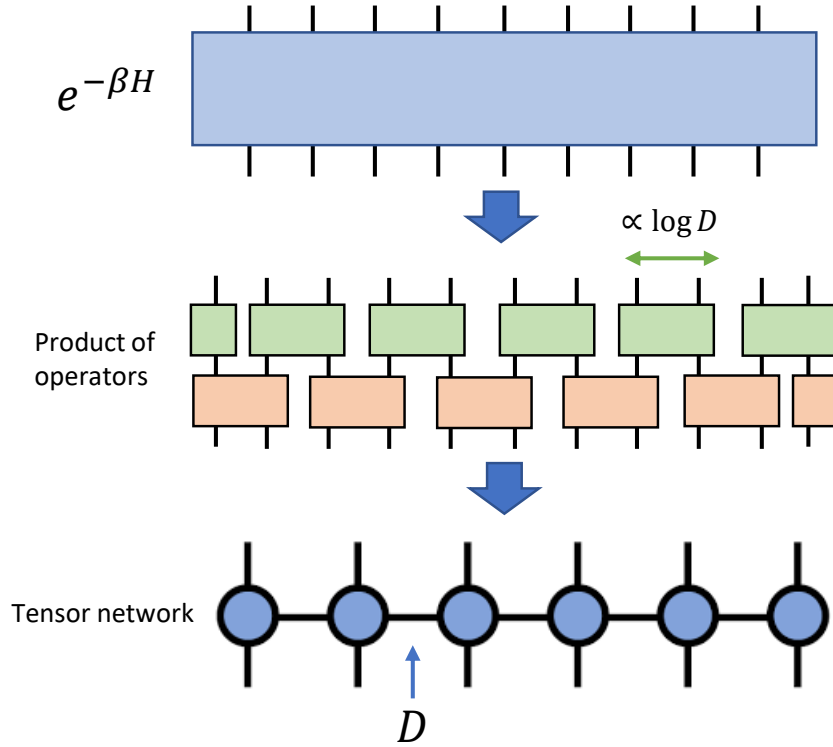


Figure 7: Schematically, the way to prove that a 1D thermal state is a tensor network is by decomposing it as a product of smaller operators. It then follows from standard methods that the bond dimension of the tensor network representation is related to the size of those operators.

Let us first briefly cover the best currently known result for 1D chains [37], in which the approximation is with the so-called Matrix Product Operators (MPO) [115]. The algorithm from the previous section gives us a hint of how this can be done: adding piece by piece from right to left, aided by the results from Sec. III B. We thus have to decompose $e^{-\beta H}$ as a product of smaller operators, in a way that the locality estimates results can be leveraged. To do this, first define $\Psi_j = e^{\beta H_j} e^{-\beta H_{j+1}}$. We then have that $e^{-\beta H} = e^{-\beta h_1} \prod_{j=1} \Psi_j$. Each operator Ψ_j can then be approximated by a localized operator $\tilde{\Psi}_j^l$ with support in a region of length l as in Eq. (41), exponentially well in l . There are $\propto N$ of those operators, and the error of each approximation can be shown to contribute additively. Thus, choosing $l \propto \log N/\epsilon$ gives the desired ϵ -good approximation to $e^{-\beta H}$.

The bond dimension can be straightforwardly assumed to be $D \leq e^{\mathcal{O}(l)} = \text{poly}(N, \epsilon^{-1})$. This is already computationally efficient. However, as shown in [37], one can instead define an operator $\tilde{\Psi}_j^l$ in which the exponential functions are approximated by their Taylor series. In that case, we can put forward results about the bond dimension required to represent polynomials of Hamiltonians [54]. This leads to an improvement of the bond dimension to $D \leq e^{\tilde{\mathcal{O}}(\sqrt{l})} = \exp\left(\tilde{\mathcal{O}}(\sqrt{\log(N/\epsilon)})\right)$. This is sub-linear in system size, much more computationally efficient. This algorithm is in fact not very far from some that have been implemented in practice [116].

In higher dimensions, the best-known method is a variation of the cluster expansion proposed in [117]. There, the expansion is treated in a slightly different way, to approximate the exponential $e^{-\beta H}$ rather than the log-partition function as in Sec. III A. Instead of counting the number of individual clusters of at most size m , one has to consider arbitrary products $\prod_{i \in W} h_i \equiv h(W)$ of

terms h_i from a multiset W of size $|W|$, that can be divided into connected clusters. Let us label the multisets $W = \{h_i\}$ in which the biggest cluster has size M to be C_M . That this is consistent with the cluster expansion can be seen from taking the exponential of Eq. (16) given the expression in terms of clusters of the powers in Eq. (20).

The following result was proven in detail in [65]. It reads

$$\|e^{-\beta H} - \sum_{W \in C_M} \frac{(-\beta)^{|W|}}{|w|!} h(W)\|_1 \leq Z \left(e^{N \frac{b(\beta)^M}{1-b(\beta)}} - 1 \right), \quad (128)$$

where $b(\beta) < 1$ for all $\beta < \beta^* = \mathcal{O}(1)$. In [118] the sum over clusters on the RHS of Eq. (128) was shown to be a tensor network (in fact, a so-called PEPO) of bond dimension $e^{\mathcal{O}(M)}$. Thus, by setting the RHS to be ε , we achieve a TN approximation to $e^{-\beta H}$ with bond dimension $\text{poly}(N, \varepsilon^{-1})$. This only holds for inverse temperatures below β^* , but the result can be extended to arbitrary temperatures simply by taking powers of the operator. This means the bond dimension grows as $D \leq \exp\left(\mathcal{O}\left(\beta \log \frac{\beta N}{\varepsilon}\right)\right)$ (see [118] for the details). This scheme has recently been numerically implemented in practice [119].

These results show that there are in principle efficient TN representations for all dimensions and all temperatures $\beta = \mathcal{O}(1)$. An important caveat is that in dimensions higher than one, a TN representation is not enough to be able to extract numerical data efficiently. This is because the contraction of TN can be a computationally demanding task by itself [120, 121]. Finally, let us note that by using the local indistinguishability from Sec. VA (or even without it in 1D [122]) it can be shown that a much smaller bond dimension is needed to simulate local properties [82].

D. Quantum algorithms for preparing thermal states

One of the most promising applications of quantum computers is the generation of exotic states of matter in complex many-body models. The expectation is that this should allow us to discover a potentially wide variety of physics, and also serve as a subroutine in certain quantum algorithms, such as those performing optimization tasks.

Because of this, a question that has been very much explored lately is that of how to prepare thermal states of local Hamiltonians with a quantum computer. This could be either a fully fledged fault-tolerant one or, by other means more suitable for the so-called NISQ (Noisy Intermediate Scale Quantum) devices, such as variational ones. In the following we review some of the currently existing ones, and also explain the ideas that highlight the complexity of the problem. We mostly focus on those that have some provable performance guarantees (regardless of whether they are efficient or not). There are many others we will not cover (such as e.g. [123–125] and others), many of which often rely on some level of heuristic arguments. These may nonetheless be more efficient in many physically relevant settings.

General considerations

We have very strong evidence pointing that preparing thermal states is, in its most general setting, not an easy task. The results on QMA hardness of the local Hamiltonian problem [13] show that there are vanishingly small temperatures (scaling quickly with system size) at which the preparation of ρ_β is QMA complete. That is, not even a quantum computer can do it efficiently [126]. This is the case even for 1D systems [127].

There are also compelling reasons to believe that this will be the case even at slightly higher temperatures: it has been conjectured that there exists local models for which all states below a

certain energy density cannot be efficiently prepared with a quantum computer [15]. The current best results along these lines are [128, 129], which show that there exist Hamiltonians for which the low energy thermal states (that is, with vanishingly small temperature) have a provably large circuit complexity lower bound. This takes the form of a lower bound in the number of elementary gates required to generate it. Stronger results also hold if one restricts the models and circuits to obey certain symmetries [130]. It is then no surprise that the most general algorithms with a provable performance have a super polynomial (in fact, exponential) circuit complexity, even already at $\beta \sim \mathcal{O}(1)$.

These points only imply that the most general algorithms will not be efficient. However, we expect that large interesting classes of models and settings will be much easier. The locality of the model, and some of its consequences from the previous sections, should simplify this task in many important settings.

Algorithms based on purifications

These algorithms work for general Hamiltonians, and could potentially be run in a fully fault tolerant quantum computer, capable of applying any quantum circuit without large errors. They are based on constructing the following purification

$$|\rho_\beta\rangle = \frac{1}{\sqrt{Z}} \sum_l e^{-\beta E_l/2} |E_l\rangle |l\rangle_{\mathcal{A}}, \quad (129)$$

where the second subsystem \mathcal{A} is made of auxiliary particles such that, upon tracing out, yield $\text{Tr}_{\mathcal{A}}[|\rho_\beta\rangle\langle\rho_\beta|] = \rho_\beta$. The way to do this is by first preparing a state $|\psi\rangle$ with $|\rho_\beta\rangle$ as a component such that

$$|\psi\rangle = \frac{1}{\mathcal{N}} |\rho_\beta\rangle + \dots \quad (130)$$

The main idea is to start with a maximally entangled state between the target system and some auxiliary qubits, and then find a scheme that implements a unitary that acts as $U \propto e^{-\beta/2} + \dots$. This can be done efficiently, for instance, with the phase estimation algorithm [105, 131, 132] or with more recent quantum simulation ideas, such as the technique based on sums of unitaries [133] (which leads to a better error dependence in many cases of interest). To obtain $|\rho_\beta\rangle$ with high precision, one must then apply amplitude amplification of the state $|\psi\rangle$. The gate complexity of this, however, grows with \mathcal{N} . This growth can be up to $\sqrt{d^{\mathcal{N}}/Z}$, which sets the leading (almost) exponential gate cost of the algorithm. In one dimension, one can instead implement this same algorithm connecting subsequent segments of the chain, which can reduce the gate complexity to a polynomial $\sim N^{\mathcal{O}(\beta)}$. Recent progress shows that the phase estimation and amplitude amplification steps in these schemes can instead be replaced by random circuits with post-selection [134], making them more amenable to current technologies. For commuting Hamiltonians, a purification in the form of a tensor network state (a PEPS) can be very efficiently prepared through an adiabatic algorithm [135]. See also the recent [136], which produces a purification of a thermal state $\propto e^{-\beta H_1}$ starting from that of another Hamiltonian H_0 , and is efficient when $\|H_0 - H_1\|$ is not too large.

Quantum Metropolis Sampling

Another class is based on the adaptation of the very well known classical Monte Carlo algorithm of Metropolis sampling [137, 138]. This is based on a stochastic update of the energy

eigenstates at each iteration, where the probability of accepting is related to the Gibbs weight, eventually converging to the thermal distribution. The first quantum version was devised in [139], which involved highly non-local updates that rendered it likely very inefficient. This problem was dealt with in [140], with a scheme based on the phase estimation algorithm to detect the energy changes, plus random local circuits that constituted the stochastic updates. A later version of this algorithm is [141], in which the stochastic updates are accepted or not according to a quantum algorithm speeding up classical Markov processes [142]. A simpler algorithm, with a similar performance, can also be found in [134]. In all these, the convergence to the thermal distribution is guaranteed by the property of detailed balance. However, it is not clear theoretically under what conditions do these converge fast (or slow) to that thermal distribution - it could be that the runtime is exponential, even in reasonably simple instances. A fast runtime is only guaranteed by proving a lower bound on the gap of the associated Markov chain, which is typically a rather hard problem, even for classical Monte Carlo schemes. See [143] for recent progress along these lines in models that obey the Eigenstate Thermalization Hypothesis [144, 145].

Variational algorithms

A currently very thoroughly studied class of mixed classical-quantum algorithms is based on the *variational* principle. The rough idea is as follows. The quantum computer is given some initial instructions through a number of parameters, and outputs a solution to a problem in the form of a quantum state, that can then be evaluated with a “cost function”. A good cost function is minimized at the exact solution, and its magnitude measures the distance to it. The classical part of the algorithm is an optimizer that uses this data to variationally find the set of instructions that minimize the cost function.

The potential advantage of this kind of method is that much of the hard computational work is placed on the classical computer: it has to perform an optimization algorithm, such as gradient descent. The quantum part only needs to execute a potentially small set of operations, such as e.g. alternating dynamics of two Hamiltonians for different times, as in the well-known QAOA algorithm [146]. The efficiency of the scheme thus relies on two crucial facts:

- It is possible to efficiently approximate the cost function with the quantum computer.
- The classical optimizer converges quickly to the global minimum, and e.g. does not get stuck in local minima or very flat regions.

Typically this kind of algorithm is aimed at finding ground states or low energy states of quantum systems, in which case the cost function is simply the energy, which is an expectation value that can in principle be evaluated easily. For thermal states, however, we need a different cost function. This is given by the free energy at inverse temperature β defined as $F_\beta(\rho) = \beta \text{Tr}[H\rho] - S(\rho)$, which, as shown in Eq. (11), is minimized by ρ_β .

An important difficulty is that, due to the logarithm, the free energy is not a linear function of the state, and cannot be computed straightforwardly. This can be dealt with by devising a quantum algorithm able to efficiently approximate the entropy. The most natural way to do this is through a series expansion to approximate the entropy, as is done in [147]. A similar option is to change the cost function altogether to an approximation whose minimum yields an approximation to ρ_β , as in [148].

The second problem, of the efficiency of the classical optimizer, has to do with the smoothness of the cost function and its sensitivity to small changes. The computational cost of computing these gradients have been studied [147], but there are so far no rigorous theoretical guarantees of

the cases in which such algorithm converges quickly. There is the possibility that one may often encounter the problem of “barren plateaus” in the optimization, which may render it inefficient [149]. However, it may well be that the structure put into the problem by the locality of the Hamiltonian may often be enough to circumvent this issue [150].

Efficient algorithms from physical features

Perhaps the main caveat of most of the aforementioned algorithms is that they are constructed for very general Hamiltonians. Thus, at least a priori, they do not make a very clear use of the physical features that we expect could simplify the problem, such as locality or any one of its consequences. It should be possible, however, to have provably more efficient algorithms in which relevant physical properties appear.

This is the case for the algorithm in [43], whose efficiency depends on two such factors: the speed of decay of CMI from Sec. IV C, and the error in the local indistinguishability from Sec. V A. The algorithm uses iterations of the recovery map that appeared in Eq. (68), which are guaranteed to yield a low error if the CMI decays quickly enough. The main idea is that one can construct local decoupled parts of the thermal state independently, and then join them together to make up the whole ρ_β via subsequent applications of the recovery map. The local indistinguishability guarantees that the local parts used in the recovery are also accurate parts of the whole thermal state the algorithm constructs. The results on the exponential decay of correlations from Sec. IV B and of exponential decay in CMI from Sec. IV C thus guarantee that there exists efficient algorithms for 1D systems and for local models at a high enough temperature $\beta \leq \beta^*$. A potential issue with this algorithm is that it requires the implementation of the recovery map, for which we do not always have explicit expressions [80]. It is encouraging, however, that quantum algorithms for one such recovery map, the Petz map, have appeared in the literature [151].

An alternative route along these lines is to find out under which conditions the dissipative dynamics (that is, when the system is coupled weakly to some external bath) associated to a Gibbs state converge quickly. Then, tools to engineer dissipative dynamics can be in principle implemented in a quantum computer [152, 153]. The challenge is to find under which conditions these dynamics have a fast convergence or mixing rate. There are some exceptions [143, 154], but rigorous results along these lines are so far mostly limited to commuting Hamiltonians, as we explain in Sec. VIII.

VIII. COMMUTING HAMILTONIANS

There is a much simpler and yet physically relevant class of Hamiltonians that merits pointing them out specifically: those in which all the $\{h_i\}$ commute with each other. This includes many interesting models for quantum many-body physics and quantum computation. Perhaps most notably, it includes all stabilizer Hamiltonians, including the toric code and other widely studied examples, as well as many other models describing various topological phases of matter.

Notice that these are not the same as classical Hamiltonians: even if we can diagonalize all the h_i simultaneously, the energy eigenbasis will in general be highly entangled. In contrast, classical Hamiltonians have a product eigenbasis. At the same time, we have

$$e^{-\beta(H-h_i)} = e^{-\beta H} e^{\beta h_i} = e^{\beta h_i/2} e^{-\beta H} e^{\beta h_i/2}, \quad (131)$$

so the tools in Sec. III B and III C are unnecessary. This means that many of the results described above take much simpler forms and easier proofs, as we now briefly explain.

Let us divide the lattice into two complementary regions D, E , with boundary ∂_{DE} , so that $H = H_D + H_E + H_I$, with $\text{supp}(H_I) \in \partial_{DE}$. Notice that

$$\text{Tr}_E[e^{-\beta H}] = e^{-\beta H_D} \text{Tr}_E[e^{-\beta(H_E+H_I)}]. \quad (132)$$

Clearly $\text{Tr}_E[e^{-\beta(H_B+H_I)}]$ has support on the region $D \cap \partial_{DE}$ only. This means that the local indistinguishability from [V A](#) holds with no error by choosing $A = D, B = E \cap \partial_{DE}, C = E \setminus B$, so that $\text{dist}(A, C)$ is roughly the width of the boundary. A similar exact result applies to the Hamiltonian of mean force. We now briefly show the proof, which is elementary and can be found in [\[155\]](#). If we define $e^{-\beta\Phi} \equiv \text{Tr}_E[e^{-\beta(H_E+H_I)}]$, we see that

$$\frac{-1}{\beta} \log(\text{Tr}_E[e^{-\beta H}]) = \alpha \mathbb{I} + H_D + \Phi, \quad (133)$$

where α is some constant, and Φ is localized in $D \cup \partial_{DE}$ and has bounded norm, as

$$H_D + H_E - h|\partial_{DE}| \leq H \leq H_D + H_E + h|\partial_{DE}| \quad (134)$$

implies that

$$e^{-\beta h|\partial_{DE}|} e^{-\beta(H_D+H_E)} \leq e^{-\beta H} \leq e^{\beta h|\partial_{DE}|} e^{-\beta(H_D+H_E)}, \quad (135)$$

which upon tracing E out and multiplying by $e^{\beta H_D}$, implies that $\|\Phi\| \leq 2h|\partial_{DE}|$.

It should also be no surprise then that the Markov property of [Sec. IV C](#) also holds exactly. This means that if we define regions A, B, C such that A, C are shielded by region B , we have that $I(A : C|B) = 0$ [\[156, 157\]](#). In fact, a converse statement holds (vanishing CMI implies the state is a thermal state of a local Hamiltonian) when the interaction graph Λ is triangle-free [\[158\]](#). As mentioned in [Sec. IV C](#), this is the quantum equivalent of the Hammersley-Clifford theorem [\[81\]](#).

All these exact results strongly suggest that algorithms such as those described in [Sec. VII](#) are much more efficient in this setting. For instance, it is immediate from a repeated application of [Eq. \(131\)](#) that the thermal states can be expressed exactly as tensor networks with constant bond dimension $D \leq e^{\mathcal{O}(k)}$. There also exists quantum algorithms for commuting Hamiltonians that are significantly more efficient than the general ones in [Sec. VII D](#) [\[135\]](#). Also, further concentration inequalities akin to [Eq. \(93\)](#) have been shown specifically for commuting Hamiltonians [\[159\]](#).

Due to their additional simplicity, there are also a number of ideas that have so far only been proven for these Hamiltonians. A noteworthy example are results on dissipative evolutions that map any given initial state to the thermal state. These processes are essentially Markovian thermalizations described by Lindblad equations, which model the interaction of a large quantum state with an external heat bath with which the coupling is weak, such that

$$\frac{d\rho}{dt} = \mathcal{L}(\rho) = -i[H, \rho] + \sum_{\alpha} L_{\alpha} \rho L_{\alpha}^{\dagger} - \frac{1}{2} \left\{ L_{\alpha} L_{\alpha}^{\dagger}, \rho \right\}, \quad (136)$$

where L_{α} are the ‘‘jump’’ operators and α indexes the energy gaps of H . The interesting cases are those for which ρ_{β} is the unique fixed point, or the unique state such that $\mathcal{L}(\rho_{\beta}) = 0$. The best known example are the Davies generators [\[160\]](#). See e.g. [\[161, 162\]](#) for introductory references on these equations and their derivation.

When the Hamiltonian is commuting, the individual operators L_{α} can in fact be taken to be local, and one can ask: what is the time it takes for such local dissipative evolution $e^{t\mathcal{L}}(\rho)$ to approach the Gibbs state? In particular, how does this time depend on system size? This can be tackled with ideas that previously appeared in the classical literature, in particular, the analysis

of the spectral gap and the log-Sobolev constant of \mathcal{L} . A bound on the spectral gap was proven assuming the decay of correlations in [163], and for specific models in [164–167]. This shows that it takes a polynomial time to thermalize. On the other hand, a bound on the so-called *log-Sobolev constant* [168] implies that instead it takes a time *logarithmic* in the system size to thermalize. This was recently proven for 1D chains [169, 170] and in [171, 172] for other models of dissipation. These results show that the dissipative processes at hand can also be seen as an efficient quantum algorithm preparing thermal states, since in principle they can be simulated efficiently with a quantum computer [152].

IX. CONCLUSIONS AND OPEN QUESTIONS

It may appear at first that studying thermal states of general complex quantum models is a very challenging task. We hope to have illustrated the fact that this is not always the case: for a large array of situations involving local Hamiltonians many non-trivial analytical statements can be made. These are both about universal physical features of the models at hand, but also about the computational complexity of the problems the physics poses. The connections found motivate a timely research program, largely inspired by quantum information theory: to understand the links between fundamental physical features and their computational complexity.

In the present context, much of the technical difficulty lies in working with the matrix exponential of any such a Hamiltonian, in which typically the individual terms do not commute. As seen in Sec. III, however, we have a number of mathematical tools to deal with these in many physically relevant regimes.

We have covered a number of statements in different areas and summarized many of the existing results on the topic. However, plenty of relevant questions are still open. We now summarize some of them, which we believe to be of particular physical or technical interest:

- The cluster expansion has been used to prove statements at high temperatures such as the exponential decay of the CMI, or the locality of the Hamiltonian of mean force. Is there an analogous technique that holds for 1D models at all temperatures, that allows us to prove those statements?
- There is a well developed mathematical theory narrowing down general conditions under which the cluster expansion converges. This follows from mathematical physics works such as [27]. This has a number of important consequences regarding decay of correlations [66], approximations to the partition function [22] or quantum algorithms [74]. It would be interesting to have a more detailed physical understanding of when this convergence happens, and narrow it down in as many regimes and classes of models as possible.
- We have only focused on local Hamiltonians with short-range interactions. A number of recent studies have developed the theory of Lieb-Robinson bounds for systems with long-range interactions, decaying with some power of the distance in the lattice [173–176]. The cluster expansion results of [24, 74] also hold in this regime. It would be interesting to study how this and other discussions of Sec. III change in this setting, and what do these imply for the thermal properties of long-range interacting systems.
- In Sec. IV we did not differentiate between classical and quantum correlations, which is likely an important distinction in this context. Can we make general statements about how quantum correlations (as measured by standard entanglement quantifiers) are more fragile at finite temperatures than classical ones? An important result in this direction is that of

[177], showing a decay of a certain kind of quantum correlations at all temperatures, based on Lieb-Robinson bounds. Related to this is the area law for the entanglement negativity shown in [61], using inequalities akin to those of Sec. III D.

- It has been conjectured [42, 43] that the current known bounds on the conditional mutual information are not optimal, and that this quantity does in fact decay exponentially at all temperatures. This has only so far been shown at high temperatures, when the cluster expansion converges [74]. We currently lack enough numerical or analytical evidence hinting towards this decay. However, if proven true, it would have interesting implications for various classical and quantum algorithms at low temperatures.
- The ideas of Sec. V, and in particular the Hamiltonian of mean force, have in the past few years features in the study of thermodynamic quantities for strongly coupled systems [83, 178–181]. Many existing results on this topic focus on simpler models than those considered here, such as individual spins coupled to quadratic baths [84, 182]. It would be interesting to understand whether the results from Sec. V have non-trivial consequences for the equilibrium and non-equilibrium thermodynamics of strongly coupled spin systems, such as those found in [183].
- The current theoretical results for tensor network descriptions of thermal states give reasonably good bounds on the bond dimension [37, 65, 118]. It would be interesting, however, to know whether the tensor networks in dimensions higher than one [65, 118] can be contracted efficiently to calculate expectation values, perhaps in cases for which the Gibbs state has exponential clustering of correlations (see [184] for a similar idea for ground states). A complementary relevant question is to fully understand whether the set of density operators represented by tensor networks in 1D is contained within the set of thermal states of local Hamiltonians [185].
- With the advent of quantum computing, there are multiple ongoing efforts aiming to find more efficient quantum algorithms for thermal sampling and partition functions. As we have seen in sec. VII D, many of the existing ones are designed for very general situations, and as such have performance bounds that will often be too conservative. Some existing schemes do make use of relevant physical features to simplify them [43, 132, 135, 143], but it seems that there is still plenty of room for exploring the kinds of regimes in which explicit and efficient algorithms can be proven. Since preparing thermal states is presumably an easier task than a general quantum computation (at least in certain regimes), it may be possible to tailor them to the limited capabilities of near-term noisy devices [134].
- Recently a number of works have narrowed down the sample and computational complexity of the problem of thermal state tomography [12, 23, 155, 186]. The basic question is: can we learn the Hamiltonian from a small number of simple (local) measurements of few copies of $e^{-\beta H}/Z$? Optimal sample and computational complexity bounds exists in the high temperature regime, in which the cluster expansion applies [23], but beyond that our theoretical understanding is not complete (for instance, in 1D). This problem has a number of applications, including the verification of quantum computation in which thermal sampling is involved [4–7], or the characterization of many-body entanglement [187, 188].

ACKNOWLEDGEMENTS

AMA would like to thank the organizers of the Quantum Thermodynamics Summer School 2021, Nuriya Nurgalieva and Lidia del Rio, since it triggered the idea of these notes. AMA also emphatically thanks the many colleagues and collaborators from which I have learnt about these topics through the years. In particular, I thank Anurag Anshu and Tomotaka Kuwahara for useful comments and discussions. AMA acknowledges support from the Alexander von Humboldt Foundation.

Appendix A: Miscellaneous proofs

Locality of operator E_A

In Sec. III B we defined the operator

$$E_A = e^{-\beta(H+A)}e^{\beta H} = \mathcal{T}e^{-\int_0^\beta ds e^{-sH} A e^{sH}}, \quad (\text{A1})$$

which is the solution of the differential equation

$$\frac{dE_A}{d\beta} = -E_A A(i\beta), \quad (\text{A2})$$

with $A(i\beta) = e^{-\beta H} A e^{\beta H}$. We can also define the localized generator

$$A^l(i\beta) = \sum_{m=0}^l \beta^m C_m(A), \quad (\text{A3})$$

and also the corresponding operator $E_A(l)$ as the solution of

$$\frac{dE_A(l)}{d\beta} = -E_A(l) A^l(i\beta). \quad (\text{A4})$$

Now from the Trotter-Suzuki decomposition

$$E_A = \lim_{L \rightarrow \infty} \prod_{j=0}^{L-1} e^{-A(i\frac{\beta j}{L})\frac{\beta}{L}} \quad (\text{A5})$$

$$E_A(l) = \lim_{L \rightarrow \infty} \prod_{j=0}^{L-1} e^{-A^l(i\frac{\beta j}{L})\frac{\beta}{L}}, \quad (\text{A6})$$

we have that

$$E_A - E_A(l) = \lim_{L \rightarrow \infty} \sum_{j=0}^{L-1} \left(A^l(i\frac{\beta j}{L})\frac{\beta}{L} - A(i\frac{\beta j}{L})\frac{\beta}{L} \right) = \int_0^\beta A^l(is) - A(is) ds, \quad (\text{A7})$$

so by the triangle inequality and Eq. (30),

$$\|E_A - E_A(l)\| \leq \int_0^\beta \|A^l(is) - A(is)\| ds \leq \beta k \|A\| \frac{(2\beta Jk)^{l+1}}{1 - 2\beta Jk}. \quad (\text{A8})$$

Proof of Eq. (53)

This can also be found in [50]. Let $F(t)$ be a differentiable and bounded operator. DuHamel's identity states that

$$\frac{d}{dt}e^{F(t)} = \int_0^1 du e^{uF(t)} \frac{dF(t)}{dt} e^{(1-u)F(t)}. \quad (\text{A9})$$

Then we have that

$$\frac{d}{dt} \log \text{Tr} (C e^{H_1+tH_2}) = \frac{\text{Tr} \left(\int_0^1 du C e^{u(H_1+tH_2)} H_2 e^{(1-u)(H_1+tH_2)} \right)}{\text{Tr} (C e^{H_1+tH_2})} \quad (\text{A10})$$

$$= \frac{\text{Tr} \left(e^{\frac{H_1+tH_2}{2}} C e^{\frac{H_1+tH_2}{2}} \int_0^1 du e^{(u-1/2)(H_1+tH_2)} H_2 e^{(1/2-u)(H_1+tH_2)} \right)}{\text{Tr} \left(e^{\frac{H_1+tH_2}{2}} C e^{\frac{H_1+tH_2}{2}} \right)} \quad (\text{A11})$$

$$\leq \left\| \int_0^1 du e^{(u-1/2)(H_1+tH_2)} H_2 e^{(1/2-u)(H_1+tH_2)} \right\|. \quad (\text{A12})$$

This follows from Hölder's inequality Eq. (6) and the positivity of C . Finally,

$$|\log \text{Tr}[C e^{H_1+H_2}] - \log \text{Tr}[C e^{H_1}]| = \left| \int_0^1 \frac{d}{dt} \log \text{Tr} [C e^{H_1+tH_2}] dt \right| \quad (\text{A13})$$

$$\leq \int_0^1 dt \int_{-1/2}^{1/2} ds \|e^{s(H_1+tH_2)} H_2 e^{-s(H_1+tH_2)}\|, \quad (\text{A14})$$

where the last step follows from the triangle inequality, Eq. (A12) and the change of variable $u - 1/2 = s$.

Proof of Eq. (109)

This can also be found in [72]. Let $p(x)$ be an arbitrary probability distribution with $\int_{-\infty}^{\infty} xp(x)dx = a$, and the condition that $p(x)$ be Lebesgue integrable. We aim to bound

$$\int_{-\infty}^{\infty} |x - a|^m p(x) dx = \int_{-\infty}^{\infty} |x|^m p(x + a) dx = \int_0^{\infty} |x|^m (p(x + a) + p(-x + a)) dx \quad (\text{A15})$$

$$= - \int_0^{\infty} x^m \frac{d}{dx} \left[\int_{|x'-a| \geq x} p(x') dx \right] dx, \quad (\text{A16})$$

where in the last step we used the fundamental theorem of calculus. This can now be integrated by parts as

$$- \int_0^{\infty} x^m \frac{d}{dx} \left[\int_{|x'-a| \geq x} p(x') dx \right] dx \quad (\text{A17})$$

$$\begin{aligned} &= - \left(x^m \int_{|x'-a| \geq x} p(x') dx \right) \Big|_0^{\infty} + \int_0^{\infty} mx^{m-1} \int_{|x'-a| \geq x} p(x') dx' dx \\ &\leq \int_0^{\infty} mx^{m-1} 2e^{-\frac{x^2}{4cA}} dx = (4c\bar{A})^{m/2} \left(\frac{m}{2}\right)!, \end{aligned} \quad (\text{A18})$$

where in the second line the first term vanishes by definition, and in the third line we used the concentration bound Eq. (93).

-
- [1] Christian Gogolin and Jens Eisert, “Equilibration, thermalisation, and the emergence of statistical mechanics in closed quantum systems,” *Reports on Progress in Physics* **79**, 056001 (2016).
 - [2] Luca D’Alessio, Yariv Kafri, Anatoli Polkovnikov, and Marcos Rigol, “From quantum chaos and eigenstate thermalization to statistical mechanics and thermodynamics,” *Advances in Physics* **65**, 239–362 (2016).
 - [3] 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,” *Physical Review Letters* **101** (2008), 10.1103/physrevlett.101.010504.
 - [4] Fernando G.S.L. Brandão and Krysta M. Svore, “Quantum speed-ups for solving semidefinite programs,” in *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)* (2017) pp. 415–426.
 - [5] 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)*, Leibniz International Proceedings in Informatics (LIPIcs), Vol. 132, edited by Christel Baier, Ioannis Chatzigiannakis, Paola Flocchini, and Stefano Leonardi (Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2019) pp. 27:1–27:14.
 - [6] Joran van Apeldoorn, András Gilyén, Sander Gribling, and Ronald de Wolf, “Quantum SDP-Solvers: Better upper and lower bounds,” *Quantum* **4**, 230 (2020).
 - [7] Fernando G.S.L. Brandão, Richard Kueng, and Daniel Stilck França, “Faster quantum and classical SDP approximations for quadratic binary optimization,” *Quantum* **6**, 625 (2022).
 - [8] Mohammad H. Amin, Evgeny Andriyash, Jason Rolfe, Bohdan Kulchytskyy, and Roger Melko, “Quantum boltzmann machine,” *Phys. Rev. X* **8**, 021050 (2018).
 - [9] Mária Kieferová and Nathan Wiebe, “Tomography and generative training with quantum boltzmann machines,” *Phys. Rev. A* **96**, 062327 (2017).
 - [10] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd, “Quantum machine learning,” *Nature* **549**, 195–202 (2017).
 - [11] Eyal Bairey, Itai Arad, and Netanel H. Lindner, “Learning a local hamiltonian from local measurements,” *Physical Review Letters* **122** (2019), 10.1103/physrevlett.122.020504.
 - [12] Anurag Anshu, Srinivasan Arunachalam, Tomotaka Kuwahara, and Mehdi Soleimanifar, “Sample-efficient learning of interacting quantum systems,” *Nature Physics* **17**, 931–935 (2021).
 - [13] Julia Kempe, Alexei Kitaev, and Oded Regev, “The complexity of the local hamiltonian problem,” in *FSTTCS 2004: Foundations of Software Technology and Theoretical Computer Science*, edited by Kamal Lodaya and Meena Mahajan (Springer Berlin Heidelberg, Berlin, Heidelberg, 2005) pp. 372–383.
 - [14] Dorit Aharonov, Itai Arad, and Thomas Vidick, “Guest column: The quantum pcp conjecture,” *SIGACT News* **44**, 47–79 (2013).
 - [15] Matthew B. Hastings, “Trivial low energy states for commuting hamiltonians, and the quantum pcp conjecture,” *Quantum Info. Comput.* **13**, 393–429 (2013).
 - [16] Martin Kliesch and Arnau Riera, “Properties of thermal quantum states: Locality of temperature, decay of correlations, and more,” in *Thermodynamics in the Quantum Regime: Fundamental Aspects and New Directions*, edited by Felix Binder, Luis A. Correa, Christian Gogolin, Janet Anders, and Gerardo Adesso (Springer International Publishing, Cham, 2018) pp. 481–502.
 - [17] The logs here are defined with base e .
 - [18] Sumeet Khatri and Mark M. Wilde, “Principles of quantum communication theory: A modern approach,” (2020), [arXiv:2011.04672 \[quant-ph\]](https://arxiv.org/abs/2011.04672).
 - [19] Samuel O. Scalet, Álvaro M. Alhambra, Georgios Styliaris, and J. Ignacio Cirac, “Computable rényi mutual information: Area laws and correlations,” *Quantum* **5**, 541 (2021).
 - [20] Mark M. Wilde, *Quantum Information Theory*, 2nd ed. (Cambridge University Press, 2017).
 - [21] David Ruelle, *Statistical mechanics: Rigorous results* (Benjamin, 1969).

- [22] Ryan L. Mann and Tyler Helmuth, “Efficient algorithms for approximating quantum partition functions,” *Journal of Mathematical Physics* **62**, 022201 (2021).
- [23] Jeongwan Haah, Robin Kothari, and Ewin Tang, “Optimal learning of quantum hamiltonians from high-temperature gibbs states,” arXiv preprint arXiv:2108.04842 (2021), [arXiv:2108.04842 \[quant-ph\]](#).
- [24] 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).
- [25] V A Malyshev, “Cluster expansions in lattice models of statistical physics and the quantum theory of fields,” *Russian Mathematical Surveys* **35**, 1–62 (1980).
- [26] Roland L Dobrushin, “Estimates of semi-invariants for the ising model at low temperatures,” *Translations of the American Mathematical Society-Series 2* **177**, 59–82 (1996).
- [27] R. Kotecký and D. Preiss, “Cluster expansion for abstract polymer models,” *Communications in Mathematical Physics* **103**, 491 – 498 (1986).
- [28] Tyler Helmuth, Will Perkins, and Guus Regts, “Algorithmic pirogov–sinai theory,” *Probability Theory and Related Fields* **176**, 851–895 (2020).
- [29] Tyler Helmuth and Ryan L. Mann, “Efficient algorithms for approximating quantum partition functions at low temperature,” (2022), [arXiv:2201.06533 \[quant-ph\]](#).
- [30] Elliott H. Lieb and Derek W. Robinson, “The finite group velocity of quantum spin systems,” in *Statistical Mechanics: Selecta of Elliott H. Lieb*, edited by Bruno Nachtergaele, Jan Philip Solovej, and Jakob Yngvason (Springer Berlin Heidelberg, Berlin, Heidelberg, 2004) pp. 425–431.
- [31] Tomotaka Kuwahara, Takashi Mori, and Keiji Saito, “Floquet – magnus theory and generic transient dynamics in periodically driven many-body quantum systems,” *Ann. Phys.* **367**, 96–124 (2016).
- [32] Gabriel Bouch, “Complex-time singularity and locality estimates for quantum lattice systems,” *J. Math. Phys.* **56**, 123303 (2015).
- [33] Huzihiro Araki, “Gibbs states of a one dimensional quantum lattice,” *Commun. Math. Phys.* **14**, 120–157 (1969).
- [34] David Pérez-García and Antonio Pérez-Hernández, “Locality estimates for complex time evolution in 1d,” (2020), [arXiv:2004.10516 \[math-ph\]](#).
- [35] The numerical constants in these functions are likely sub-optimal, but this is very rarely important for applications.
- [36] Alexander Avdoshkin and Anatoly Dymarsky, “Euclidean operator growth and quantum chaos,” *Physical Review Research* **2** (2020), [10.1103/physrevresearch.2.043234](#).
- [37] Tomotaka Kuwahara, Álvaro M. Alhambra, and Anurag Anshu, “Improved thermal area law and quasilinear time algorithm for quantum gibbs states,” *Phys. Rev. X* **11**, 011047 (2021).
- [38] Andrew M. Childs, Yuan Su, Minh C. Tran, Nathan Wiebe, and Shuchen Zhu, “Theory of trotter error with commutator scaling,” *Physical Review X* **11** (2021), [10.1103/physrevx.11.011020](#).
- [39] Seth Lloyd, “Universal quantum simulators,” *Science* **273**, 1073–1078 (1996), <https://www.science.org/doi/pdf/10.1126/science.273.5278.1073>.
- [40] Matthew B. Hastings, “Quantum belief propagation: An algorithm for thermal quantum systems,” *Physical Review B* **76** (2007), [10.1103/physrevb.76.201102](#).
- [41] Isaac H. Kim, “Perturbative analysis of topological entanglement entropy from conditional independence,” *Physical Review B* **86** (2012), [10.1103/physrevb.86.245116](#).
- [42] Kohtaro Kato and Fernando G. S. L. Brandão, “Quantum approximate markov chains are thermal,” *Communications in Mathematical Physics* **370**, 117–149 (2019).
- [43] Fernando G. S. L. Brandão and Michael J. Kastoryano, “Finite correlation length implies efficient preparation of quantum thermal states,” *Communications in Mathematical Physics* **365**, 1–16 (2019).
- [44] Aram W. Harrow, Saeed Mehraban, and Mehdi Soleimanifar, “Classical algorithms, correlation decay, and complex zeros of partition functions of quantum many-body systems,” *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing* (2020), [10.1145/3357713.3384322](#).
- [45] Matthew B. Hastings and Tohru Koma, “Spectral gap and exponential decay of correlations,” *Communications in Mathematical Physics* **265**, 781–804 (2006).
- [46] Matthew B. Hastings, “An area law for one-dimensional quantum systems,” *Journal of Statistical Mechanics: Theory and Experiment* **2007**, P08024–P08024 (2007).
- [47] Matthew B. Hastings, “Locality in quantum systems,” (2010), [arXiv:1008.5137 \[math-ph\]](#).

- [48] Matthew B. Hastings, “Gapped quantum systems: From higher dimensional lieb-schultz-mattis to the quantum hall effect,” (2021).
- [49] S. Ejima and Y. Ogata, “Perturbation theory of kms states,” *Annales Henri Poincaré* **20**, 2971–2986 (2019).
- [50] Marco Lenci and Luc Rey-Bellet, “Large deviations in quantum lattice systems: One-phase region,” *Journal of Statistical Physics* **119**, 715–746 (2005).
- [51] Rajendra Bhatia, *Matrix analysis*, Vol. 169 (Springer Science & Business Media, 2013).
- [52] Eric Carlen, “Trace inequalities and quantum entropy: an introductory course,” *Entropy and the quantum* **529**, 73–140 (2010).
- [53] J. Eisert, M. Cramer, and M. B. Plenio, “Colloquium: Area laws for the entanglement entropy,” *Reviews of Modern Physics* **82**, 277–306 (2010).
- [54] Itai Arad, Alexei Kitaev, Zeph Landau, and Umesh Vazirani, “An area law and sub-exponential algorithm for 1d systems,” (2013), [arXiv:1301.1162 \[quant-ph\]](https://arxiv.org/abs/1301.1162).
- [55] Anurag Anshu, Itai Arad, and David Gosset, “An area law for 2d frustration-free spin systems,” (2021), [arXiv:2103.02492 \[quant-ph\]](https://arxiv.org/abs/2103.02492).
- [56] G. Vidal, J. I. Latorre, E. Rico, and A. Kitaev, “Entanglement in quantum critical phenomena,” *Physical Review Letters* **90** (2003), [10.1103/physrevlett.90.227902](https://doi.org/10.1103/physrevlett.90.227902).
- [57] Fernando G. S. L. Brandão and Michal Horodecki, “Exponential decay of correlations implies area law,” *Communications in Mathematical Physics* **333**, 761–798 (2014).
- [58] F. Verstraete and J. I. Cirac, “Matrix product states represent ground states faithfully,” *Phys. Rev. B* **73**, 094423 (2006).
- [59] Yimin Ge and Jens Eisert, “Area laws and efficient descriptions of quantum many-body states,” *New Journal of Physics* **18**, 083026 (2016).
- [60] Michael M. Wolf, Frank Verstraete, Matthew B. Hastings, and J. Ignacio Cirac, “Area laws in quantum systems: Mutual information and correlations,” *Phys. Rev. Lett.* **100**, 070502 (2008).
- [61] Nicholas E. Sherman, Trithep Devakul, Matthew B. Hastings, and Rajiv R. P. Singh, “Nonzero-temperature entanglement negativity of quantum spin models: Area law, linked cluster expansions, and sudden death,” *Phys. Rev. E* **93**, 022128 (2016).
- [62] Daniel Gottesman and Matthew B. Hastings, “Entanglement versus gap for one-dimensional spin systems,” *New Journal of Physics* **12**, 025002 (2010).
- [63] Marko Znidaric, Tomaz Prosen, and Iztok Pizorn, “Complexity of thermal states in quantum spin chains,” *Physical Review A* **78** (2008), [10.1103/physreva.78.022103](https://doi.org/10.1103/physreva.78.022103).
- [64] H Bernigau, M J Kastoryano, and J Eisert, “Mutual information area laws for thermal free fermions,” *J. Stat. Mech.* **2015**, P02008 (2015).
- [65] M. Kliesch, D. Gross, and J. Eisert, “Matrix-product operators and states: NP-hardness and undecidability,” *Phys. Rev. Lett.* **113**, 160503 (2014).
- [66] Jürg Fröhlich and Daniel Ueltschi, “Some properties of correlations of quantum lattice systems in thermal equilibrium,” *Journal of Mathematical Physics* **56**, 053302 (2015), <https://doi.org/10.1063/1.4921305>.
- [67] Andreas Bluhm, Â ngela Capel, and Antonio Pérez-Hernández, “Exponential decay of mutual information for gibbs states of local hamiltonians,” *Quantum* **6**, 650 (2022).
- [68] R. L. Dobrushin and S. B. Shlosman, “Completely analytical interactions: Constructive description,” *Journal of Statistical Physics* **46**, 983–1014 (1987).
- [69] Fernando G. S. L. Brandão and Marcus Cramer, “Equivalence of statistical mechanical ensembles for non-critical quantum systems,” (2015), [arXiv:1502.03263 \[quant-ph\]](https://arxiv.org/abs/1502.03263).
- [70] Anurag Anshu, “Concentration bounds for quantum states with finite correlation length on quantum spin lattice systems,” *New Journal of Physics* **18**, 083011 (2016).
- [71] Hal Tasaki, “On the local equivalence between the canonical and the microcanonical ensembles for quantum spin systems,” *Journal of Statistical Physics* **172**, 905–926 (2018).
- [72] Tomotaka Kuwahara and Keiji Saito, “Eigenstate thermalization from the clustering property of correlation,” *Physical Review Letters* **124** (2020), [10.1103/physrevlett.124.200604](https://doi.org/10.1103/physrevlett.124.200604).
- [73] Patrick Hayden, Richard Jozsa, Dénes Petz, and Andreas Winter, “Structure of states which satisfy strong subadditivity of quantum entropy with equality,” *Communications in Mathematical Physics* **246**, 359–374 (2004).

- [74] 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** (2020), 10.1103/physrevlett.124.220601.
- [75] Isaac H. Kim, “Markovian matrix product density operators : Efficient computation of global entropy,” (2017), [arXiv:1709.07828 \[quant-ph\]](https://arxiv.org/abs/1709.07828).
- [76] Dénes Petz, “Sufficient subalgebras and the relative entropy of states of a von neumann algebra,” *Communications in mathematical physics* **105**, 123–131 (1986).
- [77] Patrick Hayden, Richard Jozsa, Dénes Petz, and Andreas Winter, “Structure of states which satisfy strong subadditivity of quantum entropy with equality,” *Communications in Mathematical Physics* **246**, 359–374 (2004).
- [78] Omar Fawzi and Renato Renner, “Quantum conditional mutual information and approximate markov chains,” *Communications in Mathematical Physics* **340**, 575–611 (2015).
- [79] Marius Junge, Renato Renner, David Sutter, Mark M. Wilde, and Andreas Winter, “Universal recovery maps and approximate sufficiency of quantum relative entropy,” *Annales Henri Poincaré* **19**, 2955–2978 (2018).
- [80] David Sutter, “Approximate quantum markov chains,” *SpringerBriefs in Mathematical Physics* , 75–100 (2018).
- [81] John M Hammersley and Peter Clifford, “Markov fields on finite graphs and lattices,” <http://www.statslab.cam.ac.uk/~grg/books/hammfest/hamm-cliff.pdf> (1971).
- [82] Álvaro M. Alhambra and J. Ignacio Cirac, “Locally accurate tensor networks for thermal states and time evolution,” *PRX Quantum* **2**, 040331 (2021).
- [83] Harry JD Miller, “Hamiltonian of mean force for strongly-coupled systems,” *Thermodynamics in the Quantum Regime: Fundamental Aspects and New Directions* **195**, 531 (2018).
- [84] AS Trushechkin, M Merkli, JD Cresser, and J Anders, “Open quantum system dynamics and the mean force gibbs state,” arXiv preprint arXiv:2110.01671 (2021), [arXiv:2110.01671 \[quant-ph\]](https://arxiv.org/abs/2110.01671).
- [85] Serena Di Giorgio and Paulo Mateus, “On the complexity of finding the maximum entropy compatible quantum state,” *Mathematics* **9**, 193 (2021).
- [86] Hugo Touchette, “The large deviation approach to statistical mechanics,” *Physics Reports* **478**, 1–69 (2009).
- [87] K. Netočný and F. Redig, “Large deviations for quantum spin systems,” *Journal of Statistical Physics* **117**, 521–547 (2004).
- [88] Fumio Hiai, Milan Mosonyi, and Tomohiro Ogawa, “Large deviations and chernoff bound for certain correlated states on a spin chain,” *Journal of Mathematical Physics* **48**, 123301 (2007), <https://doi.org/10.1063/1.2812417>.
- [89] Yoshiko Ogata, “Large deviations in quantum spin chains,” *Communications in Mathematical Physics* **296**, 35–68 (2010).
- [90] Yoshiko Ogata and Luc Rey-Bellet, “Ruelle-lanford functions and large deviations for asymptotically decoupled quantum systems,” *Reviews in Mathematical Physics* **23**, 211–232 (2011), <https://doi.org/10.1142/S0129055X11004291>.
- [91] Fernando G. S. L. Brandão, Marcus Cramer, and Mădălin Guță, “Berry-Esseen theorem for quantum lattice systems and the equivalence of statistical mechanical ensembles,” <http://quantum-lab.org/qjp2015/talks/125-Brandao.pdf> (2015).
- [92] R. Lima, “Equivalence of ensembles in quantum lattice systems,” *Annales I.H.P. Physique théorique* **15**, 61–68 (1971).
- [93] R. Lima, “Equivalence of ensembles in quantum lattice systems: States,” *Communications in Mathematical Physics* **24**, 180 – 192 (1972).
- [94] Markus P. Müller, Emily Adlam, Lluís Masanes, and Nathan Wiebe, “Thermalization and canonical typicality in translation-invariant quantum lattice systems,” *Communications in Mathematical Physics* **340**, 499–561 (2015).
- [95] Takashi Mori, “Weak eigenstate thermalization with large deviation bound,” arXiv preprint arXiv:1609.09776 (2016), [arXiv:1609.09776 \[quant-ph\]](https://arxiv.org/abs/1609.09776).
- [96] Tomotaka Kuwahara and Keiji Saito, “Polynomial-time classical simulation for one-dimensional quantum gibbs states,” (2018), [arXiv:1807.08424 \[quant-ph\]](https://arxiv.org/abs/1807.08424).
- [97] Larry Stockmeyer, “The complexity of approximate counting,” in *Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing*, STOC ’83 (Association for Computing Machinery, New

- York, NY, USA, 1983) p. 118–126.
- [98] More specifically, it is in the class BPP^{NP} , see [100].
 - [99] Leslie Ann Goldberg and Heng Guo, “The complexity of approximating complex-valued ising and tutte partition functions,” (2017), [arXiv:1409.5627 \[cs.CC\]](#).
 - [100] Sergey Bravyi, Anirban Chowdhury, David Gosset, and Pawel Wocjan, “On the complexity of quantum partition functions,” (2021), [arXiv:2110.15466 \[quant-ph\]](#).
 - [101] Sergey Bravyi, David P. Divincenzo, Roberto Oliveira, and Barbara M. Terhal, “The complexity of stoquastic local hamiltonian problems,” *Quantum Info. Comput.* **8**, 361–385 (2008).
 - [102] Sergey Bravyi and David Gosset, “Polynomial-time classical simulation of quantum ferromagnets,” *Physical Review Letters* **119** (2017), [10.1103/physrevlett.119.100503](#).
 - [103] Elizabeth Crosson and Aram W. Harrow, “Rapid mixing of path integral monte carlo for 1d stoquastic hamiltonians,” *Quantum* **5**, 395 (2021).
 - [104] Elizabeth Crosson and Samuel Sleazak, “Classical simulation of high temperature quantum ising models,” arXiv preprint [arXiv:2002.02232](#) (2020), [arXiv:2002.02232 \[quant-ph\]](#).
 - [105] David Poulin and Pawel Wocjan, “Sampling from the thermal quantum gibbs state and evaluating partition functions with a quantum computer,” *Physical Review Letters* **103** (2009), [10.1103/physrevlett.103.220502](#).
 - [106] Chris Cade and Ashley Montanaro, “The Quantum Complexity of Computing Schatten p-norms,” in *13th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2018)*, Leibniz International Proceedings in Informatics (LIPIcs), Vol. 111, edited by Stacey Jeffery (Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2018) pp. 4:1–4:20.
 - [107] Anirban N. Chowdhury, Rolando D. Somma, and Yigit Subasi, “Computing partition functions in the one-clean-qubit model,” *Physical Review A* **103** (2021), [10.1103/physreva.103.032422](#).
 - [108] Zeph Landau, Umesh Vazirani, and Thomas Vidick, “A polynomial time algorithm for the ground state of one-dimensional gapped local hamiltonians,” *Nature Physics* **11**, 566–569 (2015).
 - [109] Itai Arad, Zeph Landau, Umesh Vazirani, and Thomas Vidick, “Rigorous rg algorithms and area laws for low energy eigenstates in 1d,” *Communications in Mathematical Physics* **356**, 65–105 (2017).
 - [110] Yichen Huang, “A polynomial-time algorithm for the ground state of one-dimensional gapped hamiltonians,” (2015), [arXiv:1406.6355 \[cond-mat.str-el\]](#).
 - [111] Jacob C Bridgeman and Christopher T Chubb, “Hand-waving and interpretive dance: an introductory course on tensor networks,” *Journal of Physics A: Mathematical and Theoretical* **50**, 223001 (2017).
 - [112] J. Eisert, “Entanglement and tensor network states,” (2013), [arXiv:1308.3318 \[quant-ph\]](#).
 - [113] Román Orús, “Tensor networks for complex quantum systems,” *Nature Reviews Physics* **1**, 538–550 (2019).
 - [114] J. Ignacio Cirac, David Pérez-García, Norbert Schuch, and Frank Verstraete, “Matrix product states and projected entangled pair states: Concepts, symmetries, theorems,” *Reviews of Modern Physics* **93** (2021), [10.1103/revmodphys.93.045003](#).
 - [115] B Pirvu, V Murg, J I Cirac, and F Verstraete, “Matrix product operator representations,” *New Journal of Physics* **12**, 025012 (2010).
 - [116] Bin-Bin Chen, Lei Chen, Ziyu Chen, Wei Li, and Andreas Weichselbaum, “Exponential thermal tensor network approach for quantum lattice models,” *Phys. Rev. X* **8**, 031082 (2018).
 - [117] Matthew B. Hastings, “Solving gapped hamiltonians locally,” *Phys. Rev. B* **73**, 085115 (2006).
 - [118] 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** (2015), [10.1103/physrevb.91.045138](#).
 - [119] Bram Vanhecke, David Devoogdt, Frank Verstraete, and Laurens Vanderstraeten, “Simulating thermal density operators with cluster expansions and tensor networks,” (2021), [arXiv:2112.01507 \[cond-mat.str-el\]](#).
 - [120] Norbert Schuch, Michael M. Wolf, Frank Verstraete, and J. Ignacio Cirac, “Computational complexity of projected entangled pair states,” *Physical Review Letters* **98** (2007), [10.1103/physrevlett.98.140506](#).
 - [121] Jonas Haferkamp, Dominik Hangleiter, Jens Eisert, and Marek Gluza, “Contracting projected entangled pair states is average-case hard,” *Physical Review Research* **2** (2020), [10.1103/physrevresearch.2.013010](#).

- [122] Yichen Huang, “Locally accurate matrix product approximation to thermal states,” *Science Bulletin* **66**, 2456–2457 (2021).
- [123] Mario Motta, Chong Sun, Adrian T. K. Tan, Matthew J. O’Rourke, Erika Ye, Austin J. Minnich, Fernando G. S. L. Brandão, and Garnet Kin-Lic Chan, “Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution,” *Nature Physics* **16**, 205–210 (2020).
- [124] Sirui Lu, Mari Carmen Bañuls, and J. Ignacio Cirac, “Algorithms for quantum simulation at finite energies,” *PRX Quantum* **2** (2021), 10.1103/prxquantum.2.020321.
- [125] J. Cohn, F. Yang, K. Najafi, B. Jones, and J. K. Freericks, “Minimal effective gibbs ansatz: A simple protocol for extracting an accurate thermal representation for quantum simulation,” *Physical Review A* **102** (2020), 10.1103/physreva.102.022622.
- [126] Alexei Yu Kitaev, Alexander Shen, Mikhail N Vyalyi, and Mikhail N Vyalyi, *Classical and quantum computation*, 47 (American Mathematical Soc., 2002).
- [127] Dorit Aharonov, Daniel Gottesman, Sandy Irani, and Julia Kempe, “The power of quantum systems on a line,” *Communications in Mathematical Physics* **287**, 41–65 (2009).
- [128] Lior Eldar, “Robust Quantum Entanglement at (Nearly) Room Temperature,” in *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*, Leibniz International Proceedings in Informatics (LIPIcs), Vol. 185, edited by James R. Lee (Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2021) pp. 49:1–49:20.
- [129] 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)*, Leibniz International Proceedings in Informatics (LIPIcs), Vol. 215, edited by Mark Braverman (Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2022) pp. 6:1–6:22.
- [130] Sergey Bravyi, Alexander Kliesch, Robert Koenig, and Eugene Tang, “Obstacles to variational quantum optimization from symmetry protection,” *Physical Review Letters* **125** (2020), 10.1103/physrevlett.125.260505.
- [131] Chen-Fu Chiang and Pawel Wocjan, “Quantum algorithm for preparing thermal gibbs states—detailed analysis,” in *Quantum Cryptography and Computing* (IOS Press, 2010) pp. 138–147.
- [132] Ersen Bilgin and Sergio Boixo, “Preparing thermal states of quantum systems by dimension reduction,” *Physical Review Letters* **105** (2010), 10.1103/physrevlett.105.170405.
- [133] Anirban Narayan Chowdhury and Rolando D. Somma, “Quantum algorithms for gibbs sampling and hitting-time estimation,” *Quantum Information and Computation* **17** (2017), 10.26421/QIC17.1-2.
- [134] Oles Shtanko and Ramis Movassagh, “Algorithms for gibbs state preparation on noiseless and noisy random quantum circuits,” arXiv preprint arXiv:2112.14688 (2021), arXiv:2112.14688 [quant-ph].
- [135] Yimin Ge, András Molnár, and J. Ignacio Cirac, “Rapid adiabatic preparation of injective projected entangled pair states and gibbs states,” *Physical Review Letters* **116** (2016), 10.1103/physrevlett.116.080503.
- [136] Zoe Holmes, Gopikrishnan Muraleedharan, Rolando D. Somma, Yigit Subasi, and Burak Şahinoğlu, “Quantum algorithms from fluctuation theorems: Thermal-state preparation,” (2022).
- [137] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller, “Equation of state calculations by fast computing machines,” *The Journal of Chemical Physics* **21**, 1087–1092 (1953), <https://doi.org/10.1063/1.1699114>.
- [138] W. K. Hastings, “Monte Carlo sampling methods using Markov chains and their applications,” *Biometrika* **57**, 97–109 (1970), <https://academic.oup.com/biomet/article-pdf/57/1/97/23940249/57-1-97.pdf>.
- [139] Barbara M. Terhal and David P. DiVincenzo, “Problem of equilibration and the computation of correlation functions on a quantum computer,” *Physical Review A* **61** (2000), 10.1103/physreva.61.022301.
- [140] K. Temme, T. J. Osborne, K. G. Vollbrecht, D. Poulin, and F. Verstraete, “Quantum metropolis sampling,” *Nature* **471**, 87–90 (2011).
- [141] Man-Hong Yung and Alán Aspuru-Guzik, “A quantum–quantum metropolis algorithm,” *Proceedings of the National Academy of Sciences* **109**, 754–759 (2012).
- [142] M. Szegedy, “Quantum speed-up of markov chain based algorithms,” in *45th Annual IEEE Symposium on Foundations of Computer Science* (2004) pp. 32–41.
- [143] Chi-Fang Chen and Fernando GSL Brandão, “Fast thermalization from the eigenstate thermalization hypothesis,” arXiv preprint arXiv:2112.07646 (2021), arXiv:2112.07646 [quant-ph].

- [144] J. M. Deutsch, “Quantum statistical mechanics in a closed system,” *Phys. Rev. A* **43**, 2046–2049 (1991).
- [145] Mark Srednicki, “Chaos and quantum thermalization,” *Physical Review E* **50**, 888–901 (1994).
- [146] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann, “A quantum approximate optimization algorithm,” arXiv preprint arXiv:1411.4028 (2014), [arXiv:1411.4028 \[quant-ph\]](https://arxiv.org/abs/1411.4028).
- [147] Anirban N Chowdhury, Guang Hao Low, and Nathan Wiebe, “A variational quantum algorithm for preparing quantum gibbs states,” arXiv preprint arXiv:2002.00055 (2020), [arXiv:2002.00055 \[quant-ph\]](https://arxiv.org/abs/2002.00055).
- [148] Youle Wang, Guangxi Li, and Xin Wang, “Variational quantum gibbs state preparation with a truncated taylor series,” *Phys. Rev. Applied* **16**, 054035 (2021).
- [149] Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven, “Barren plateaus in quantum neural network training landscapes,” *Nature Communications* **9**, 4812 (2018).
- [150] M. Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J. Coles, “Cost function dependent barren plateaus in shallow parametrized quantum circuits,” *Nature Communications* **12**, 1791 (2021).
- [151] András Gilyén, Seth Lloyd, Iman Marvian, Yihui Quek, and Mark M Wilde, “Quantum algorithm for petz recovery channels and pretty good measurements,” arXiv preprint arXiv:2006.16924 (2020), [arXiv:2006.16924 \[quant-ph\]](https://arxiv.org/abs/2006.16924).
- [152] M. Kliesch, T. Barthel, C. Gogolin, M. Kastoryano, and J. Eisert, “Dissipative quantum church-turing theorem,” *Physical Review Letters* **107** (2011), [10.1103/physrevlett.107.120501](https://doi.org/10.1103/physrevlett.107.120501).
- [153] Hong-Yi Su and Ying Li, “Quantum algorithm for the simulation of open-system dynamics and thermalization,” *Phys. Rev. A* **101**, 012328 (2020).
- [154] Kristan Temme, “Lower bounds to the spectral gap of davies generators,” *Journal of Mathematical Physics* **54**, 122110 (2013), <https://doi.org/10.1063/1.4850896>.
- [155] Anurag Anshu, Srinivasan Arunachalam, Tomotaka Kuwahara, and Mehdi Soleimanifar, “Efficient learning of commuting hamiltonians on lattices,” https://people.eecs.berkeley.edu/~anuraganshu/Learning_commuting_hamiltonian.pdf (2021).
- [156] M.S. Leifer and D. Poulin, “Quantum graphical models and belief propagation,” *Annals of Physics* **323**, 1899–1946 (2008).
- [157] David Poulin and Matthew B. Hastings, “Markov entropy decomposition: A variational dual for quantum belief propagation,” *Physical Review Letters* **106** (2011), [10.1103/physrevlett.106.080403](https://doi.org/10.1103/physrevlett.106.080403).
- [158] Winton Brown and David Poulin, “Quantum markov networks and commuting hamiltonians,” (2012), [arXiv:1206.0755 \[quant-ph\]](https://arxiv.org/abs/1206.0755).
- [159] Giacomo De Palma and Cambyse Rouzé, “Quantum concentration inequalities,” arXiv preprint arXiv:2106.15819 (2021), [arXiv:2106.15819 \[quant-ph\]](https://arxiv.org/abs/2106.15819).
- [160] E. B. Davies, “Markovian master equations,” *Communications in Mathematical Physics* **39**, 91–110 (1974).
- [161] Heinz-Peter Breuer, Francesco Petruccione, *et al.*, *The theory of open quantum systems* (Oxford University Press on Demand, 2002).
- [162] Angel Rivas and Susana F Huelga, *Open quantum systems*, Vol. 10 (Springer, 2012).
- [163] Michael J. Kastoryano and Fernando G. S. L. Brandão, “Quantum gibbs samplers: The commuting case,” *Communications in Mathematical Physics* **344**, 915–957 (2016).
- [164] R Alicki, M Fannes, and M Horodecki, “On thermalization in kitaev’s 2d model,” *Journal of Physics A: Mathematical and Theoretical* **42**, 065303 (2009).
- [165] Anna Kómar, Olivier Landon-Cardinal, and Kristan Temme, “Necessity of an energy barrier for self-correction of abelian quantum doubles,” *Physical Review A* **93** (2016), [10.1103/physreva.93.052337](https://doi.org/10.1103/physreva.93.052337).
- [166] Kristan Temme, “Thermalization time bounds for pauli stabilizer hamiltonians,” *Communications in Mathematical Physics* **350**, 603–637 (2017).
- [167] Angelo Lucia, David Pérez-García, and Antonio Pérez-Hernández, “Thermalization in kitaev’s quantum double models via tensor network techniques,” arXiv preprint arXiv:2107.01628 (2021), [arXiv:2107.01628 \[quant-ph\]](https://arxiv.org/abs/2107.01628).
- [168] Michael J. Kastoryano and Kristan Temme, “Quantum logarithmic sobolev inequalities and rapid mixing,” *Journal of Mathematical Physics* **54**, 052202 (2013).
- [169] Ivan Bardet, Ángela Capel, Li Gao, Angelo Lucia, David Pérez-García, and Cambyse Rouzé, “Entropy decay for davies semigroups of a one dimensional quantum lattice,” arXiv preprint arXiv:2112.00601 (2021), [arXiv:2112.00601 \[quant-ph\]](https://arxiv.org/abs/2112.00601).

- [170] Ivan Bardet, Ángela Capel, Li Gao, Angelo Lucia, David Pérez-García, and Cambyse Rouzé, “Rapid thermalization of spin chain commuting hamiltonians,” arXiv preprint arXiv:2112.00593 (2021), [arXiv:2112.00593 \[quant-ph\]](https://arxiv.org/abs/2112.00593).
- [171] Ivan Bardet, Ángela Capel, Angelo Lucia, David Pérez-García, and Cambyse Rouzé, “On the modified logarithmic sobolev inequality for the heat-bath dynamics for 1d systems,” *Journal of Mathematical Physics* **62**, 061901 (2021), <https://doi.org/10.1063/1.5142186>.
- [172] Ángela Capel, Cambyse Rouzé, and Daniel Stilck França, “The modified logarithmic sobolev inequality for quantum spin systems: classical and commuting nearest neighbour interactions,” arXiv preprint arXiv:2009.11817 (2020), [arXiv:2009.11817 \[quant-ph\]](https://arxiv.org/abs/2009.11817).
- [173] Chi-Fang Chen and Andrew Lucas, “Finite speed of quantum scrambling with long range interactions,” *Physical Review Letters* **123** (2019), [10.1103/physrevlett.123.250605](https://doi.org/10.1103/physrevlett.123.250605).
- [174] Minh C. Tran, Chi-Fang Chen, Adam Ehrenberg, Andrew Y. Guo, Abhinav Deshpande, Yifan Hong, Zhe-Xuan Gong, Alexey V. Gorshkov, and Andrew Lucas, “Hierarchy of linear light cones with long-range interactions,” *Physical Review X* **10** (2020), [10.1103/physrevx.10.031009](https://doi.org/10.1103/physrevx.10.031009).
- [175] Tomotaka Kuwahara and Keiji Saito, “Strictly linear light cones in long-range interacting systems of arbitrary dimensions,” *Physical Review X* **10** (2020), [10.1103/physrevx.10.031010](https://doi.org/10.1103/physrevx.10.031010).
- [176] Minh C. Tran, Andrew Y. Guo, Christopher L. Baldwin, Adam Ehrenberg, Alexey V. Gorshkov, and Andrew Lucas, “Lieb-robinson light cone for power-law interactions,” *Physical Review Letters* **127** (2021), [10.1103/physrevlett.127.160401](https://doi.org/10.1103/physrevlett.127.160401).
- [177] Tomotaka Kuwahara and Keiji Saito, “Exponential clustering of bipartite quantum entanglement at arbitrary temperatures,” arXiv preprint arXiv:2108.12209 (2021), [arXiv:2108.12209 \[quant-ph\]](https://arxiv.org/abs/2108.12209).
- [178] M. Perarnau-Llobet, H. Wilming, A. Riera, R. Gallego, and J. Eisert, “Strong coupling corrections in quantum thermodynamics,” *Phys. Rev. Lett.* **120**, 120602 (2018).
- [179] Peter Talkner and Peter Hänggi, “Open system trajectories specify fluctuating work but not heat,” *Physical Review E* **94** (2016), [10.1103/physreve.94.022143](https://doi.org/10.1103/physreve.94.022143).
- [180] Philipp Strasberg and Massimiliano Esposito, “Measurability of nonequilibrium thermodynamics in terms of the hamiltonian of mean force,” *Phys. Rev. E* **101**, 050101 (2020).
- [181] Peter Talkner and Peter Hänggi, “Colloquium: Statistical mechanics and thermodynamics at strong coupling: Quantum and classical,” *Rev. Mod. Phys.* **92**, 041002 (2020).
- [182] J. D. Cresser and J. Anders, “Weak and ultrastrong coupling limits of the quantum mean force gibbs state,” *Phys. Rev. Lett.* **127**, 250601 (2021).
- [183] Álvaro M. Alhambra, Georgios Styliaris, Nayeli A. Rodríguez-Briones, Jamie Sikora, and Eduardo Martín-Martínez, “Fundamental limitations to local energy extraction in quantum systems,” *Phys. Rev. Lett.* **123**, 190601 (2019).
- [184] M. Schwarz, O. Buerschaper, and J. Eisert, “Approximating local observables on projected entangled pair states,” *Physical Review A* **95** (2017), [10.1103/physreva.95.060102](https://doi.org/10.1103/physreva.95.060102).
- [185] Chi-Fang Chen, Kohtaro Kato, and Fernando GSL Brandão, “Matrix product density operators: When do they have a local parent hamiltonian?” arXiv preprint arXiv:2010.14682 (2020), [arXiv:2010.14682 \[quant-ph\]](https://arxiv.org/abs/2010.14682).
- [186] Cambyse Rouzé and Daniel Stilck França, “Learning quantum many-body systems from a few copies,” (2021), [arXiv:2107.03333 \[quant-ph\]](https://arxiv.org/abs/2107.03333).
- [187] Christian Kokail, Rick van Bijnen, Andreas Elben, Benoît Vermersch, and Peter Zoller, “Entanglement hamiltonian tomography in quantum simulation,” *Nature Physics* **17**, 936–942 (2021).
- [188] Christian Kokail, Bhuvanesh Sundar, Torsten V. Zache, Andreas Elben, Benoît Vermersch, Marcello Dalmonte, Rick van Bijnen, and Peter Zoller, “Quantum variational learning of the entanglement hamiltonian,” *Phys. Rev. Lett.* **127**, 170501 (2021).